FACULTY OF LAW, ARTS & SOCIAL SCIENCE
SCHOOL OF SOCIAL SCIENCE

A New Approach to Calculate and Forecast Dynamic Conditional Correlation –The Use of a Multivariate Heteroskedastic Mixture Model

CHENG LU

Thesis for the degree of Doctor of Philosophy

November 2010

A New Approach to Calculate and Forecast Dynamic Conditional Correlation –The Use of a Multivariate Heteroskedastic Mixture Model

Much research in finance has been directed towards forecasting time varying volatility of unidimensional macroeconomic variables such as stock index, exchange rate and interest rate. However, comparatively little is devoted to modelling time varying correlation. In this research, we extend the current literature on correlation modelling by reviewing existing time-series tools, performing empirical analysis and developing two new conditional heteroscedastic models based on mixture techniques. Specifically, Engle's standard DCC is augmented with an asymmetric factor and then modified so that disturbances (conditional returns) can be modelled using multivariate Gaussian mixture distribution and multivariate T mixture distribution. A key motivation of proposing mixture models is to account for the bi-modality observed in unconditional distribution of realized correlation. Besides, the ultimate purpose of incorporating this assumption to a multivariate GARCH is to account for a variety of stylized features frequently presented in financial returns such as volatility clustering, correlation clustering, leverage effect, fat tails, skewness and leptokurtosis. Since the model flexibility given this assumption can be greatly enhanced, after a thorough comparison we find significant evidence of outperformance of our models over other alternative models from a range of perspectives. Besides, in this research we also study a new type of correlation model using multivariate *skew-t* as basis for quantifying the density values of conditional returns. Note that, the ADCC *skew-t* and AGDCC *skew-t* model analyzed in this research are both new to the financial literature.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

This thesis would not have been possible without the encouragement and assistance from numerous peoples. First of all, I would like to express my sincerest gratitude to Prof. Taufiq Choudnry, who took over as my supervisor at the middle stage of my PhD. His comments on this thesis are highly appreciated. Besides, I also thank Dr. Gita Persand for her valuable guidance and suggestions during the start of my research work. While I was writing the first paper, the constructive discussion with Dr. James Chong from University of California Institute is proved to be crucial for all latter works. Meanwhile, I also owe a particular debt to Prof. María Concepción Ausín and Prof. Pedro Galeano in Universidad Carlos Ш de Madrid for their assistance on the programming for my second and third papers and Mr. David Baker at ISS for his help to launch my work on IRIDIS system. Moreover, thanks are also given to several colleagues and friends who provide consistent encouragement during my stay in Southampton. They are Dr. Karina Sudderson, Dr. Shuang Liu, Dr. Wei Yang and Dr. Jian Luo.

Finally, and most of all, I am grateful to my family, especially to my grandma and mother (Mrs Liu, WenYin and Mrs Liu, QingLin) for their patience, understanding and trust during all these years. No achievements of mine would be possible without their helps and encouragements. No difficulties would be overcome without their guide and patience. No words can really express my sincere respect to them now or forever. Tears and bloods will not wipe anything. What leaves on a stone will still remain. Difficulty may destroy a man, but it may also develop a man. It was them who lit me a torch while I was feeling cold and dark and it was them gave me the hope and will to survive when I was for the time in my life feel the desperate... For me, they are more than close relatives. They are my true friends and sometimes even spiritual guide. I will dedicate the rest of my life to love them.

谨以此文献给我最亲爱的外婆 (刘文英女士)
和尊敬的母亲(刘庆琳女士)

Chapter 1

Introduction

1.1 Research problem and Research aim

In finance, many time series of asset returns are characterized by serial dependence. It is due to the evidence documented in the literature that supports the findings of a positive autocorrelation in the variation of their conditional second moment. About three decades ago, researchers started to realize that volatility, just like return, can also be modelled as a time varying variable and its process tends to show persistent patterns. Since then, countless effort was put into exploiting traditional time series tools to modelling its dynamics and this trend has continued to the present day, although most attention previously paid to the univariate returns, has recently shifted to the multivariate context. For example, in the 1970s the main time series tool for modelling conditional return was the Auto-Regressive-Moving-Average (ARMA) model. Later, this technique was developed by Engle (1982) and generalized by Bollerslev (1986) to propose the famous GARCH framework, whose variants and extensions even today still dominate most of the literature on volatility forecasting.

Recently, benefiting from the reinforcement of globalization and advances in technology, much evidence shows that, not only volatility, co-movements of returns in different markets and of different asset classes are also becoming more and more significant, and univariate volatility is not only serially dependent on its own lagged term, but also correlated with others over time. Given this feature, the necessity of modelling covariance, as well as correlation, both to be time-varying is then highlighted. As Bauwens and Laurent (2002) illustrate, *"…recognizing this commonality through a multivariate model can lead to obvious gain in efficiency and more relevant financial decision-making than working with separate univariate models…"* Based on this motivation, a number of multivariate models are then proposed in the literature to capture the correlation dynamics. Among those most widespread tools, it is the Engle's (2002) DCC that successfully attracted most of the attention.

Since correlation in various financial applications is now an indispensible input and its importance nowadays is even more clearly recognized, the main aim of this research is then to examine, based on Engle's work, the efficiency of existing tools for modelling its dynamics and develop some new ones which can allow for more flexibility (either distributional or economical) so that hopefully a model capable of producing more accurate forecasts of the future correlation can be found. Implicitly, this research is motivated by questions like, 'in a financial market what really is a good estimate to depict association between returns?', or, in a similar vein, 'how can we develop an appropriate correlation model which can generate forecast both efficiently and accurately enough to predict the future correlation?', and 'how can these generated correlations be applied in the real world to generate economical benefits?'

1.2 Economic Contents

Since correlation has become the objective of this research, it is necessary to note some economic contents of this statistic and understand why it is important to generate accurate forecast of it. In finance, although in countless studies it has been proved that calculating this coefficient is not only necessary but also indispensable, its usefulness is frequently highlighted in only four major areas. These are portfolio selection, risk management, asset pricing and propriety trading.

First, concerning the asset allocation, correlation is a major input of Markowitz (1952)'s portfolio selection model to compute the portfolio variance on the aggregation level. Many hedge fund traders and investment managers use this coefficient to access the 'risk-return profile' of different assets included in a large opportunity set and decide which one to pick and the optimal weight to invest so that the overall holding risk of portfolio can be minimized and the corresponding return maximized.

Similarly, in risk management, to generate the next day's VaR one needs an accurate forecast of the entire covariance matrix. Nowadays, since a realized portfolio may contain hundreds or even thousands of assets including equities, derivatives and synthetic instruments, there is then an urgent need to find a flexible and cheap method for calculating large correlation (or covariance) matrices, to a given accuracy.

Besides, this coefficient can also be applied occasionally for pricing and hedging purposes. For instance, some exotic structures whose payoff depends on more than one underlying factor (e.g., interest rate spread or equity basket options) need correlation as an input to determine their fair prices. Meanwhile, if one wants to hedge these products, this association measure would also become indispensible. In even broader terms, this coefficient has been proved crucial not only for pricing some specific products but also for a range of assets. This is because even the fundamental asset pricing model such as Capital Asset Pricing Model (CAPM) needs this statistic as input to determine the unsystematic risk of a single asset relative to the whole market (see Sharpe, 1964).

In addition, due to the recent recognition of dynamic property for correlation, this coefficient is also considered as a risk factor, just like the time-varying volatility. Theoretically, in a derivative market where all other risks apart from this factor can be 'perfectly' hedged, an experienced trader can, by exploiting the difference of market expectation on this particular variable, make riskless profits. As a result, a new trading strategy called 'correlation trading' is then formed and recently (especially during the credit crunch) it has successfully attracted a lot of researchers' and practitioners' interests. Typically, when market risk (price changes of a traded asset) is the primary source of extracting return (or the sole factor to be hedged), strategies of trading correlation are analogous to those developed for trading volatility in equity markets. However, when credit risk (default of a credit product) is managed and exploited, a different class of trading method then needs to be used. Here, concerning this feature, since it is not like others which have been thoroughly reviewed and highlighted in the financial literature, an illustration is provided below to fill the gap.

Correlation trading

Generally speaking, there are two types of correlation trading strategies in financial market; equity-type ones and credit-type ones. As for the first, a correlation forecast, once generated, is often not directly inserted to a pricing model to exploit the price difference of a specific product but rather put through a filtering mechanism to calculate an intermediate quantity (variance-covariance matrix) so that, in the multivariate context, common volatility trading strategy can be performed. For example, we can use a correlation forecast of two currency pairs to determine the volatility forecast of their cross-products. Then, by inserting this volatility forecast into a standard currency option-pricing model, a forecast of the future prices of an ATM currency straddles can be derived. This price, after being compared with realized market prices, can be utilized to determine the opportunity of profitability (See Chong, 2004, for overpricing and underpricing of currency options). However, in the credit market, since correlation, for a variety of instruments, is a major input for the pricing formulas, profits then can be made directly from the mispricing of these products. This strategy, compared to the previous one, reflects the true virtue of trading 'correlation' expectation in the market. To see the details of how to perform these strategies, in the following passage we divide our discussions into two subsections.

a. Correlation trading in equity market

First, concerning the strategies adopted in equity market, as illustrated earlier, since they are inherently close to the volatility trading strategies, a proper understanding of the latter is beneficial to understand our current aim.

Volatility trading, as can be directly inferred from its name, is to trade the difference between market expectation and user's expectation on future volatility of a specific asset. Just like trading equity directly, the simplest way of trading volatility is to develop a linear 'contract' where the underlying instrument solely depends on the volatility of the target asset (or let us say that the payoff of this contract is an explicit function of volatility) so that profits can be made directly from trading this contract; e.g., buy the contract when we expect the volatility to rise and sell it when we expect it to fall.

In theoretical analysis, validity of this innovation has already been discussed. Brenner and Galai (1993) proposed a so-called realized volatility index and gave the futures and options written on it. Fleming, Osdiek and Whaley (1993) described the construction of an implied volatility index (VIX) whose derivative contracts are provided in Whaley (1993). In empirical applications, as a response to the immense demand, nowadays realized contracts of these volatility indices have also been introduced and listed in exchanges. For example, OMLX, a London-based subsidiary of Swedish exchange OM, launched the volatility futures in 1997, and Deutsche Terminborse (DTB) launched the VIX future contract in 2002.

Here, apart from utilizing an explicit contract, volatility can also be traded by combining a static position in a derivative product (option) and a dynamic (time-varying) position on the same underlying. For example, a common hedge fund trading strategy is to exploit the mispricing of convertible bonds listed in a financial exchange (or OTC). The strategy of holding a convertible and simultaneously delta-hedging the position is usually called convertible arbitrage. And the purpose is to find the risk-free profit from the mispricing of calls or puts that were embedded in the target convertibles due to the divergence of market expectation on volatility. Here, it is important to note that the hedging error (profit/loss) of this strategy is not totally determined by the gamma (second derivative of option prices to volatility) but theta (first derivative of option prices relative to time) as well. Thus, even if a profit is made, the result does not, as a whole, correspond to the forecast of volatility. To pursue a more 'purified' trade, one then needs a more volatility-specific contract (e.g., volatility swap). For example, Neuberger (1990) showed that by delta-hedging a contract paying log of the prices resultant hedging errors would accumulate to only the difference between realized volatility and fixed variance used in the delta hedge. That is, for this contract

$$P/L = \int_{T}^{T'} (\Sigma_h^2 - \Sigma_t^2) dt \tag{1.1}$$

, the holding period is now from T to T'. Σ_h^2 here denotes the implied volatility of target contract at T. Similarly, Dupire (1993) proved that a calendar spread of two log contracts would also serve the same purpose, as the payoff would equal to the variance difference between two maturities.

So far, the above trading methods are all exploiting the divergence of market expectation on future volatility, nothing has been said about how to use correlation as input to implement these strategies. Clearly, to achieve this goal, a bridge between volatility and correlation needs to be built in the first place. Often, this can be done by finding a triangular relationship between multiple-assets so that portfolio theory can be utilized. For example, in a three-currency trio, any currency pair can be regarded as an authentic portfolio comprising the other two. Thus, their univariate volatility and cross-correlation are related to each other and can be calculated interchangeably after a proper transformation of portfolio variance equation. Since volatility can be calculated using correlation as input, aforementioned strategies then can be adopted. For more details on this issue, in Chapter 5 we review some literature concerning trading correlation in equity market. To see its applications in the foreign exchange market, another example is given in the same Chapter.

b. Correlation trading in the credit market

Unlike the volatility trading strategies which have been repeatedly tested and implemented in equity markets for decades, in credit market, strategies of trading correlation are developed only very recently and the industry-standard model for pricing its base asset (CDO) was proposed only after we stepped into the 21st century. Although these products' appearance in the financial world is quite late, interests generated on trading them are massive, probably due to the rapid development of credit derivative markets in recent years.

In the credit market, a common way to trade correlation is to through a portfolio-based contract whose price is an explicit function of default correlation between individual credits included in this portfolio. Typical products of this type are synthetic CDO, Nth to default basket (NTD), CDO^2 and CDS index such as iBoxx. Here, we present an example using synthetic CDO. Depending on the level of default risk that expected cash flows of a CDO can bear, usually this product can be divided into three tranches: senior, mezzanine and equity. Senior tranche (credit) qualifies for an Aaa (Moody's rating) because defaults must wipe out both mezzanine and equity tranches before investors suffer any loss. The mezzanine layer, which has only the equity shield against losses, often carries a Baa rating while the unrated equity tranche then bears the risk of first dollar loss. Given these specifications, traders are now able to bet directly on the expectation of future realized default correlation by either longing or shorting a CDO that includes the target credit (for example by buying or selling a specific single tranche of a synthetic CDO) as a component. However, it should be noted that in the real terms delta-hedging and gamma-hedging are still indispensible when such products are traded, or the profit/loss could be affected by other contaminating factors.[1]

Consider now a credit exposure which has been properly dynamic-hedged and an expectation of future correlation generated, in order to realize this expectation, one can either long the equity tranche or short the senior tranche to long the 'correlation', or alternatively, for shorting 'correlation', one can either short the equity tranche or long the senior tranche. Now, we use an example to illustrate this strategy. In 2005, Standard and Poor and Moody's both dropped their ratings on the debt of General Motors and Ford below investment grade. At that time, a potentially profitable correlation bet would then be to long the equity tranche of a CDO and short the senior or mezzanine tranche. This is because, if the defaults stayed low, the return on the equity tranche would outstrip losses on the senior or mezzanine tranche. However, if defaults pick up, gains on the short position of senior or mezzanine would then at least offset losses on the equity tranche. Forming such a strategy implies that the market is now expecting the default correlation in a CDO to rise due to the simultaneous downgrade of two giant auto-manufactory firms (See FTSE Global Market, 2005).

1.3 Research scope

Now, we illustrate the scope of this research. Since the main task of this thesis is now to analysis various aspects of correlation in equity and foreign exchange market and we intend to achieve this goal by extending Engle's work to propose a more generalized framework than existing multivariate GARCH models for forecasting future correlation, the following strategy is adopted. First, based on Engle (2002), we let the dynamic covariance between two different assets follow a standard DCC-style evolving process. Then, an asymmetric factor, similar to

[1] Hedging the exposure of a credit derivative needs to adopt a similar procedure (entering into an identical offsetting position) that used in the equity derivative market. Say that a dealer has a long (short) position on a single tranche of a CDO, that is, he sold (bought) a protection. To hedge the marked-to-market risk resulted from a potential movement of credit spread on a single credit, he needs to buy (sell) the protection on this particular name that is included in an identical tranche. Here, delta of a credit is the amount of protection the dealer buys (sells) on that name to hedge the linear spread risk. Only small movements in the credit spread can be immune after dynamic delta-hedging. To protect the curvature of marked to market risk; one also needs to perform gamma hedging so as to isolate the spread convexity risk.

the one used to model leverage effect in the volatility process, is incorporated to the target dynamics. Here, we consider enhancing the flexibility of correlation models using mixture distributions. The proposed model is then defined as ADCC-MGM if returns are assumed following multivariate Gaussian mixture (MGM) distribution.[2] Besides, to allow for extreme events, we also consider the case where innovations are multivariate T mixture (MTM) distributed. Thus, an even more generalized framework can be constructed. That is ADCC-MTM. It should be noted that investigations into these conditional heteroskedastic mixture models are very rare in financial literature. To our best knowledge, the only research performed so far is by Bauwen, Hafner and Rombouts (2006).

As just mentioned, in fitting correlation dynamics, we use mixture models and it is mainly due to the flexibility concerns as a variety of stylized features can be steadily captured. However, at this stage it is also necessary to note another motivation of making this assumption. That is, unconditional distribution of realized correlation tends to show ‘multi-modality’. As for this feature, a detailed illustration with evidence will be given in Chapter 5. However, for now our emphasis is only on the generality of our new models. Indeed, ADCC-MGM and ADCC-MTM are so generalized that they can nest a variety of conditional correlation models. More importantly, they can be used to answer some unique questions like ‘Is the broad market now generating diverging (or new) opinions on future correlation, future volatility or future returns’ or ‘Might the co-movement between equity index of say European nations and that of the US change to another regime after the credit crunch?’. Besides, in more general terms, these mixture models can also be used to analyze linear interdependence, contagion issues and spillover effects.

Concerning their inferences, estimation of a multivariate GARCH is often performed by maximizing a log-likelihood function assuming Gaussian innovations because consistency of the resultant estimators can be ensured provided that conditional mean and variance are correctly specified (See Lee and Hansen, 1994, for convergence of QML in univariate setting and Jeantheau, 1998, for the multivariate case). However, here, to allow for more generality we adopt a Bayesian approach.[3] Specifically, a Monte Carlo Markov Chain (MCMC) technique, namely the Griddy Gibbs sampler, is chosen to calculate the mixture models’ inferences where each parameter of ADCC-MGM and ADCC-MTM is approximated using

[2] For asymmetric correlation, we mean that the correlations between different return series may appear to be dependent on the prevailing direction of the market. That is, one can expect to observe a higher correlation during the market crashes than in normal circumstances.

[3] Parameter uncertainty in Bayesian inference is allowed because parameter values in this paradigm are illustrated through a distributional form. More details on this issue will be illustrated in Chapter 5.

values of a series of random draws simulated from a specific kernel. The reason for choosing this numerical algorithm is to allow for the parameter uncertainty. Since estimated parameter values can now be illustrated through a density form, we can use this algorithm to obtain distributional characteristics of future correlation, future volatility and even future returns. However, in the classical inferential framework, even with a data-augmentation enhanced EM algorithm this task is still impossible.

Apart from the mixture models, in this research we also study a variety of alternative DCCs and examine their model performances from a range of perspectives including portfolio optimization and risk management. Here, concerning these competitors, it is especially worth noting two models, which we propose by combining the generality of AGDCC of Cappoiello *et al.* (2004) in capturing the covariance dynamics and flexibility of multivariate *skew-t* of Bauwens and Laurent (2002) in accounting for skewness, fat tails and high peakedness of a conditional distribution. As with mixtures, these models can substantially increase the flexibility of a standard DCC and, to our best knowledge, are also the first time studied and estimated in empirical research.

1.4 Structure of the thesis

Based on the goal and scope illustrated above, this thesis is now divided into three major parts. In the first part, we review various correlation measures and dynamic models developed to capture their evolving process and some inferential methods for estimating these models. Then, an empirical analysis is performed with emphasis put onto using existing time series tools and a market-implied information source for forecasting future correlation. Finally, we also initiate our own way for estimating correlation between different assets by exploiting a parametric and a semi-parametric (mixture models) technique. Clearly, some of the above issues are intrinsically related to each other. Therefore, overlapping illustration is unavoidable. However, we have tried to minimize this as much as possible.

The rest of this thesis is organized as follows.

In Chapter 2, to obtain a thorough knowledge of correlation coefficient, the target of this research, we start the description of it from the beginning. First, issues like its conception, assumptions and empirical potentials are stated. Then, two time series models and one stochastic model, all in their multivariate versions, for modelling correlation dynamics, are presented. Besides this, we also give a short summary of various stylized features shown in asset returns, and three methods to deal with them, so that the motivation and tools of extending existing models can be obtained. Meanwhile, some introductory illustrations on the inferential methods for estimating GARCH models are also provided.

In Chapter 3, since mixtures are now to become an integral part of this thesis, we review the formation, development history of this type of model and present some of its implementational issues and estimation methods. Meanwhile, since inference is to be calculated using a Bayesian method and this approach is intimately tied to the stochastic simulation techniques, we review some MCMC tools **in Chapter 4** with emphasis specifically put onto the Griddy Gibbs sampler. Note that, these two Chapters serve the similar purpose as Chapter 2 since majority of the contents are devoted to reviewing existing methodologies.

In Chapter 5, we use foreign exchange market as an example to perform empirical analysis of forecasting performance of a variety of existing correlation models. After analysis, an interesting finding is worth mentioning here. That is, unconditional distribution of realized correlation shows bimodality. This feature has important implication in finance because it provides a way to reveal the divergence of market views on future correlation. Given this rationale, a spontaneously solution to enhance the traditional correlation dynamics is then to incorporate its original structure to a new mixture model. And this step is taken in the next chapter.

In Chapter 6, we combine the aforementioned feature (bimodality in bivariate distribution), with some new ones (excess kurtosis, skewness, asymmetric correlation) to add to a standard DCC to form a so-called ADCC-MGM model and ADCC-MTM model. After presenting the specifications, we show how to estimate these models from a Bayesian's perspective. Specifically, for each parameter we start by giving a prior assumption for each of its marginal densities (mostly assumed uniform) and then obtain their posterior sampling kernels. A specific sampling sequence is given for each model and we also show how to generate correlation forecasts, return forecast, minimized variance and VaR based on them.

In Chapter 7, we report the posterior simulation results and forecasting performances of mixture models using two sets of simulated data and three sets of empirical data. Besides this, model performances of a variety of other DCC variants, including ADCC-*skew-t* and AGDCC-*skew-t,* are also analyzed and compared to one another.

Finally, **in Chapter 8** we provide the conclusion of this thesis with implications shown and directions for future studies presented.

1.5 Summary

In this Chapter, we introduce the main scope of this research. It includes presenting the motivation, economic contents, aim and structure of the whole thesis. Specifically, we are interested in analyzing the correlation dynamics presented in various financial assets. And our main aim is to device a new system, which is based on the current time-series modelling structure, for forecasting future correlation (or covariance) both accurately and efficiently. To achieve this goal, we implement two strategies. One is to utilizing the mixture modelling technique to incorporate a pre-specified distributional assumption to an enhanced DCC. The other is to combine a skewed version of standard distribution to another existing correlation model so as to form a new dynamics. Concerning the inference, we use maximum likelihood as well as a Bayesian approach to calculate (or approximate) the parameter values. And it is confirmed that, after enhancing the model sophistication, forecasting performance of standard DCC model does improve a lot.

Chapter 2

Literature review (part one)
- Correlation and its associated models

Introduction

The main task of this Chapter is to illustrate some preliminary issues concerning the correlation coefficient. Specifically, we will introduce its conception, assumptions and review some recent developments on its associated models (two time series and one stochastic) and their empirical applications in different financial markets. Besides this, we also present a brief overview of various inferential methods for estimating GARCH models to highlight the difference between maximum likelihood and the Bayesian approach that will both be implemented in our later empirical analysis.

2.1 Conception and assumptions

2.1.1 Conception

In statistics, correlation is defined as a quantity depicting the linear relationship between two or more random variables. Since it can tell how much one is proportional to another, in a variety of multivariate analyses this statistic is found useful.[4] If, for example, $(X,Y)^T$ are two relating variables with non-zero finite variance, their correlation $\rho(X,Y)$ can be computed using

$$\rho(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \tag{2.1}$$

where *Var(X) Var(Y)* represent the sample variance, *Cov(X,Y)=E(XY)-E(X)E(Y)* denotes the sample covariance. Under strictly increasing linear transformation, it satisfies $\rho(\alpha X+\beta,\gamma Y+\sigma) = sign(\alpha\gamma)\rho(X,Y)$ for any real numbers α,β,γ and σ.

2.1.2 Assumptions

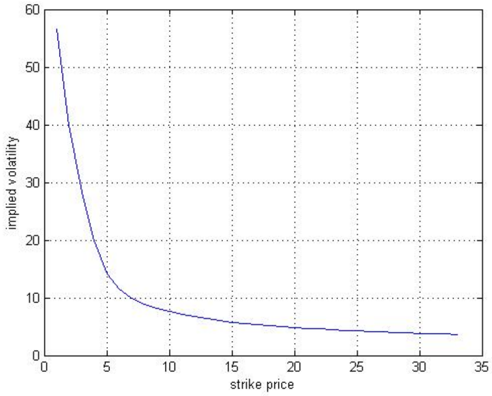
In equation (2.1), for the estimated correlation to be valid, usually three conditions need to be satisfied. First, causality between variables of interest needs to be tested and confirmed to ensure there is a realistic relationship between them. This step is essential because two variables, even without any inherent linkage, could still lead to non-zero correlation due to the pure coincidence. Second, to generate a valid correlation, it is required that underlying observations of two variables follow normal distribution not only individually but also jointly.[5] In finance, although one can argue from theoretical perspectives that, according to the *Law of large number,* multivariate Gaussian is a valid assumption for conditional distribution of asset returns, their unconditional distributions are frequently found to be non-Gaussian. Thirdly, it is important to stress that correlation coefficient can only be used to capture the linear dependence. Concerning this issue, consider now an example: If one is asked to calculate the correlation between X and |X|, instinctively, an immediate answer might be that these two variables have a non-zero correlation. Indeed, they are dependent on different domains, but, as a whole, are actually yielding zero correlation. Through this comparison, it is clear that correlation is actually a narrow-ranged dependence measure in statistics to depict relationships, and it is only defined on a linear space. Correct interpretation of this feature is important

[4] Proportional here means linearly related; that is, how much can the relationship be approximated by a straight line?

[5] The process of checking univariate normality is very easy. For example, one can rely on either Kolmogorov-Smirnov test or Shapiro-Wilk normality test to test the hypothesis. However, for multivariate normality, its associated test statistic is then far more difficult to calculate. Usually, we can, by performing visual analysis of the sample data, shed some light on this issue. For example, if the scatter-plot of bivariate data presents clear evidence of elliptical contour, then this data is very likely to be multivariate Gaussian distributed.

because, for many financial variables, their true co-movement is actually non-linear. Consider the typical volatility smile presented in the prices of an equity option for instance. Implied volatility is negatively correlated to the strike prices through a concave function. That is, implied volatility decreases rapidly when the strike price is relatively low, but much more slowly when strike goes high. Since the gradient of each point on this curve is different, we cannot rely on the simple linear analysis (correlation/a regression line) to properly depict the relationship between these two variables.

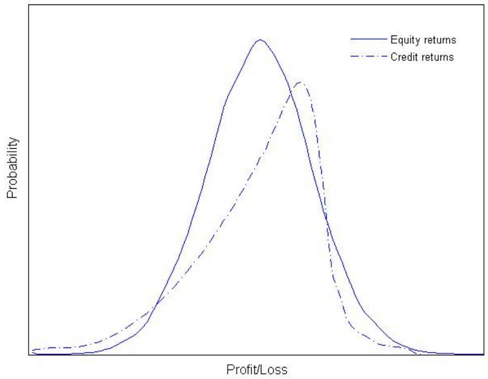
Typical volatility smile of an ATM equity option



Given the above assumptions, it is not difficult to note that the validity of using correlation in the real financial world is very easy to challenge. Actually, this is indeed the case, but not for all situations. For example, in credit market, returns of most instruments such as bonds, CDS and CDO apparently do not fit a normal distribution. They are inherently skewed because creditors usually have a strong probability of making a relatively modest profit on the interest of debt and a small chance of losing a large part of the initial outlay. In terms of a probability curve, these characteristics are translated into a thick left tail and an upside limit. However, in foreign exchange and equity markets, while the exact Gaussian is also seldom observed, massive evidence confirms that non-normality of conditional return is often caused by fat tails instead of excess skewness. For instance, for currency returns, little evidence can be found to support the significant asymmetry in their conditional (or unconditional) distribution and density of their returns usually presents an apparent bell shape. Given this feature, it is then fair to say that using linear correlation in these markets is theoretically more valid than in credit market.[6] To obtain a clearer view of the structural difference of probability density of credit return and FX/equity return, see below.

[6] For example, like Riskmetrix, 'Creditmetrix' is also a JP-Morgan-based institution which provides the industrial solution for credit research, analysis and trading. In their original model, they use the correlation between assets returns rather than the credit returns (a linear function of the credit spread) to calculate the default correlation of two or more credit instruments "*…This is probably because asset returns are more*

Probability density of asset returns (equity/credit)



As will be shown in later Chapters, since all of the empirical data used in this research are selected from either foreign exchange market or equity market and are, in most cases, assumed to be Gaussian distributed, correlation for our cases is then reckoned a valid statistic to compute. However, bear in mind that, theoretically, more prudent dependence measures (such as ranking statistic of Spearman's rho, Kendall's Tau and other copular variants) are also widely available in the literature. A full explanation of these alternative measures is beyond the scope of this research, we thus only provide a brief illustration of their mechanisms and characteristics in Appendix I.

2.2 Multivariate Correlation models

Above, we have described some introductory issues concerning the correlation coefficient. In the following, we illustrate three multivariate tools for modeling its dynamics. Since volatility and correlation are two inherently-related variables and a substantial amount of literature has already been dedicated for estimating univariate volatility, emphasis of this section are put onto illustrating those models using multivariate extensions of univariate volatility techniques for quantifying the correlation's evolving process.

Specifically, to incorporate the dynamic property, in the subsequent sections we gradually relax the assumptions (a constant covariance matrix and a constant correlation) implied in equation (2.1). That is, first, we let the evolving process of covariance (not correlation), as a whole, be generated from a specific dynamic mechanism. Thus, correlation is allowed to be time-varying. However, this is because covariance is now dynamic (EWMA, VECH, BEKK and SV). Then, we relax this assumption by modelling the variance process of each time series

closely related to the equity returns which tend to be more normally distributed…", See McGinty and Beinstein (2004).

in multivariate data separately and allow correlation itself to evolve dynamically. In this case, again, correlation is time-varying, but now due to its own dynamics (DCC and its variants).

2.2.1 Exponential Weighted Moving Average (EWMA)

Now, we illustrate EWMA model. The EWMA model is a common risk management tool initially developed by JP Morgan's risk management team to estimate time-varying volatility and covariance. Consider now a series of pseudo asset returns r_t or N dimensional R_t, this method computes univariate volatility by

$$\sigma_t^2 = (1-\lambda)r_{t-1}^2 + \lambda\sigma_{t-1}^2 \tag{2.2}$$

and multivariate covariance using

$$\Sigma_t = (1-\lambda)R_{t-1}R_{t-1}^{'} + \lambda\Sigma_{t-1} \tag{2.3}$$

where variance/covariance of the next day is computed by using squared return and variance/covariance observed today. Here, if (2.2) and (2.3) are initialized by setting σ_0 and Σ_0 equal to the sample variance/covariance, one can easily obtain a recursion function for calculating in-sampling volatilities. That is,

$$\sigma_t^2 = \sum_{j=0}^{\infty}(1-\lambda)\lambda^j r_{t-1-j} \quad \text{or} \quad \Sigma_t = \sum_{j=0}^{\infty}(1-\lambda)\lambda^j R_{t-1-j}R_{t-1-j}^{'} \tag{2.4}$$

In so doing, current smoothed values are then the exponentially weighted moving average of past squared returns. Hence, EWMA is also called exponential smoother.

Conditional variance for its k-day ahead aggregated return is $k\Sigma_{t+1}$, which means this model now assumes a flat-term structure for future volatility and will perform like a random walk in generating time-varying variance/covariance. However, note that most of the empirical findings in financial literature suggest that volatility is unlikely to follow random walk and it is undesirable to have a flat-term structure for forecasting purposes because all variance/covariance forecasts, once generated, will be the same for all forecast horizons of interest. Plus that volatility dynamics are now driven by a no-need-to-estimate parameter λ (no empirical fitting is needed), this method is then criticized by some researchers as being an insufficiently prudent approach for calculating volatility, and using industrial standard 0.94 for λ is not only arbitrary but also inefficient, though very easy.[7]

[7] In Riskmetric, decay factor λ is set to be a constant. 0.94 for daily data and 0.97 for weekly data.

However, it does not mean that, in covariance modelling, using this approach then cannot generate any significant benefits. At least, the positive definitiveness of covariance matrix can be inherently guaranteed since squared return and covariance in equation (2.3) are both insured to be positive semi-definitive. In addition, EWMA is also very easy to implement. For more details on this model, see the methodology section in Chapter 5.

2.2.2 GARCH series models

Since the temporal aggregation assumed in EWMA is implausible, there is then a motivation to propose a more flexible structure for modelling variance/covariance dynamics. In the univariate context, Bollerslev (1986), based on Engle's (1982) work, introduced a generalized version of ARCH model by combining the parsimony in parameters and flexibility in lag structure of conditional variance. The GARCH series model he proposed offers a convenient framework for modeling some key dynamic features of asset returns including volatility clustering, mean-reversion and long memory. Since covariance modeling through a multivariate GARCH is usually based on techniques developed in the univariate analysis, we briefly review some univariate GARCH literature in the following section before correlation modelling through a multivariate GARCH is highlighted and illustrated.

a. Variance modelling via univariate GARCH

First, we present the most parsimonious form of univariate volatility evolving process suggested in Bollerslev (1986). That is a GARCH (1,1),

$$\sigma_t^2 = \varpi + \alpha r_{t-1}^2 + \beta \sigma_{t-1}^2 \tag{2.5}$$

After repeated substitution, one can derive a seemingly EWMA type of dynamics from (2.5) as the current volatility is again an exponentially weighted moving average of past square returns.

$$\sigma_t^2 = \frac{\varpi}{1-\beta} + \alpha \sum_{j=0}^{\infty} \beta^j r_{t-1}^2 \tag{2.6}$$

However, note that there are crucial differences between these two approaches. In the GARCH model, parameters are estimated by a rigorous inferential method, unlike EWMA in which the parameters are set in an ad-hoc fashion. Besides, volatility stationarity is guaranteed since an expansion of (2.6) would eventually lead dynamics to converge to a constant long-run value ($\varpi / 1 - \alpha - \beta$). However, for EWMA, only a random walk with noises is assumed (see Harvey, 1989). Moreover, by using GARCH model one can at least obtain a volatility term-structure more realistically than the flat shape assumed in EWMA and higher-order specification can also be more easily incorporated.

Given so many advantages, numerous variants of GARCH model are then proposed in the literature and these new developments usually go into two directions. One is to increase the flexibility by changing the assumption of conditional return distribution so that extreme events in financial markets can be more easily captured than in the normal (Gaussian) environment. For example, Engle and Bollerslev (1986), in their treatment of asset return, used the *t* distribution to replace Gaussian to account for the excess-kurtosis (fat tails). Lee and Tse (1991) used Hermite polynomials to enhance a symmetric distribution and propose a so-called Gram-Charlier expansion method. Liu and Brorsen (1995) tested asymmetric stable density; Knight, Satchel and Tran (1995) implemented the double gamma distribution; Harvey and Siddique (1999) considered the use of a non-central student *t* distribution. (See also Brannas and Nordman, 2001, for a recent example of using log-generalized gamma distribution and a Pearson IV distribution with a univariate GARCH). Second, this type of model is also frequently extended in response to the leverage effect in volatility. That is, in financial markets, especially in equity markets, negative returns usually boost volatility by more than a positive return of the same absolute magnitude. To account for this effect, Nelson (1991) proposed the exponential GARCH (EGARCH) by adding natural logarithm to conditional variance. Glosten, Jagannathan and Runkel (1993) modified the variance equation by inserting a new lag-term to variance equation so that conditional volatility follows one process when innovations are positive and another otherwise. Furthermore, generalizations of their model (GJR) are also proposed in the literature. For instance, in Hagerud (1996) and Gonzalez-Rivera (1996), the authors added a logistic smooth transition function to volatility evolving process so that GJR can be obtained as a special case. Similar methods of using Taylor expansion or putting emphasis on conditional standard deviation instead of variance to account for the leverage effects were also suggested (see Sentana, 1995, for Quadratic ARCH and Zakoian, 1994, for Threshold GARCH). However, here, as far as the flexibility is concerned, it is then worth mentioning Ding, Granger and Engle's (1993) Asymmetric power ARCH, because their model is so generalized that all asymmetric GARCH models mentioned above can be nested.

b. Covariance/Correlation modelling via MGARCH

It is not difficult to note that the aforementioned GARCH literature focus on only univariate volatility. However, if the task is to model the volatility dynamic of a portfolio containing multiple assets, it is then necessary we could extend univariate GARCH techniques to multivariate versions for analysis.

Here, to propose a multivariate GARCH (MGARCH), usually two things need to be noted. One is to ensure the positive definitiveness of resultant covariance. The other is to keep the proposed model as parsimonious as possible. Concerning the second issue, it is important because computational cost is often a major concern for MGARCH models (i.e., our proposed correlation mixture models). Thus, it is often desirable we can device a function parsimonious enough for every covariance/correlation evolving process and tune the numerical algorithm before inference is actually calculated. In order to achieve this task, usually we can by performing a proper trimming in the parameter matrix to reduce the model dimensionality so that overall estimation cost can be alleviated to an acceptable level.

In the following, we briefly review several typical MGARCH models. As stated earlier, although the development of GARCH from univariate to multivariate has intrinsically allowed the calculation of correlation as an inner product of variance-covariance matrix, this statistic itself, in most of the early researches, was often assumed to be either fixed or following a stable deterministic process. For example, Bollserlev (1990), in his multivariate GARCH, once modelled the correlation using a constant. However, most empirical studies that attempted to verify his findings have failed to confirm the validity of this assumption. In fact, a large number of researchers find it quite reasonable to attest that correlations usually increase in periods of high volatility and that both magnitude and persistence of this statistic is affected by volatility, suggesting that this coefficient is more likely a time-varying variable.

To account for this feature, financial researchers then start to propose various generalizations of univariate GARCH. For example, Bollserlev *et al.,* (1988) and Engle and Kroner (1995) proposed solutions like VECH and BEKK which assume covariance to evolve according to

$$\text{VECH:} \qquad vec(\Sigma_t) = vec(C) + Avec(R_{t-1}R_{t-1}^{'}) + Bvec(\Sigma_{t-1}) \tag{2.7}$$

and

$$\text{BEKK:} \qquad \Sigma_t = CC^{'} + AR_{t-1}R_{t-1}^{'}A^{'} + B\Sigma_{t-1}B^{'} \tag{2.8}$$

where *vec*(.) is a column operator converting upper triangular elements of a N dimensional symmetric matrix into a N(N+1)/2×1 column vector and A, B are N(N+1)/2 squared symmetric matrix. Indeed, in above equations, correlation is now allowed to change over time. However, it is worth noting that its time-varying property was given only because the covariance matrix, as a whole, is now assumed to evolve dynamically. Besides, implementation of these models often involves various difficulties such as the curse of dimensionality and negative-definiteness. For example, VECH model is frequently associated with a very large parameter vector (21 parameters need to be estimated for calculating correlation in a bivariate case) with no guarantee of positive definitiveness for its resultant covariance. In the case of BEKK, although it can partially resolve the VECH's problem (positive definitiveness) by introducing a different parameterization, non-linear constraints usually have to be imposed in order to ensure the covariance stationarity. Taken together, empirical potentials of these MGARCH models are then rather limited. And it is not until recently a significant breakthrough was observed in this streamline of literature.

Engle (2002) generalized the CCC model of Bollserlev (1990) to put forward the Dynamic Conditional Correlation (DCC) model in which the variance-covariance matrix can be decomposed into two separate functions for modelling. One corresponds to univariate volatility; the other corresponds to time-varying correlation. Note that this separation is a crucial step to differentiate DCC from other MGARCH models because a decentralized estimation procedure which can resolve the large system problem is now provided (DCC can be used to analyze a large portfolio). For example, in other forms of MGARCH models, estimation is usually performed by maximizing the log-likelihood function with respect to the whole parameter set including those governing the univariate volatility process and those governing the covariance evolving process. However, for DCC, an appropriate univariate GARCH is fitted to each asset return in the first place (models will be different from asset to asset). Then, these returns, after being standardized by the estimated GARCH volatility, are fitted to another GARCH so that evolving process of an arbitrary covariance matrix can be modelled and finally correlation matrix after transformation can be obtained. Given this feature, a multivariate problem is then successfully decomposed to a series of univariate problems and it is reasonable to expect a substantially lower estimation cost. Besides, the correlation's dynamic property is now given without the help of any intermediate product (covariance).

Given these advantages, non-linear generalizations of standard DCC were then brought into light by various authors. For example, to allow for asymmetric response of conditional correlation to past shocks, Sheppard (2002) introduced ADCC by incorporating two factors. One is an asset-specific correlation news impact curve; the other is an asymmetric factor. To ensure the positive definitiveness, Cappoiello, Engle and Sheppard (2003) introduced the structure breaks and a BEKK-type parameterization. A similar property in Hafner and Franses (2003) is guaranteed by squaring the values of all correlation parameters. Besides these, here it is also worth mentioning Cajigas and Urga's (2005) AGDCC model in which asset returns are assumed to be asymmetric Laplace distributed. Note that their model is so generalized that all DCC variants mentioned above can be nested. Recently, new developments and refinements of standard DCC are still being proposed in the literature. For instance, to allow for the multivariate thresholds, Andrino and Trojani (2005) proposed the tree-structured DCC. To perform the sectorial asset allocation, Billio, Caporin and Gobbo (2006) introduced a block-diagonal structure to relax the common dynamics. And, by exploiting the Engle and Lee's (1999) idea of using different component specifications to quantify short- and long- sources that affect volatility dynamics, Colacito, Engle and Ghysels (2009) introduced the DCC-MADIS model. To see more details on these newly invented DCC variants, an illustration is provided in Appendix II.

2.2.3 Multivariate stochastic volatility models

Thus far, our discussion has explicitly focused on using time series models to capture the co-movement between multiple assets based on the assumption that covariance (or correlation directly) follows autoregressive processes. However, these time varying co-movements can also be captured using unobserved component models which assume the covariance (or correlation) to vary stochastically.

In the univariate context, stochastic volatility (SV) model introduced and popularized by Harvey, Ruiz and Shephard (1994) and Jacquier, Polson and Rossi (1994) has already been confirmed as a success in explaining the jump-diffusion process of volatility. Through either quasi-maximum likelihood or a Bayesian approach, its inference can be easily calculated. However, in the multivariate settings, as in the case of GARCH, it is then very difficult to generalize SV to allow for time-varying correlation.

As Bos and Gould (2007, p2) illustrate, *"…each possible choice for the parameterisation implies a certain restriction in either the space of the possible covariance or correlations. Also, allowing e.g. all correlations to evolve dynamically over time, can lead to a high number of parameters, even for a relatively low number of assets …"*

To the author's knowledge, very few pieces of literature are devoted to this topic and the only known contribution of studying multivariate SV with stochastic correlation is made by Yu and Meyer (2006). In their model, univariate return of each asset in a portfolio is assumed to have a SV type variance whilst correlation is modelled independently by a transformed random walk.[8] Concerning their inference, in the article by Harvey, Ruiz and Shephard (1994) multivariate SV model with constant correlation is estimated by quasi-maximum likelihood (QML) after the model is linearised so that standard Kalman filtering techniques can be

[8] Yu and Meyer (2006) used a rescaled sigmoid function to transform a random walk process to calculate correlation so that the resulting value is bounded in (-1, 1). If q_t now represents this random walk process, stochastic correlation is then modelled by $\rho_t = (e^{q_t} - 1)/(e^{q_t} + 1)$, where $q_t = q_t + \eta_{t;}\ \eta_t \sim N\,(o, \delta_t)$

adopted. However, while stochastic correlation is introduced, a more generalized way to deal with the non-linear state space model called ‘Single Source of Error’ (SSOE) then needs to be adopted (see Ord, Snyder, Koehler, Hyndman and Leeds, 2005, for details).

2.3 Inferential methods for GARCH models

Above, we have described three ways of proposing dynamic correlation models. Two are using time-series structures. The remaining is exploiting the stochastic theorem. Letting aside the flexibility, since in the real terms computational cost of estimating a state space model is frequently found substantially higher than fitting a time series model, it is then preferred we can use the first way to give arise to a new DCC. And it spontaneously becomes the target of this research. In this section, to meet this need, we provide an introductory description of some inferential methods for estimating MGARCH models after they are proposed. More detailed illustration on this topic can also be found in section 3.6 and Chapter 4.

Given a distributional assumption, inference of GARCH models is usually calculated by maximum likelihood (ML) or quasi-maximum likelihood (QML) through numerical approximation on the target log-likelihood derivatives. A specific optimization tool such as Newton-Raphson will be applied iteratively to search for a global optimum (if possible) for the parameter of interest until the convergence of the resulting estimator. To perform this task, it is often required that first-order derivatives of log-likelihood function (Gradient), as well as second-order derivates (Hessian), for each parameter can be found. Although the gradient function, given an analytical density form, is easy to generate, empirically, numerical differentiation of Hessian matrix especially for a MGARCH model is troublesome. To alleviate this difficulty, a popular method is then to exploit a result from Berndt *et al* (1974)’s studies on the system of simultaneous equations to replace the exact Hessian with an (asymptotically equivalent) matrix of outer products (OP) of Gradients. Often, to achieve the convergence, this method called BHHH requires a larger number of iterations than Newton-Raphson, but a much simpler calculation at each step. Fiorentini *et al,* (1996) took a further step to circumvent the non-trivial numerical approximation by obtaining a closed-form approximation of Gradient and Hessian for each parameter in a univariate GARCH. However, in order to locate the global maximum for the log-likelihood function, a mixed-gradient algorithm, which combines the estimated information matrix with the exact Hessian, is then needed. Among other works, here it is worth noting the asymptotic quasi-maximum likelihood (QML) estimator of Lee and Hensen (1994). In the univariate context and under lower-lever conditions, these authors proved that consistency of QML estimator can be ensured even if unconditional return is found not Gaussian-distributed (provided that conditional mean and conditional variance are now correctly specified). Similar evidence for MGARCH models is also provided (see Jeantheau, 1998 for consistency and Gourieroux, 1997 for asymptotic normality).

Apart from the classical inferential method (ML), inference of MGARCH models can also be studied using a MCMC algorithm. This stochastic simulation technique is usually performed in a Bayesian framework. Unlike ML, its aim is not to find a point estimator that can globally maximizing the log-likelihood function, but to reproduce the joint distribution of the whole parameter set. Since quantification of the resultant estimator is now given through a distributional form, parameter uncertainty attached to the model response is allowed. Besides, efficiency of the estimator is also ensured, but now by *Law of large Number* and *Central Limit theor*em.[9]

As Geweke (2005 p23) puts it, "*…Bayesian approach provides not only a more fluent communication between the investigator and potential results but greatly expands the choices of the models by considering uncertainty of parameters…*"

Given the capability of solving high-dimensional problems, the Bayesian method is, however, much less frequently applied in statistical literature to estimate quantitative models compared to ML. This is mainly due to the high computational cost associated with its implementation. For example, in the early days although a Bayesian statistician can steadily resolve a complex estimation task by either sampling a high-dimensional density directly or transforming this task into a series of unidimensional jobs, the appearance of a posterior density that was difficult to manipulate analytically was very common. Given this problem, one then had to use numerical approximation techniques rather than direct sampling to generate each new draws. In this case, calculating a high-dimensional integral was then often required and this task, for a non-analytical sampling kernel, was especially troublesome.[10] However, thanks to the innovations in stochastic simulation techniques and modern computational facilities, this problem was resolved after the monographs by Metropolis *et al.,* (1953); Hastings (1970); Geman and Geman (1984) and Gelfand and Smith (1990). Since the introduction of their MCMC techniques, simulation of a non-analytical function no longer needs to rely on a series of independent draws from the density of interest, but can use the realization of a specific

[9] In the simulation framework, Law of large number supplies the result that the more simulated values, the better the approximation. Central limit theorem offers a measure for the approximation error.

[10] Empirically, the joint posterior density (the kernel to be simulated) is usually high-dimensional. This is because, even for a very simple model, it usually contains more than two parameters.

Markov Chain, a series of dependent points, to approximate the target distribution. Given this relaxation of the assumption, the extent and potentials of the Bayesian approach in statistical learning are then considerably widened.

As cited from McLachlan and Peel (2000 p53), "*…with the advent of inexpensive, high speed computers and the simultaneous rapid development in posterior simulation technique such as the Markov Chain Monte Carlo (MCMC) methods for enabling Bayesian estimation to be undertaken, practitioners are now increasingly turning to Bayesian methods for the analysis of complicated statistical model…*"

Concerning its use in conditional heteroskedastic models, several attempts have been made in the literature and massive evidence were found confirming the informativeness of resultant Bayesian inferences. For example, Geweke (1989) used the importance sampling technique of Hammersley and Handscomb (1964) to estimate a univariate GARCH with Gaussian innovation. A similar attempt using student *t* for modelling conditional return is considered in Kleibergen and Van Dijk (1993). Besides, in the univariate context the Metropolis-Hasting algorithm is applied in Geweke (1993) to simulate posterior draws for IGARCH, while a Griddy-Gibbs sampler of Ritter and Tanner (1992) is used in Bauwens and Lubrano (1998) to estimate a MGARCH. Here, it is especially worth noting the work of Bauwens *et al.,* (1998) where, in the multivariate context, the authors conducted a thorough comparison of posterior results generated from three different MCMC techniques for estimating GARCH models. One is importance sampling; the other two are Metropolis Hastings (MH) algorithm and Gibbs sampler respectively. After several experiments, the authors found the importance sampler could provide an accurate estimate of the conditional moments, but was less precise in approximating marginal densities. Training of MH on GARCH often failed to explore enough of the tail behaviours. Only the Griddy-Gibbs sampler can produce most of the posterior characteristics accurately using a moderate number of random draw, although robustness of their resultant estimators does not come free. However, implementation of this algorithm is usually associated with massive computational time (See Chapter 4 for a more detailed explanation of this algorithm and other MCMC techniques).

Besides, MCMC algorithms are also found having a lot of potential in estimating state space models (or latent factor models). In particular, much research in this area has been performed to analyze stochastic volatility (SV) models. For example, Chib *et al.,* (2002) used Bayesian approach to estimate a high dimensional SV. Cappuccio, Lubian and Raggi (2004) provided recent evidence of using three different MCMC techniques suggested in Jaquier *et al.,* (1994,

1999) and Tierney and Mira (1999) to calculate the inference for a standard SV model where conditional innovations are assumed to be skew-GED distributed.

Apart from the typical forms of heterogeneity that have been thoroughly analyzed in the literature (like standard GARCH and univariate SV), recently there is another growing body of works which favour mixing exotic stochastic processes with simpler ones. For example, by mixing a standard autoregressive process, such as GARCH, with a flexible distributional assumption, one can propose a generalized volatility/correlation model so that the heteroskedastic, leptokurtic and heavy-tailed features of the financial time series can be simultaneously accounted. Taking the mixture distribution for instance, it is then natural to consider its use in conjunction with a MGARCH. This attempt in the literature has already been made and will be reviewed in the next Chapter. As for our purposes here, we only want to stress the fact that inferences of this type of models is often calculated by Bayesian approach since estimation of a large parameter set and a complicated likelihood function are now concurrently required. For instance, Ausín and Paleano (2005) used a variant of Gibbs sampler to estimate a univariate GARCH with Gaussian mixture distributed errors. The authors introduced a contaminating factor to link the variance of two component distributions so that probability of extreme events, which is determined by a high-variance Gaussian, can relate to the probability of normal events that are controlled by another low-variance Gaussian. A more generalized covariance evolving process assuming mixture distributed innovations is studied in Bauwens, Hafner and Rombouts (2006), where a diagonal VECH model this time is used.

2.4 Summary

In this Chapter, we provide some introductory descriptions of the correlation. First, some basic issues on this statistic including its conception and assumption are illustrated. Then, three types of models for capturing its dynamic property are presented. Among them, two are using time series tools. One is exploiting the stochastic theorem. Since the aim of this research is to propose a new DCC type model based on the Engle (2002)'s work, we describe the virtue of two inferential methods for estimating MGARCH models once they are proposed. Concerning the details of the motivation of proposing these new developments and Bayesian methods of estimating them, we illustrate and review them in the next two Chapters.

Chapter 3

Literature reviews (part two)
-Finite Mixture model and its estimation techniques

Introduction

The main purpose of this chapter is to review various aspects of the finite mixture model. This model is an integral part of this thesis and it lays the foundation for the conditional heteroskedastic correlation mixture models to be proposed in chapter 6. In the first section, we review some typical methods for tackling non-Gaussian features exhibited in the financial time-series and give arise to the motivation of using mixture model in this thesis to enhance the distributional characteristics to be assumed in our correlation evolving process. Then, in the next two sections, we respectively illustrate the main probabilistic properties, development history, mixing strategies and some implementational issues of this type of model and give two examples of it, namely, the multivariate Gaussian mixture (MGM) and multivariate T mixture (MTM). Finally, various techniques for estimating them are also briefly discussed. Specifically, we start by describing some introductory optimization tools proposed in the early days. Then, a comprehensive overview of iteration-based algorithms for fitting mixture models is provided. For those techniques developed after the 1970s, emphasis is put onto the classical-inference based EM algorithm and Bayesian-inference based Monte Carlo sampling methods.

3.1 Methods to tackle non-Gaussian features

As can be recalled from the last chapter, we have mentioned a current trend for proposing generalized correlation model is to mixing a standard autoregressive process (like a GARCH) with a flexible and plausible distributional assumption. Since our research work is partially based on this virtue, it is then beneficial to know the contributions that have already been made on this streamline of the literature. Here, to review these works, we start by illustrating some stylized features that are frequently exhibited in financial returns because these features provide the exact motivation of extending existing correlation models. And methods for tackling them can be directly transformed as a tool for developing new DCC variants.

First, a well-known feature of financial returns is their heavy-tailed distribution. In many foundational theories of mathematical finance, e g, option-pricing model of Black and Scholes (1973), portfolio theory of Markowitz (1952) and CAPM (APT) asset-pricing model, returns are unanimously assumed to be multivariate Gaussian distributed. Although, as a reasonable first approximation to the reality, it can give arise to a lot of tractable forms, empirically this conjecture is often found severely underestimating the probability of extreme events. In particular, during the aftermath of 1987's market crash and 2007's credit crunch, the deficiency of using Gaussian as a valid assumption for risk models is then clearly recognized. Besides, it is widely-accepted that high-frequency returns could also show asymmetry and high peakedness. However, an interesting finding is that these features could vary systematically from market to market. For example, FX returns are usually found high-peaked but approximately symmetric around zero whilst in equity market pronounced evidence of negative skewness is then discovered.

Given these features, to account for them is always very important for any financial models because their appearances are often directly related to the theoretical validity of the model inferences. In order to tackle them, usually we have three choices. One is to assume a proper stochastic process other than the general diffusion (with time-varying volatility and possibly mean-reverting) for conditional returns. Second is to fit a given parameter function or apply a so-called expansion method to reconstitute the conditional distribution being modelled. Finally, we can also use a semi-parametric technique (mixture modelling technique). In the following, we respectively describe these solutions.

a. Using a stochastic process to capture the stylized features

First, since unconditional returns are usually found non-Gaussian, many researchers then argue we could move beyond the traditional lognormal assumption by assuming a more appropriate stochastic process for price dynamics. In the literature, there are several works, which extend the traditional geometric Brownian motion, worth noting here. They are the pure jump process of Cox and Ross (1976), jump-diffusion developed by Merton (1976), and Lévy process suggested in Benhamou (2000). Using any of these processes for modelling conditional return can yield leptokurtosis and fatter tails than Gaussian in resultant distribution.

Take jump-diffusion as an example. This stochastic process models return using a Poisson mixture of Gaussian distribution so that total changes in asset price can be decomposed into 'normal' and 'abnormal' components.[11] The 'normal' component is modelled by a general diffusion process (Geometric Brownian motion) which is set up to capture the stock price dynamics without spikes. Discontinuous 'abnormal' component is given by a Poisson process which is applied only when a more-than-marginal change is observed. To define the Poisson component, usually three parameters are needed. They are frequency of a jump, its expected size, and the possible standard deviation of this jump within a short period of time. To calibrate the model, Beckers (1981) employed the method of cumulants; Ball and Torous (1983) studied the maximum likelihood; Henson and Westman (2002) applied the un-weighted least square.

Concerning the pure jump, it is a special case of jump-diffusion when the diffusion component in the later process is set to be constant. As for the Lévy process, its generating mechanism is the most flexible of the three. Since both continuous diffusion and discontinuous jumps can be included, this process provides the most generalized method at hand for modelling asset returns stochastically. For its applications in finance, see Benhamou (2000) for its implementation in option pricing and Gander and Stephens (2005) for its uses in stochastic volatility modelling.

b. Using expansion method or a parametric function

[11] Here, it is important to make a clear distinction between Poisson mixture of Gaussian and the finite Gaussian mixture, to be illustrated in later chapters of this thesis, since they are inherently related to each other. The Poisson mixture of Gaussian, according to Beckers (1981), models the density function $p(x)$ of daily asset returns using $\sum_{n=0}^{\infty}\frac{e^{-\lambda}\lambda^n}{n!}\phi(\mu,n\Sigma)$. However, a finite Gaussian mixture is based on Bernoulli mixing. That is, for a M component mixture, resulting density is now written as $p(x)=\sum_{n=0}^{M}\phi_n(\mu,\Sigma)$. Indeed, these two mixture densities are very close to each other. For example, for small value of λ , Poisson mixture and Bernoulli mixture are practically indistinguishable. This is because the sum of a series of *i.i.d.* Bernoulli variables will statistically approximate a binomial distribution, which will converge to the Poisson process if the number of these *i.i.d* variables included is now very large.

Above, we have just illustrated a method of capturing stylized features of financial returns through a direct modification of the stochastic process. However, since most of the modelling task is only to account for the non-Gaussian features in a distributional form, using this assumption to change the virtue of overall return dynamics is, then, clearly too restrictive. Here, a more straightforward and cheaper solution is to fit a given parametric or non-parametric function to conditional returns.

For example, if target returns only present features showing small deviations from Gaussian, we can apply a so-called cumulant expansion method (Edgeworth series or Gram-Charlier series). The virtue of this method is to augment a base density (say Gaussian) with an infinite sum of its cumulants (a series of Hermite polynomials) so that the base density can be reconstituted, showing small deviation in tail behaviours. This approach has been empirically proved useful in modelling weakly non-linear growth of fluctuations. However, a serious shortcoming is that its augmented *p.d.f* (probability density function) may sometimes be ill-defined. For example, it could assign non-zero probability to negative densities. Although the positive definitiveness of the resultant covariance matrix still can be ensured if, for example, one expands a symmetric distribution like Gamma using a series of Laguerre polynomials, empirical use of these methods in modelling financial returns is rare because characteristics of non-Gaussians presented in unconditional return distribution are usually significant (in the form of a much fatter tail and leptokurtosis).

To account for more leptokurtosis, countless researchers then start to use a parametric function, more generalized than those standard ones (Gaussian), for modelling return dynamics. Concerning this task, in literature there is a wide class of distributions one can choose. Apart from the elementary examples that have been repeatedly investigated, such as Beta, Gamma, Student *t*, Laplace and Lognormal, analyzing generalized forms of these simple distributions has also attracted a lot of interests. For example, Bookstaber and McDonald (1987) proposed the Generalized Beta distribution of the second kind (GB2) whose density presents a lognormal-style distribution shape. Karian, Dudewicz and McDonald (1996) introduced the Generalized Lambda distribution whose density also allows for a variety of shapes. Other potentially interesting ones include Generalized Exponential distribution of Nelson (1991), Asymmetric Exponential distribution of Fernández, Osiewalski and Steel (1995), Double Weibull distribution proposed by Mittnik and Rachev (1993), Double Exponential distribution suggested in Granger and Ding (1995) and Hyperbolic distribution given by Kuechler, Neumann, Soerensen and Streller (1999) (see also Engle and Gonzalez-Rivera, 1991; Hafner and Rombouts, 2004 for a non-parametric extension).

As can be easily noted, most of the densities mentioned above were proposed more than a decade ago and can be applied only in a univariate model; nowadays, however, since, in the financial context, more attention has been paid to the multivariate problems (i.e., using MGARCH models to fitting time-varying covariance matrix) there is an urgent need to introduce higher moments directly into a multivariate distribution. For example, by exploiting a result from Azzalini (1985), Bauwens and Laurent (2002) introduced the Multivariate Skew-Student *t* distribution. A generalization of the Multivariate Elliptical distribution is proposed in Branco and Dey (2001). Among others, here it is worth noting the Asymmetric Multivariate Laplace (AML) distribution of Kotz, Kozubowski and Podgorski (2003) because higher moments (both skewness and kurtosis) of their density are now incorporated by only one additional parameter. Since most of the modern financial models themselves are often associated with a very complicated specification, parsimony of this density is then clearly an advantage over other alternatives (see Hanson and Zhu, 2004; Sepp, 2004; Heyde and Kou, 2004; Cajigas and Uever, Urga, 2005; and Komunjer, 2005 for its applications). [12] Besides, if the flexibility is the only concern, it is then worth mentioning the Generalised Hyperbolic (GH) distribution of Barndorff-Nielsen (1977) (see Bibby and Sørensen, 2003; Barndorff-Nielsen and Sheppard, 2001 for an overview of its development). Note that, this density is so generalized that even AML is a limiting case of it. However, as a price to pay, its associated estimation cost is also massive. Thus, it is not surprising that this model is seldom applied in empirical analysis. However, several papers contributing to its developments are still worth mentioning here. For example, Mencia and Sentana (2004) analysed a GH distribution in a multivariate conditionally heteroskedastic dynamic regression model. Schoutens (2003) developed its use in the Lévy and Ornstein-Uhlenbck (OU) process. For a complementary review of other multivariate asymmetric distributions, see also chapter 7.

c. Using a mixture modeling technique

In addition to a single parametric or non-parametric function, in fitting multivariate returns some empirical studies have also confirmed the effectiveness of using a finite mixture distribution. The investigation of this distribution has a long history in statistics and its use can generate a lot of appealing characteristics. For example, as illustrated in McLachlan and Peel (2000 p46), *"...by adding up a sufficient number of component distributions any multivariate*

[12] The recent models used to account for the high moments (3rd and 4th moments) usually include at least two additional parameters. One is to capture the skewness, the other is to capture the leptokurtosis. However, in AML only one parameter is enough to capture both these two high moments. Therefore, its specification is parsimonious.

distribution can be approximated to arbitrary accuracy, moreover, an exact 'copy' of the original can also be expected if an infinite mixture of different contributions is used…" Since many stylized features such as multi-modality, skewness, excess kurtosis and heavy tails can be simultaneously included, using this method for modeling conditional returns then seems an ideal solution to increase the flexibility of a standard DCC although the specifications they give could be very complicated.

Empirically, most financial researchers are inclined to use Gaussian as component to construct standard mixture models (probably due to its numerically tractable density form). For example, to our best knowledge, Vlaar and Palm (1993) provided the first attempt to model innovations of a univariate GARCH to be Gaussian mixture-distributed. Ausin and Galeano (2005), based on Bai, Russell and Tiao (2003), performed a similar piece of work where a contaminating factor for modelling variance in different environments is included. In the multivariate context, Haas, Mittnik, and Paolella (2004), by extending the work of Wong and Li (2000), developed two distinct ways of proposing mixtures of Gaussian. One is to mix Gaussian distribution. The other is to mix Gaussian variables. Bauwens, Hafner and Rombouts (2006) provided the most recent evidence of incorporating this density to a covariance stationary VECH model.

Although it is known that, by mixing different Gaussians, a variety of density shapes can be easily reproduced; in empirical analysis the number of components included in such a mixture seldom exceeds two, due to the numerical cost concern. In such cases, to increase the flexibility, it is then preferred to introduce a more generalized density than Gaussian as component to construct the mixture. For example, Maclachlan and Peel (2000), in their research, proposed the *t* mixture model. Casarin (2003) studied the stable mixture. Haas *et al.,* (2005) introduced the Paretian mixture.

3.2 Finite Mixture Model

Above, we have highlighted some advantages of using mixtures for tackling financial return's non Gaussian features. Here, a point needs to be stressed is using this distributional form not only can yield some traditional benefits such as incorporating fat tails, more importantly, it can also allow for the multi-modality that usually cannot be captured by other methods. As to be shown in later chapters, unconditional distribution of realized return, volatility and correlation often present multiple peaks (see chapter 5). It is probably because heterogeneous groups of market participants are now simultaneously forming their expectation of how future market will move. Since their opinions are usually different from each other, in a distributional form such typical sign showing the divergence of expectation is then reflected through the multimodality. And it is nature to consider using a finite mixture model to tackle this problem. Given this motivation, in the following we describe how to build such a model.

First, we give its definition. Finite mixture model (FM), as can be directly inferred from its name, is a model where probability density of observations is formed by a discrete mixture of a finite number of single densities. Since its response data is generated by at least two different dynamic processes, this model provides a flexible, convenient and semi-parametric method for modelling sophisticated distributions.[13]

Consider a d-dimensional time series $\{y_t\}_1^T$ with T observations, if its probability density $\Phi(y)$ filtered by the past information set F_{t-1} is now given by a mixture of M component and each component is allowed to have its own distributional form. After data augmentation,[14] $\Phi(y)$ is then written as

$$\Phi\left(y_t \mid F_{t-1}\right) = \sum\nolimits_{m=1}^{M} \pi_m p_m\left(y_t \mid \varphi_m\right) \tag{3.1}$$

where $p_m\left(y_t \mid \varphi_m\right)$ denotes the density function of m^{th} component, φ_m represents its corresponding parameter set and π_m represents the weight parameter that satisfies $\left\{\pi_m > 0, \Sigma\pi_m = 1\right\}$ for all $m = 1 \cdots M$.

3.2.1 Development History

Since this model is very flexible in accounting for distributional characteristics, statisticians have been using it for a long time and the first attempt was made by the famous biometrician Karl Pearson in his classical 1894 paper where a moment-matching technique is used to fit a two-component normal mixture. However, after that, it suddenly lost its appeals among researchers and evaporated from the literature for a fairly long time. And it was not until Rao (1948) that this topic was reactivated again. This is because, in early days, estimation of all models had to be done by manual calculation. Indeed, estimating such a sophisticated model was inevitably a laborious task.

[13] The reason why finite mixture (FM) modelling is categorized as a semi-parametric technique is explained in Jordan and Xu (1995). Briefly, when the density functions of all components can be specified before mixing, the model is regarded as obtaining a parametric form. However, if the number of components is allowed to grow, it then leads to a non-parametric model. Here, a niche between both sides is then classified as semi-parametric.

[14] Data augmentation is a technique of introducing component labels to sample data so as to construct the full information set. Once this approach is adopted, usually one can obtain knowledge such as the fact that a specific observation is generated by a particular component in the mixture. For a detailed illustration of this technique, see Chapter 5.

However, after 1970, due to technological advances and development of some elegant iterative techniques, the advantage of using finite mixture was then re-addressed. For example, in classical inferential framework, Day (1969) and Wolfe (1970) formularized the first analytical maximum likelihood (ML) estimation procedure for Gaussian mixtures. By augmenting the existing observations with a latent variable, Dempster *et al.* (1977) made a revolutionary contribution by introducing a so-called Expectation Maximization (EM) algorithm for fitting various mixtures (see Aitkin and Aitkin, 1996; Titterington, Smith, Makov, 1985; and Mclachlan and Basford, 1988, for details). Similarly, in the Bayesian frameworks, the inferences of these models are also studied and a stochastic simulation technique called MCMC was developed by Tanner and Wong (1987) and Gelfand and Smith (1990) to perform the mixture learning.

Nowadays, benefiting from the availability of much cheaper computing facilities, the extent and potential of mixture model are even more widely extended. Its applications now can be traced to many different areas for modeling random phenomena. For instance, in statistics, apart from the traditional use of mixture models in cluster analysis, this technique is now also applied to survival analysis, discriminant analysis and image construction. For a more detailed review of these issues, see Everitt (1996), McLachlan and Peel (2000) and Dias (2004).

3.2.2 Standard Mixtures and Hybrid Mixtures

Given equation (3.1), we have two ways to construct a mixture model. One is to choose all components from the same parametric family to build a so-called standard mixture. Meanwhile, we can also select components from different distributional groups to form a hybrid mixture. As far as the flexibility is concerned, the hybrid way is usually considered as a better choice than its alternative because different styles of distributional characteristics can be simultaneously included. However, in practice most researchers are still inclined to use standard mixture for modeling heterogeneity because its associated computational sophistication is much lower. And among various choices it is those whose components are formed by distributional variants included in the exponential distribution family that are used the most in empirical researches (for example, Gaussian mixture). In the following, we respectively describe these two ways of forming mixtures.

a. Standard Mixtures

First, for constructing standard mixtures, in statistics we have a variety of choices. However, as far as popularity is concerned, it is then especially worth mentioning the multivariate Gaussian mixture (MGM) because this model is the one that is most frequently applied by different financial researchers. Although its specification is relatively simple, given sufficient number of components included, its flexibility is usually considered as enough for capturing the high moments of asset return to a given accuracy. And one can observe a lot of its applications in empirical researches. For example, in the context of stochastic modeling, Labidi and An (2000) used MGM to analyze the equity index returns. McNeil, Nyfeler and Frey (2001) applied it to model credit product returns. In risk management, current version of RiskMetrics^{TM} employed a two-component MGM to evaluate the market-risk models. A similar approach to calculate Value at Risk (VaR) is adopted by Venkartaraman (1997). Besides, this mixture is also used in several cases to resolve the asset allocation problems. For example, in Buckley *et al.,* (2002) the authors used Gaussian mixture to fit the returns of a hedge fund portfolio and then generated the optimal investment weights for each asset.

In recent years, increasingly, apart from the above example, attempts are also made to propose standard mixtures using components other than Gaussian. For example, McLachlan and Peel (2000) proposed the multivariate T mixture; Kuester, Mittnik and Paolella (2005) studied the multivariate GED mixture. Haas, Mittnik, Paolella and Steude (2005) introduced the multivariate stable Paretian mixture. Here, concerning these models, it is necessary to note that their respective advantages are different, although generality is roughly the same. The first two are especially good at accounting for tail-behavior, whilst the third outperforms others only from a theoretical perspective. Specifically, as pointed out by Mandelbrot (1963, p5) and Fama (1965), since "*…stable Paretian is the only valid distribution that can arise as a limiting distribution for the sums of i.i.d random variates…*", there is then a motivation to use this distribution as a theoretically valid assumption to propose mixture for modeling return dynamics, as the logarithm of asset return itself is known to follow additive principle based on the central limit theorem.[15] (See also Mittnik and Rachev, 1993a and b; Rachev, Kim and Mittnik, 1999; and Rachev, 2003, for more details on stable Paretian).

b. Hybrid Mixtures

Compared to the standard mixture, hybrid mixing is a strategy which can outlines the true virtue of 'mixture'; however its empirical applications are not as numerous as its alternative.[16]

[15] Return of a financial asset given the prices at time 0, P_0, and at time 1, P_1 is generally depicted in the form of $R_1 = (P_1 - P_0)/ P_0$. However for the ease of capturing stylized factor, it is also popularized by various authors adding a logarithm to the above function. Meanwhile, the difference between returns can be even more magnified after being multiplied by a constant throughout the sampling period. Thus, in a log-return series, the return is said to follow the additive principle; for example, the weekly return is the summation of the *i.i.d* daily return. The daily return is then the summation of even higher frequency returns if it is empirically available.

[16] 'Mixture' in the common sense is to make a combination of things with different characteristics. However in the finance literature, its application seems more concentrated on the mixture of the same distribution but with

And to our best knowledge, it was not until 1970s this type of model was formally introduced in the finance context.

DuMouchel (1973) was the first to use a univariate mixture of normal and stable Paretian to model distribution of common stock prices. He found that the excess kurtosis and fat-tails that frequently characterize the return distributions could be remarkably well captured by his model. A similar strategy is adopted in Bones *et al* (1974), where evidence supporting the superiority of hybrid mixing was found again. Here, concerning their models, it is necessary to note that the mixture components were chosen based on the traditional 'stable' law; that is, the density can allow for possibly different behaviors in different segments of sampling data or, in a similar vein, it is expected that, within different segments, one will only be able to observe minor changes with small probability. Although this theorem was favored by researchers in early days such as Mandelbrot (1963) and Fama (1965), recent investigations show that more coherence to the empirical data can be achieved when sudden breaks or jumps are also taken into account. For instance, we can apply the 'geometric stable law' to asset return so that stability of a dynamic process is preserved only before the occurrence of an unexpected shock. Rachev and SenGupta (1993) tested this hypothesis and proposed an alternative to DuMouchel (1973) by suggesting a combination of Laplace and Weibull distributions. By replacing Gaussian component with the geometrically stabled Laplace distribution and stable Paretian with Weibull distribution, they found the significant evidence of outperformance of their model over DuMouchel (1973)'s.[17]

Besides this, hybrid mixtures are also analyzed in some recent studies. For example, in Haas, Mittnik and Paolella (2005), the authors proposed two different ways to construct such models. One is to exploit a result from Kanji (1985) and Jones and Mclachlan (1990) to combine two components discretely. The other is built based on the principle that conditional return is a weighted sum of two differently distributed random variables. Here, to better understand their difference, we use, as an example, Gaussian and Laplace as components, to see their resultant density functions. As for the first, since Gaussian and Laplace are now mixed in the traditionally discrete way, its resultant density after mixing can be easily written as

$$\textit{First Mixture}: \qquad f_1(x) = \frac{\pi_1}{\sqrt{2\pi}} e^{-x^2/2} + \frac{1-\pi_1}{2} e^{-|x|} \quad \pi_1 \in [0,1] \qquad (3.2a)$$

multiple components.

[17] Gaussian distribution follows traditional 2-stable law. Laplace, or double exponential distribution, follows 2-geometric stable law. For a more detailed illustration of this issue, see Robbins (1948), Gnednenko and Fahim (1969).

where π_1 denotes the proportion of Gaussian-generated observations in all training data and each observation is generated from either a Gaussian density or a Laplace density. However, concerning the second, its density is then given by

$$\begin{aligned} &\textit{Second Mixture}: \quad X = \pi_1 G + (1-\pi_1)L; \\ &f_2(x) = \frac{1}{2(1-\pi_1)} e^{\pi_1^2/2(1-\pi_1^2)} A\left[\Xi\left(B+C\right) + \Xi\left(-B+C\right)\right] \qquad (3.2b) \\ &A = e^{x/1-\pi_1}; \quad B = -x/\pi_1; \quad C = -\pi_1/1-\pi_1 \end{aligned}$$

where π_1, although still called the weight parameter, now denotes the proportion of Gaussian variate in the calculation of each observation and Ξ represents the *c.d.f* of a standard normal distribution. Obviously, these two functions are now far from similar. One is calculated using linear combination whilst the other is generated using first derivative of *c.d.f.*

3.2.3 Implementational Issues

In this section, we give assumptions for constructing finite mixture model and illustrate some of its implementational issues such as the number of components to be included, identifiability of each observation and parameter restrictions to be added. All of these issues are important for ADCC-MGM and ADCC-MTM to be proposed in Chapter 6.

a. Independence assumption

First, it is necessary to mention a theoretical assumption for constructing mixtures. That is, the response data needs to be assumed at least locally independent. This conjecture is a relating but weaker assumption than *i.i.d.* The only difference between them is a conditional argument. Local independence indicates the statistical irrelevance of different observations conditioned on a series of component labels. However, such requirement is not needed in the later case (see Dias, 2004). Since component labels now play the essential part in understanding the inference, we describe them below.

According to Everitt (1996), a component label is a latent variable that conveys the information about a particular observation, say y', that is generated by which component in the mixture. If this knowledge is acquired, then, according to local independence each observation will be provided with a specific label and appear independent of one another and target likelihood function of the joint density can be written as a multiplication product of all its marginal densities so that different inferential approaches can be adopted to estimate parameters.

Here, one of the most intriguing advantages of using this assumption is training data not required to be independent before conditioning. Thus, y_t could be a totally *i.i.d* time series or dependent before conditioning but independent thereafter. As for the second case, mixture model then burgeons into another popular framework, Hidden Markov Model (HMM) of Baum and Petrie (1966). Note that, this class of model provides an alternative to finite mixtures and since its invention has also attracted a lot of interests (see Bye and Schechter, 1986, for latent Markov model, and Chib, 1996, for Markovian mixture model). [18]

b. Number of Mixture Components

From equation (3.1), we can easily know that two things usually need to be determined before constructing a mixture model. The first is to choose the number of component to be included, then, the distribution functions for each. Concerning the first issue, although a number of theoretical researches have already been done, a common criterion for choosing M is still not found. Thus, in majority of the cases this task is still mainly performed by visual analysis. For instance, if no prior information is available, M is usually chosen by accessing the number of modes found in histogram plot of sample data. However, a clear drawback of this approach is the components then need to be fairly wide apart in order to be detected. To obtain a more objective result, information-based model selection criteria, such as AIC of Akaike (1973) and BIC of Schwarz (1978), are then needed. As Roeder and Wasserman (1997, p23) argued, "*…When a normal mixture model is used to estimate a density non-parametrically, the density estimates that use BIC to select the number of components in the mixture is usually consistent…*" However, in the finance context, countless authors confirmed that, often, a two-component mixture is already flexible enough to capture the stylized characteristics exhibited in asset returns. Thus, assessment of component number is then usually not a major task. To see more on this particular issue, a good overview can be found in McLachlan and Peel (2000).

c. Identifiability of mixture component

Besides, for a mixture model, to calculate its inference, one also needs to ensure the identifiability of each component. That is, before an iterative procedure is adopted, knowledge of component label for each observation needs to be acquired first. In a hybrid mixture, this task is very easy because distributional functions of each component are already given differently, which intrinsically allows the demarcation of group data. However, the problem does arise when standard mixtures are estimated. Since only one parametric function is to be

[18] By assuming that the latent variable follows a Markovian process, usually a first order HMM is flexible enough to capture all characteristics of a finite mixture model.

inserted, labels of a component once decided may still switch again. Thus, the task of identification could become very troublesome in an unconstrained setting.[19]

In this case, a parameter restriction then usually needs to be imposed to resolve this problem. For example, Aitkin and Rubin (1985) favored constraining the weight parameter π_m to follow an ascending order $\pi_1 \le \pi_2 \le \cdots \le \pi_M$ so that each component in a standard Gaussian mixture can be numerically identified. A similar approach is adopted in Bauwen and Lubrano (2006), where a descending order is considered. Here, if only two components are allowed, one can also use $\pi_1 > 0.5$ to replace $\pi_1 > \pi_2$ (see Galeano and Ausin, 2005 for example). Besides, in several cases, this attempt is also made through the restriction imposed on the weight parameters, e.g. to let $\mu_1 \le \mu_2 \le \cdots \le \mu_M$ so that means of different components can be identified. However, in the literature, such applications are far less frequently applied than the previous one because evaluation of mean parameter is usually found more complicated than weight parameter in either classical or Bayesian inferential framework.

d. Parameter Restriction

Finally, since mixture model even in its most parsimonious form is very likely to be associated with a complicated log-likelihood function, it is then usually preferred a proper trimming of the target parameter set of interest could be considered or certain subjective restrictions imposed. Although such trimming and restrictions will inevitably lead to loss of generality, the reduced computational burden is often considered as more valuable for empirical analysis. For example, Bauwens, Hafner and Rombouts (2006) assumed the mean vector of their training data to equal zero so that, in a two-component mixture, mean parameters of only one component needs to be calculated stochastically, whilst the other analytically. In so doing, sophistication caused by the numerical sampling then can be partially alleviated.

3.3 MGM and MTM

Now, we present the density functions of two standard mixtures to be applied in our latter correlation modelling. One is Multivariate Gaussian Mixture (MGM). The other is Multivariate T Mixture (MTM). Here, we choose Gaussian and *t* as components to construct

[19] In this thesis, to confirm the existence of ‘interchanging identifiability’, we perform a posterior sampling without imposing any restriction on the weight parameters. After experimenting, we find that neglecting this problem leads to seriously biased results. For example, the posterior draws of weight parameters, π_m, is trapped in a very narrow space after a mild number of iterations and their values hardly change thereafter. Since the parameter space cannot be explored completely, it is then extremely difficult to have a clear identification of which mixture component really determines the next data.

mixtures because of their intuitive simplicity, numerical tractability and model flexibility.

a. Multivariate Gaussian Mixture (MGM)

Analysis of the MGM model has a long history in statistics. Using this type of mixture has many advantages. For example, as McLachlan and Basford (1988, p45) illustrated, "*...any continuous distributions can be approximated arbitrarily well by a finite mixture of Gaussian distributions with common variance...*" To define its specification, one only needs to replace distribution function $p_m\left(y_t \mid \varphi_m\right)$ in equation (3.1) with a Gaussian *p.d.f*, say $\phi_m\left(y_t \mid \varphi_m\right)$. Then, a *d*-dimensional *M*-component MGM can be given as,

$$\begin{aligned} f\left(y_t \mid F_{t-1}\right) &\sim \sum_{m=1}^{M} \pi_m \phi_m\left(y_t \mid \varphi_m\right) \\ &= \sum_{m=1}^{M} \pi_m \left(2\pi\right)^{-d/2} \left|\Sigma_{mt}\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(y_t-\mu_m\right)' \Sigma_{mt}{}^{-1}\left(y_t-\mu_m\right)\right\} \end{aligned} \tag{3.3}$$

where μ_m, Σ_{mt} denotes the mean and time-varying covariance of m^{th} Gaussian component and Σ_{mt} here is required to be a *d-by-*d symmetric, positive definitive matrix.

b. Multivariate T mixture (MTM)

Although, by using a large number of components, one can be assured that the tail behavior of resultant Gaussian mixture would be very flexible. An immediate cost of performing this strategy is the substantial increase of sophistication in its inference calculation. Therefore, in order to account for the extreme events in a more cheap way, we might need to consider using a more generalized distribution than Gaussian to construct mixture, but not increasing the numbers of components to be included.

Here, an easy solution is to choose a Multivariate T mixture (MTM). This model can provide a cheap and robust generalization to Gaussian Mixture. Not only is a heavier tail allowed, MTM can also obtain MGM as a limiting case whenever its degree of freedom parameter approaches infinity.[20] Since *t* itself is often considered a scaled mixture of normals, MTM constructed by using this distribution as component can then be regarded as a 'Mixture of Mixture' (see Tukey, 1960, for using a contaminated Gaussian mixture to construct *t,* and Huber, 1964, who used an integration technique to provide its generalization. In Appendix III, we have described the hierarchical mixture formation of a standard *t* in more detail).

[20] For example, consider a *d*-variate random variable $y_t = \sqrt{\Sigma_t}\varepsilon_t$, if the innovation ε_t is now *i.i.d* multivariate *t* distributed according to $t_v\left(0,1\right)$, y_t then follows the same distribution but the variance is $\Sigma_t \cdot v/(v-2)$ for all *v*>2. When the degree of freedom parameter $v \to \infty$, y_t tends to be Gaussian distributed, since $\lim_{v\to\infty} v/(v-2)=1$, and the variance of y_t is just equal to Σ_t.

Now, to construct such a model, as before, we only need to substitute $p_m\left(y_t \mid \varphi_m\right)$ with a *t* density $t_m\left(y_t \mid \varphi_m\right)$. Then, a *d*-dimensional *M*-component T mixture model can be easily given as,

$$y_t \mid F_{t-1} \sim \sum_{m=1}^{M} \pi_m t_{v_m}\left(y_t \mid \mu_m, \Sigma_m\right)$$

$$f\left(y_t \mid \varphi\right) = \sum_{m=1}^{M} \pi_m \frac{\Gamma\left(\left(v_m+d\right)/2\right)}{\left(\pi \cdot v_m\right)^{d/2} \Gamma\left(v_m/2\right)\left|\Sigma_{mt}\right|^{1/2}} \left(1+\frac{\left(y_t-\mu_m\right)' \Sigma_{mt}^{-1}\left(y_t-\mu_m\right)}{v_m}\right)^{-\frac{v_m+d}{2}} \qquad (3.6)$$

where v_m is a positive scalar denoting the degree of freedom parameter of m^{th} *d*-variate multivariate *t* component and $\Gamma(\cdot)$ is the Gamma function satisfying $\Gamma(x)=\int_0^{\infty} t^{x-1} e^{-t} dt$.[21]

3.4 Mixture Model Estimation Techniques

In this section, we start to illustrate the mixture model estimation techniques. It is a major aim of this chapter. First, a brief illustration of some simple methods is provided. Then, a comprehensive overview of iteration-based algorithms is given and we put the emphasis on classical inference-based techniques and Bayesian inference-based algorithms. As has been illustrated in chapter one, since we are going to use a MCMC algorithm (Bayesian) to estimate correlation mixture models we dedicate the next chapter to a detailed illustration of issues concerning this simulator. However, for now we only provide an overview of alternatives to this technique and use one of them to estimate ADCC-*skew t* and AGDCC-*skew t,* also proposed in this research. For a similar review, see Titterington *et al.,* (1985), Everitt and Merette (1990), McLachlan and Peel (2000) and Dias (2004).

3.4.1 Simple Methods

In a mixture model, given that the number of components is known, there are a lot of techniques that can be used to estimate its parameters. Among various alternatives, early methods such as graphical analysis, method of moments and minimum distance are easy to implement although their resulting estimators sometimes are found inefficient.

[21] For any positive integer *x* inserted to a Gamma function, we can use $\Gamma(x)=(x-1)!$ to calculate its resultant value. However, if *x* is very large, sometimes an approximation function (Stirling) then needs to be used. That is, $\Gamma(x) \approx e^{-x} x^{x-1/2} \sqrt{2\pi}\left(1+\frac{1}{12x}+\frac{1}{288x^2}+o\,|\,x\,|^{-3}\right)$

In Section 3.1 we have stated that the first attempt to fit the mixture model was made by the famous statistician Pearson. In his classic 1894 paper, five parameters in a heteroskedastic Gaussian mixture model were calculated by solving a ninth degree polynomial using method of moment. Undoubtedly, finding the roots for a nonic by manual computation in 1894 required a lot of effort. Although its estimation procedure is sophisticated, in the last century using these methods has still attracted some interests. For example, Quandt and Ramsey (1978) used moment generating function to calculate inference of a two-component mixture model (for similar works, see also Lindsay and Basak, 1993, and Furman and Lindsay, 1994). A more popular way is to adopt a so-called minimum distance strategy. The virtue of this technique is to calculate the parameter value so that the distance between empirical distribution and the proposed mixture is minimized. Since its resultant inference would be very sensitive to the method chosen for computing the distance, it is often beneficial to use a variety of measures. For example, Choi and Bulgren (1968) examined the Wolfowitz distance; Yankowitz (1969) studied Levy distance; Macdonald (1971) used Cramer-von Mises distance. For a more comprehensive review of the characteristics of these distance measures, see Titterington *et al.* (1985).

3.4.2 Classical-inference based iterative methods

Compared to the simple methods, if the task is to estimate a model having complex specifications, a better choice is to adopt an iteration-based inferential approach because this type of method is capable of producing statistically more efficient estimators.[22] If the convergence of parameter values can be confirmed, seldom, substantial approximation errors will be generated. However, as a price to pay, intensive computational work then becomes inevitable. Thus, it is worth mentioning it is the recent advent of high-speed computing facilities that really accelerates the development of these algorithms in the mixture model context. In the following, since the focus of this thesis is on mainly Bayesian inference, we only briefly illustrate several typical classical inferential methods.

3.4.2.1 Maximum Likelihood (ML)

First, for estimating mixture distribution, we start the illustration of classical-inference based methods with maximum likelihood or ML. The aim and estimation procedure of this approach is given below. Consider a mixture distributed random variable y with totally T observations;

[22] The reason why iteration-based method can produce a more practical and efficient estimation procedure than the simple methods is because an analytical solution for the parameter estimation is generally difficult to find in a mixture distribution. As cited in Titterington (1996),'*…the main reason for the huge amount of literature on estimation methodology for mixtures lays in the fact that explicit formula for the parameter estimates are typically unavailable….*'.

the objective of ML is to find an estimator for parameter φ, say $\hat{\varphi}$, that, in regular conditions, can maximize the likelihood function $L(y\,|\,\varphi)$ or the log-likelihood function $\ell(y\,|\,\varphi)$.[23]

Use the mixture distribution defined in equation (3.1) for example, since $\ell(y\,|\,\varphi)$ is given by

$$\begin{aligned}\ell(y\,|\,\varphi) &= \log L(y\,|\,\varphi) \\ &= \sum\nolimits_{t=1}^{T} \log\left(\sum\nolimits_{m=1}^{M} \pi_m p_m\left(y_t\,|\,\varphi_m\right)\right)\end{aligned} \tag{3.7}$$

the task of ML is then to find a $\hat{\varphi}$ that can satisfy $\hat{\varphi} = \arg\max_{\varphi} \ell(y\,|\,\varphi)$, or equivalently, solve the function $\partial\ell(\varphi\,|\,y)/\partial\varphi = 0$. The asymptotic covariance matrix of this estimator is defined as the inverse of the observed Fisher information matrix, $I^{-1}(y\,|\,\varphi)$, where $I(y\,|\,\varphi) = -\partial\ell^2(y\,|\,\varphi)/\partial\varphi\partial\varphi^T\,|_{\varphi=\hat{\varphi}}$, and $\hat{\varphi}$ is considered as a valid estimator (MLE) if it can globally maximize equation (3.7). Here, since $\hat{\varphi}$ generally does not have an explicit solution, maximization step is usually performed by adopting an iterative procedure such as Newton-Raphson algorithm. Besides, when some parameter restrictions also need to be imposed, a non-linear sequential routine is then required to augment the process (see *FSQP* algorithm of Lawrence and Tits, 2001, for example).

As for the consistency, efficiency and asymptotical normality of the target estimator, Wald (1949) confirmed all these properties in his research. However, it is also important to note several exceptions here. For example, in some cases we may find $\ell(y\,|\,\varphi)$ is unbounded over the parameter space, thus it is impossible to find a single global maximum for equation (3.7). Hence one may need to look for a new local maximum that can also satisfy the same regularity conditions. However, the difficulty remains if multiple local maximums are coexisting. In this case, a proper selection among alternatives then could become another difficult task since additive separability of parameters has already been destroyed in a standard mixture models. To generate such an estimator, a proper initial value obtained from the prior investigations or graphical analysis on the training data is then indispensable.

Concerning the empirical evidence of applying ML to fitting mixture distribution, Rao (1948) made the first attempt to use Fisher method of scoring to estimate a two-component normal mixture with equal variances. Later, his iterative procedure was studied in Hasselblad (1966),

[23] Likelihood function $L(y\,|\,\varphi)$ in equation (3.1) is obtained as the probability density of observed data. It is considered as a function of model parameters φ. Since log-likelihood function $\ell(y\,|\,\varphi)$ is a monotonic function of $L(y\,|\,\varphi)$, the estimator $\hat{\varphi}$ that maximizes the $\ell(y\,|\,\varphi)$ thus will be the same as the one that maximizes $L(y\,|\,\varphi)$.

Day (1969) and Wolfe (1970), where an explicit formula for parameter estimators is derived and applications are extended to mixture of all variants in the exponential distribution family. Recently, Lindsay and Roeder (1992) and Bohning (1999) discussed the use of a non-parametric extension of ML to estimate mixture distribution. Although it is now confirmed that these ML methods are all utilizing a more efficient estimation procedure than early methods, it was not until the seminal paper of Dempster *et al.* (1977) that using classical inferential techniques was really stimulated.[24] Not only is the iterative scheme of ML formalized in a more general context, their EM algorithm also helps establish the convergence of MLE on a theoretical basis.

3.4.2.2 Expectation Maximization algorithm (EM)

For EM algorithm, in the literature there are a substantial amount of works dedicated to this topic. As a generic method for computing MLE based on the incomplete information set, this iterative method has been applied in a variety of statistical problems such as solving mixture distribution, variance component estimation and factor analysis. Since its contribution to classical statistics is substantial, we present in the following a detailed description of its estimation procedure along with some illustrations of its advantages and drawbacks in applications.

The EM algorithm is a technique strongly rooted in the missing information principle introduced by Orchard and Woodbury (1972) and subsequently developed by Beale and Little (1975). Its basic idea is to exploit the reduced complexity of ML after data augmentation. That is, by augmenting the current observable data with a hidden space, computation of MLE is then expected to be much easier for the new 'complete information'. Generally, the observed data in EM algorithm is called 'incomplete data' and the augmented part of these data is referred to as the 'missing data'. Here, note that these 'missing data' are not always missing in the real world, most of the time it is just a convenient technical device.

Once the complete information set is formed, the algorithm then works iteratively by alternating between E-step (Expectation) and M-step (Maximization). Formally, let y and z denote the observed information and missing data respectively and $\varphi^{(h)}$ be the current state of parameter; since log-likelihood function of complete data (y, z) given by

$$\log L_c(\varphi) = \sum \log p(y, z \mid \varphi) \tag{3.8}$$

[24] Early investigations of mixture distribution only use univariate sample data. However, with the advent of EM algorithm, such investigations are now also available in multivariate context.

is now unobservable, this method then solves the incomplete-data likelihood $L(\varphi \mid y)$ and obtains parameter estimates of the next state by replacing $\log L_c(\varphi)$ with its conditional expectation given y and current fit for φ. The E- and M- step in this procedure is given by

E- step: Calculate the conditional expectation of an auxiliary Q-function

$$\begin{aligned} Q(\varphi \mid \varphi^{(h)}) &= E\left[\log L_c(\varphi) \mid y, \varphi^{(h)}\right] \\ &\propto \log p(\varphi) + E\left[\log p(y,z \mid \varphi) \mid y, \varphi^{(h)}\right] \\ &= \log p(\varphi) + \int p(z \mid y, \varphi^{(h)}) \log p(y,z \mid \varphi) dy \end{aligned} \tag{3.9}$$

M-step: Update the parameter set to $\varphi^{(h+1)}$ by maximizing $Q(\varphi \mid \varphi^{(h)})$

$$\varphi^{(h+1)} = \arg\max_{\varphi} Q(\varphi \mid \varphi^{(h)}) \tag{3.10}$$

To ensure the monotonicity of this algorithm, it is often required in M step $Q(\varphi^{(h+1)} \mid \varphi^{(h)}) \geq Q(\varphi \mid \varphi^{(h)})$. Then, the whole iterations can proceed until the convergence is suggested by certain stopping criterion, i.e. $\| \varphi^{(h+1)} - \varphi^{(h)} \| \leq \varepsilon$ where $\varepsilon > 0$.

Here, before proceeding, it is necessary to note several advantages of the EM. Apart from its ease of implementation and numerical stability that are frequently documented in textbooks, another important aspect of this algorithm is that it can be used to input the missing values (obtained in the E- step). Besides, under regularity conditions, the global convergence of this algorithm can be ensured. However, using this inferential method can also generate drawbacks. For example, unlike ML, using the EM algorithm cannot provide an estimator for observed Fisher information matrix as a by-product in maximization step. Thus, we cannot generate an automatic estimate of standard error for $\hat{\varphi}$. Besides, in some cases EM may converge very slowly due to the lack of an analytical solution in either E- steps or M- steps. In this case, it is then preferred to use a simulation-based approach to enhance the algorithm. For example, Tanner and Wei (1990) introduced the Monte Carlo EM (MCEM) algorithm. Nielsen (2000) suggested using the stochastic EM (SEM).

As for its implementations, many classical papers and textbooks have illustrated an example (see Titterington *et al.* 1985 and McLachlan and Krishnan, 1997). Here, to obtain a practical view, we give the details (algorithm and codes) of how to use EM to estimate an *M*-component standard Gaussian mixture in Appendix IV. For more applications of this algorithm, see also Rachev and SenGupta (1993) for using GEM, a variant of EM, to estimate a hybrid mixture of Laplace and Weibull distribution, and Liu (1997), McLachlan and Peel (1998) and Lee *et al.* (2004) for using ECM to calculate the inference of a multivariate *t* mixture.

3.4.3 Simulation-based Bayesian approach

Above, we briefly illustrate two classical-inference based estimation techniques. Now, we describe how to use another rational method, a Bayesian approach, to learn unobservable parameters φ in a mixture model. The main aim of Bayesian inference is to simulate a series of random draws from a sampling kernel that corresponds to φ so that the true parameter value can be approximated using empirical summaries of these simulated values after initial draws are discarded. For financial models, since φ is often a parameter set containing many different elements, a high-dimensional integration technique is usually required for sampling purpose. Besides, in some cases, since these kernels may not have an analytical form, numerical approximation is also needed. Due to these difficulties, in early days using this method to estimate a sophisticated model was then found very difficult. However, after 1990 situation improved a lot, benefiting from the fast development of MCMC algorithms and the advent of modern high-speed computers. Recently, numerous researchers have successfully opened new interest in this inferential method and it is frequently applied in countless researches to estimate mixture models.

In order to obtain a brief idea of the development history of this method, it is necessary to note several monographs that have made the crucial contributions to its build-up and extensions. The origin of the Bayesian inference can be traced back to Thomas Bayes's essay, published in 1763. Initial development of this method was far from easy and its theoretical foundation was continuously challenged by numerous frequentists. For example, the founder of likelihood inference, Fisher, was particularly hostile to the use of Bayesian methods and often critical. In the middle of the last century, in response to the obvious deficiencies in classical inference, scholars such as Jeffreys (1961), Good (1950), and Lindley (1961) opened new interest in Bayesian methods. Unfortunately, the solutions provided by these authors, although good, could not be used to solve mathematical forms that were analytically intractable. To resolve this difficulty, a new revolutionary simulating technique, Markov chain Monte Carlo (MCMC), was then created. Starting from the fundamental work of Metropolis *et al,* (1953) and Hastings (1970), MCMC algorithm, since its introduction, has attracted a lot of interests and obtained massive empirical potentials. Important works concerning this sampling technique include Geman and Geman (1984), Gelfand and Smith (1990), Gilks *et al.* (1996), Robert and Casella (1999) and Carlin and Louis (2000).

Since the main purpose of this thesis is to calculate the Bayesian inference, in the following, several typical simulation techniques for conducting this inference are described in detail. Here, we divide these techniques into two categories. First, for those which can only be used to generate *i.i.d* draws, they are described in the traditional Monte Carlo framework and illustration is provided in the following sub-sections. Second, if resultant draws are forming Markov chains, algorithms are categorized into MCMC framework and we describe them in the next chapter. Before proceeding, some preliminary issues about the Bayesian inference are illustrated.

3.4.3.1 Preliminary issues on Bayesian inference

a. Bayesian vs. Frequentist

To understand the Bayesian method, it is always necessary to start with its difference from classically inferential approaches. Theoretically, there are many ways in which we can highlight these differences. For example, the probability statement in these two inferential paradigms is interpreted differently. From a frequentist's point of view, probability is regarded as an objective measure, a limiting relative frequency that represents the long-run behavior of a non-deterministic outcome.[25] However, according to Bayesian statisticians, it is then considered as a subjective quantity which heavily depends on the researcher who is assessing it. For example, while calculating the Bayesian inference, one always needs to assume a prior distribution for the parameter of interest before posterior simulation can be performed. Besides, difference between Bayesian and non-Bayesian can also be addressed by how they interpret the nature of parameters. In classic inference, parameter of a model is considered as a fixed, deterministic quantity. However, from a Bayesian's viewpoint, this unobservable quantity then becomes a variable. Although some possible values may still be suggested, usually a probability distribution will be associated to encode the uncertainty of parameters.

b. Advantages and assumptions

According to the illustration above, one may have already noted an important advantage of Bayesian methods over its competitor. That is, the parameter uncertainty is allowed. Unlike ML, Bayesian inference can use empirical summaries of a series of random draws to approximate the statistical characteristics of the true parameter value. Since the posterior result is depicted in a distributional form, more inferential information can be incorporated compared to that generated by using classical inference, where only a point estimator is often derived. Although it is not guaranteed that this distributional information can always make a substantial

[25] Laplace (1814) proposed the earliest version of this definition. Later, Neyman and Pearson formalized his idea and introduced extension. Although their interpretations provide an intuitively simple way to think of probability, to obtain an estimate of it, imposition of an assumption is necessary. That is, one can generate an infinite series of trials, replications, or experiments on the event of interest using the same search design. However, practically, as Kendall (1949) and Placket (1966, p26) put it, "…*Frequently, it is however not possible to obtain a large number of outcomes from exactly the same event-generating systems*…"

contribution, it does become valuable when non-standard posterior densities are observed. For example, when the posterior density presents significant asymmetry or multi-modality, we cannot then rely on the classical inferential method to recognize the risk of parameter uncertainty, which may cause serious underestimation and overestimation of the forecasts calculated from the model. Besides, since all scientific models are proposed according to the modeller's own understanding of the 'truth', Bayesians' paradigm provides the most overt presentation of model assumption because its probability statement is also based on a subjective measure (assuming a prior distribution). For a more detailed account of these advantages, see also Berger (1986), Efron (1986) and Gill (2002).

To conduct this inference, here it is worth noting some assumptions. First, when posterior sampling is performed, it is required that sampling kernels are all parametric functions. Although the analysis of non-parametric Bayesian modeling is also growing rapidly nowadays, we only review and apply the likelihood-based Bayesian method in this thesis. Second, since unknown parameters are all treated as having distributional qualities rather than being fixed, it is assumed that we can specify a proper prior distribution for these parameters. In case choosing a prior density is difficult due to the lack of relevant information, a distribution showing equal weighting is then used. Finally, sample data are assumed to be locally independent.

3.4.3.2 Posterior Updating Scheme

Now, we start to illustrate the details of posterior sampling scheme. First, consider a model with observations y distributed according to a parametric probability density $p(y|\varphi)$ where φ denotes the parameter set of interest.[26] Since the goal of inference is now to derive a probability statement of $p(\varphi|y)$ by exploiting the information in $p(y|\varphi)$, it is then required we apply the Bayes Theorem as an information processor so that

$$p(\varphi \mid y) = \frac{p(\varphi, y)}{p(y)} \tag{3.12}$$

Here, since the marginal density $p(y)$ can be retrieved by integrating out φ from the joint density of $p(\varphi)$ and $p(y|\varphi)$, that is $p(y) = \int p(\varphi)p(y \mid \varphi)d\varphi$, (3.12) can be rewritten as

[26] In some literature, model specification which was cast in the form of a conditional argument on a probability distribution can also be written as $p\,(y|\,\varphi, H)$ where H denotes modeller's background state of the information, which encompasses all hypotheses and existing knowledge before collecting data. Since this additional conditioning on H is required throughout the Bayesian theorem, to ease the expression, we omit its presence in the notations.

$$p(\varphi \mid y) = \frac{p(\varphi)p(y \mid \varphi)}{\int p(\varphi)p(y \mid \varphi)d\varphi} \tag{3.13}$$

In the above equation, note that $p\ (y)$ does not depend on φ suggesting that it provides no relevant inferential information about the likely value of φ. Thus, by terming this quantity a normalizing constant and eliminating it, we derive a compact and succinct form of $p\ (\varphi \mid y)$[27]

$$p(\varphi \mid y) \propto p(\varphi)p(y \mid \varphi) \tag{3.14}$$

where $p(\varphi)$ is a probabilistic form of prior information assumed by modellers on φ, called prior density. $p(\varphi \mid y)$ is called posterior density because the updated information is derived only after all training data have been learned.

According to (3.14) it is now clear that the posterior density is proportional to the unnormalized post-data inference. If either $p(\varphi)$ or $p(y|\varphi)$ is widely dispersed relative to the other, it will then have less of an impact on the final probability statement. This natural weighting scheme reflects the relative levels of uncertainty in these two densities. Empirically, since $p(\varphi)$ only encompasses the modeller's subjective knowledge, it is the $p(y|\varphi)$ that frequently plays the critical role in determining the shape of posterior density. The influence of this function on the posterior information becomes greater as the number of new observations increases. That is because the more observations involved in updating, the less influence exerted by our own conviction $p(\varphi)$.

Since, in Bayesian statistics, the likelihood function $L(\varphi \mid y)$ and $p(y|\varphi)$ is interchangeable, that is $L(\varphi \mid y) \equiv p(y|\varphi)$, $p(\varphi \mid y)$ can also be interpreted as a quantity jointly determined by the prior density and likelihood function. Thus, the equation (3.14) can also be rewritten as,[28]

$$\text{Posterior density} \propto \text{Prior density} \times \text{Likelihood function} \tag{3.15}$$

a. Prior distribution

Above, we have given two factors that simultaneously determine the posterior sampling scheme. Now, in this subsection, we describe the importance of assuming a proper prior density. Here, $p(\varphi)$ is termed as a prior density because its distributional form is given before each sample is incorporated to updating. If sufficient prior knowledge is available, this density can be defined on a very small domain with a parametric form. However, in most cases, only

[27] $\propto$ here denotes 'proportional to'

[28] Under mild conditions, Gelman *et al.* (1995) proved that the posterior density derived from (4.15) can convey 'more precise and sharper' information than the modeller's prior knowledge on φ.

very limited information is obtainable at the early stages of estimation. Thus, we often need to rely on a vague probabilistic statement for $p(\varphi)$. For example, we can let the priors be uniformly distributed. Briefly, it is a non-informative density giving equal or nearly equal weight to all possible values in target space Θ. Besides, we can also use reference prior, diffuse prior and many others for the same purpose.[29] It is necessary to note that, once such priors are assumed, its density value is usually a constant which can be eliminated in the posterior density. Thus, all that relates to the posterior information is only the likelihood function, and the result of Bayesian inference is very close to those generated by applying classically inferential techniques.

Meanwhile, there are also other things that need to be noted when specifying a proper prior. In equation (3.14)**,** we illustrated an example of generating posterior result for a one-parameter model. However, it is common that parameter set of interest may contain multiple elements. In this case, joint prior density of φ is then often handled in a way that all its marginals are assumed to be independent of one another so that prior information of one parameter will not contribute the posterior updating process of another. And $p(\varphi)$ is simply the multiplication of all individual priors. On the left-hand side of equation (3.14), since the goal of inference is now to generate posterior draws for all elements in φ, only marginal density of $p(\varphi \mid y)$ will be analyzed. In the following, we describe how to derive this joint posterior density and evaluate these marginal densities.

b. Posterior simulation

In a multi-parameter model, since the posterior $p(\varphi \mid y)$ is a joint density, sampling kernels to be evaluated are then marginals that correspond to each element in φ. Theoretically, if the state space is finite, these marginal densities can be assessed by integrating out all elements other than that of the interest from the joint density. For example, if φ_1 is the one of interest, its marginal can be defined as

$$\begin{aligned} p(\varphi_1 \mid y) &= \int p(\varphi \mid y) d\varphi_{-1} \\ &\propto \int p(\varphi) p(y \mid \varphi) d\varphi_{-1} \end{aligned} \tag{3.16}$$

[29] Reference prior is proposed in Bernardo (1979). Diffuse prior is suggested by a symmetric distribution with a very large variance. Berger (1985) discerned the location parameter from the precision parameter and presented 10 different ways to propose prior for hyper-parameters. For example, for a standard normal distribution, he suggested that a ζ^{-1} shaped prior for the precision parameter ζ, which is the inverse of scale parameter σ in N(μ , σ) is appropriate.

where φ_{-1} indexed all parameters except φ_1 in the whole set of φ. Meanwhile, we can also use

$$\begin{aligned} p(\varphi_1 \mid y) &= \int p(\varphi_1, \varphi \mid y) d\varphi \\ &= \int p(\varphi \mid y) p(\varphi_1 \mid \varphi, y) d\varphi \end{aligned} \tag{3.17}$$

to obtain the same result. The virtue of the second method is to apply numerical integration directly to the whole parameter set φ. However, since the first is more closely related to the MCMC algorithms to be illustrated in the next chapter, we apply it to depict the posterior sampling scheme.

Given the above equations, now it may seem very straightforward that the numerical integration is actually a plausible method to evaluate the posterior density once each marginal is properly defined. However, in practice using this technique is not only difficult but also costly. This is because target posteriors are often given non-analytically. Besides, the integrals included in them are most of the time defined as high-dimensional (since the models now have more than one parameter). Thus, even if an integration solution is proposed, it is often problem-specific. For example, Woznikowaski (1991) developed an analytical method to calculate the high-dimensional integration. Since the technique he introduced requires the target function to be drawn from a particular distribution, his method is then not suitable for the general Bayesian learning.

To circumvent this difficulty, a feasible way is to evaluate the posterior by applying a simulation technique to a sampling kernel that corresponds to the target density so that a series of random draws, whose limiting distribution approximates the density of interest, can be generated. Take the updating process suggested in (3.16) for example. The task of evaluating $p(\varphi_1 \mid y)$ now can be translated to simulating a series of random draws of φ_1 so that their stationary distribution can approximate $p(\varphi_1 \mid y)$. And it can be performed by firstly drawing a random sample, say m^{th} value of φ, from

$$\varphi^{(m)} \sim p(\varphi \mid y) \tag{3.18}$$

and then inserting $\varphi^{(m)}$ to (3.17) to obtain a $\varphi_1^{(m)}$ which follows

$$\varphi_1^{(m)} \sim p(\varphi_1 \mid \varphi^{(m)}, y) \tag{3.19}$$

Geweke (1989) argued that one can use the importance sampling technique of Hammersley and Handscomb (1964) with a standard optimisation method to generate a random sample for (3.18); however, a proper tuning is usually required when this approach is adopted. In the following year, Gelfand and Smith (1990) applied an image reconstruction technique suggested in Geman and Geman (1984) to perform the same task. They generate $\varphi^{(m)}$ by using the kernel updated by parameter values of last state, so that this new value is drawn from

$$\varphi^{(m)} \sim p(\varphi \mid \varphi^{(m-1)}, y)\,,\ where\ \ \varphi^{(m)} \xrightarrow{d} p(\varphi \mid y) \qquad (3.20)$$

Here, since the posterior results are approximated by a series of random draws and illustrated in a distributional form, compared to the classically inferential techniques, Bayesian inference are then able to present more informative results. In the following, we present a detailed illustration of how to use simulation technique to achieve this inferential task.

3.4.3.3 Monte Carlo Simulation Techniques

In this section, we describe several traditional Monte Carlo techniques of simulating *i.i.d* sequence of $\{\varphi^{(m)}\}$ whose density can approximate the posterior density of interest or just be $p(\varphi \mid y)$.

a. Direct sampling

For some kernels, since the inverse of their distribution functions (*c.d.f*) may have an explicitly parametric form, we can simulate a sequence of *i.i.d* samples for the target parameter by simply applying the *direct sampling technique*, as, for example, in the so-called *conjugate* situation where the posterior density is of the same distributional type as the prior density. Generating a random sample is easy because the sampling kernel to be evaluated now is only a modification of the prior density after all coefficients that characterize the conjugate class of probability distributions are updated (see Box and Cox, 1973, for a practical example using normal distribution, and Robert and Casella, 1999, for a general theory of conjugation for exponential distribution family). Although this method is mathematically convenient, situations like the conjugacy are extremely rare when empirical learning is performed. In all except several illustrative cases, posterior results usually cannot be generated analytically.

b. Acceptance and Rejection Sampling

From equation (3.15), it is known that posterior density is now jointly determined by two functions. Even if prior density $p(\varphi)$ is assumed to be uniformly distributed so that its density values can be absorbed in normalization constant, the chance of posterior density being complicated by a non-trivial likelihood function is still very high. In common situations where an analytical sampling kernel cannot be found, one then has to rely on a more sophisticated simulator to generate new updates for φ.

Here, a typical solution is to apply an *acceptance and rejection sampling* (ARS) technique. This method is initially attributed to three pioneers in the simulation area, Von Neumann, Metropolis and Ulam. Its basic idea, which is not difficult to conceptualize, is to simulate *i.i.d* samples from a source density $p(s)$ that is similar to the target density, rather than from $p(\varphi \mid y)$ itself. Here, note that the sense in which this source density is similar to the posterior is crucial. Depending on the efficiency of resultant simulators, ARS usually can be divided into acceptance sampling technique and the importance sampling technique.

b.1 Acceptance Sampling

First, regarding the acceptance sampling, we depict its sampling process using an example. Suppose we now let $\kappa(\varphi \mid y) = c_I p(\varphi \mid y)$ be a sampling kernel of posterior density, and $\kappa(\varphi \mid s) = c_s p(\varphi \mid s)$ be a sampling kernel of source density $p(s)$; if the bound of these two kernels r now satisfies the condition $r = \sup_{\theta \in \Theta} \kappa(\varphi \mid y) / \kappa(\varphi \mid s) < \infty$, the m^{th} draw of φ is then generated by applying the following pseudocode.

Acceptance sampling

```
1. Draw $u$ from a uniform distribution [0, 1]
2. Draw a candidate value $\varphi^*$ from $p(\varphi \mid s)$
3. If $u > \kappa(\varphi^* \mid y) / r \cdot \kappa(\varphi^* \mid s)$, go to step 1
4. Otherwise, $\varphi^{(m)} = \varphi^*$
```

Here, if the source density is correctly specified, one can prove that the samples drawn from the source density will always show the same distributional characteristics as those generated by sampling from posterior density, and the efficiency of this simulator is determined by the frequency of acceptance (See Geweke, 2000).

However, there is a difficulty; in most cases it is very hard to find such a good source density. Although, in some very special cases, it is certainly possible that we can find a $p(s)$ that can perfectly match the posterior density (hence r=1 and step 3 in the above loop can be omitted since all new draws now will be accepted with a fixed probability of one and no rejection will occur), such cases are very rare in empirical learning. For example, when a non-trivial likelihood function is incorporated to posterior density, finding an appropriate source density for $p(\varphi \mid y)$ is then usually a very difficult task (correlation mixture models proposed in this research are good examples of this).

b.2 Importance sampling

As have been briefly inferred from the above illustration, if the *acceptance and rejection* rule is applied to determine the appropriateness of a new random draw, in all except the ideally efficient sampling, $\{\varphi^{(m)}\}$ is always a fraction of all simulated values generated from the source density. Since the rejection is statistically unavoidable, efficiency of the ARS simulator then sometimes could arise as a concern. For example, in some cases the domain from which new draws are simulated could be much more disperser than that of interest, thus, it may take ARS a longer-than-usual time to finally locate a candidate draw which can be accepted. To improve this efficiency, another technique that also burgeons into standard procedure of ARS is then often used.

By placing more emphasis on the 'important' regions where posterior density is concentrated, Hammersley and Handscomb (1964) proposed a so-called *importance-sampling* technique where simulation is performed in the most relevant areas. Briefly, its basic idea is to incorporate a time-varying weighting scheme to the simulation process. By allowing the ratio of posterior density to source density as a function of candidate values, this simulator differs from acceptance sampling in that the fixed bound r is now replaced by a variable $\varpi(\varphi) = \kappa(\varphi \mid y) / \kappa(\varphi \mid s)$. Unlike the boundary condition imposed before, we no longer have to obtain the exact value of this bound, but just need to make sure $\varpi(\varphi) < \infty$. From this perspective, it is then very easy to note an advantage of this simulator. Since only the existence of an upper bound for $\varpi(\varphi)$ needs to be verified, finding a proper source density for importance sampling is then much easier.

Besides, under this approach, using empirical summary of the candidate draws to approximate the true parameter values is also very easy. As noted, in acceptance sampling, since the ratio of posterior density to source density is deterministic (a fixed value r), we use the accepted samples directly to make this approximation. However, when importance sampling is applied, these samples need to be adjusted by $\varpi(\varphi)$ before being input. For example, if M candidate values for φ have been generated, mean and variance of true parameter value are then approximated by

$$\mu(\hat{\varphi}) = E\left[p(\varphi \mid y)\right] = \frac{\sum_{i=1}^{M} \varphi^{(m)} \varpi(\varphi^{(m)})}{\sum_{i=1}^{M} \varpi(\varphi^{(m)})} \qquad \Sigma(\hat{\varphi}) = Var\left[p(\varphi \mid y)\right] = \frac{\sum_{i=1}^{M} [(\varphi^{(m)} - \mu)\varpi(\varphi^{(m)})]^2}{\sum_{i=1}^{M} \varpi(\varphi^{(m)})} \qquad (3.21)$$

Here, it should be noted that, although these candidate values are now used to approximate the distributional characteristics, they do not constitute a random sample from the real posterior density. This is because ultimately these values are simulated from source density.

As for the sampling efficiency, clearly, importance simulator now provides a statistically better solution than acceptance simulator because all candidate values are drawn from the most relevant area. However, as one of the ARS, it still has the same drawbacks as the others. For example, simulation result obtained from using this technique is very sensitive to the source density chosen to approximate the posterior. To propose a reliable source density, although various criteria have been already discussed, simple methods like moment matching, Laplace approximations, mixtures, and re-parameterisation are all found insufficiently flexible to accommodate the general problems. For example, when the target kernel has a complicated form, finding a good source density for it is then often considered as an impossible task. In such cases, researchers usually are inclined to try several different distributions for $p(s)$ until an optimal solution is found. Here, if a poor choice is made, the immediate cost is a very low acceptance rate. This happens because only a few candidates will be drawn directly from the high probability region (high mass).[30] Compared to the others, the weights $\varpi(\varphi)$ of these points are often much higher. Thus, the accepted samples for $p(\varphi \mid y)$ may be just reduced to these points. Since the difficulty of finding a proper $p(s)$ is massive in Bayesian statistics, a proper tuning is usually required when this technique is used.[31]

As for its implementations, Kloek and Van Dijk (1978) made the first attempt to use importance simulator to calculate Bayesian inference. A more extensive treatment of this technique with proofs was provided in Geweke (1989). Recently, several variants of this sampler are also proposed in the literature. For example, Evans (1991) introduced a so-called ‘adaptive importance sampling’ technique. Dagum *et al.* (1995) introduced the stopping rule theorem and Neal (1996) proposed the annealed importance sampling. For a more detailed summary and overview of these simulators, see Gelman *et al* (1995), Tanner (1996) and Robert and Casella (1999)

c. Hybrid sampling

In the last subsection, we presented the advantages and drawbacks of two ARS simulators. Acceptance sampling is easy to apply but inefficient to perform; importance sampler is a more efficient simulator whilst its implementation requires the calculation of a weighting function. Since the only difference between these two techniques is their formation of weighting scheme, a hybrid approach that yields the relative advantages of both then can be developed.

[30] In Metropolis *et al* (1953), this area refers to the places where high probability of acceptance is concentrated.

[31] Tuning here is the attempt to try different distributions for source density $p(s)$. Usually, adopting this strategy will increase the computational cost of obtaining inferential results.

To illustrate this approach, consider now that the existence of a theoretical bound for $\kappa(\varphi \mid y)/\kappa(\varphi \mid s)$ has be proved whilst its exact value has not yet been determined, to identify when to use importance simulator and when to use acceptance simulator, an arbitrary bound for the weighting scheme needs to be assumed in the first place. Say, if this bound is now given a finite value *b*, one then perform acceptance simulator to generate new draws whenever $\sup_{\theta\in\Theta} \kappa(\varphi \mid y)/\kappa(\varphi \mid s) \leq \max(b,1)$ is satisfied (this bound is now defined as either *b* or one). However, if the random draw φ^* satisfies $\kappa(\varphi^* \mid y)/\kappa(\varphi^* \mid s) \in [b,+\infty)$, importance simulator is then applied.

Since an analytic characterization of posterior density in general Bayesian learning is very difficult to find, even with a hybrid approach implementation of ARS algorithms may still encounter various difficulties. For example, the major problem, as has been illustrated already, is to find a proper source density that could closely approximate the posterior. Sometimes, even if such a density is given, simulation of new draws might still be trapped in a tiny region of probability space. That is, most of the new points are drawn from a small area whose volume is a tiny fraction of the whole. In this case, we would then need a simulator which can direct the searching of random samples to the most relevant areas as well as can be performed very efficiently. In particular, the MCMC algorithm to be described in the next chapter is exactly such a technique. Not only is the source density no longer required, this type of simulator can also be applied to tackle the problem of non-analytical kernel.

3.8 Summary

In this chapter, we start by illustrating some stylized features presented in the financial time series and then describe several ways to tackle them. Among these features, we concentrate on the non-Gaussian characteristics such as heavy tails and leptokurtoses and point out using mixture distribution is an ideal solution to accommodate them. Since building mixture model is a main aim of this research, we illustrate the probabilistic properties, development history, mixing strategies and implementational issues of this type of model and give two examples of it. Besides, we also describe several techniques that can be used estimate them. Specifically, for the classical inferential approach, emphasis is put onto the maximum likelihood and EM algorithm. For Bayesian inference, an introductory illustration of its aim, sampling process and estimation procedure is provided. However, concerning the details of its simulator, description is given in the next chapter.

Chapter 4

Literature review (part three)
-Markov Chain Monte Carlo (MCMC) algorithm

Introduction

In this chapter, we describe the *Markov Chain Monte Carlo* (MCMC) algorithms. As a naïve method for performing stochastic simulation, this technique provides a rational solution to calculating the Bayesian inference by leading the search of candidate values for each parameter to a high probability region in an efficient manner. In the following sections, we provide a comprehensive overview of the aim and sampling process of this technique and discuss several issues concerning its implementations. Specifically, emphases are put onto two of the most widely used simulators. One is Metropolis-Hasting algorithm of Hasting (1970). The other is Gibbs sampler of Geman and Geman (1984). Since the task of this thesis is to use a variant of standard Gibbs sampler to estimate correlation mixture models, for this particular simulator we illustrate its settings and sampling procedure in details. Besides, several diagnostic tests for examining the convergence for resultant draws are also reviewed.

4.1 Development history and Markov Chains

First, it is beneficial to briefly review the development history of MCMC algorithms. The origin of this technique is attributed to Metropolis (1953) who laid the foundation of using a sequence of dependent points to investigate the equilibrium properties of large systems of particles (e.g. molecules in a gas). Later, Hastings (1970) generalized his method to propose the famous Metropolis-Hasting algorithm. Through these studies, although the bridging relationship between stochastic simulation and inference calculation was found, it was not until Geman and Geman’s (1984) and Gelfand and Smith’s (1990) work that implementational potentials of MCMC were fully recognized in the Bayesian context. This is because an important solution for alleviating the computational burden for Bayesian inference is finally raised. Since then, countless researches are dedicated to developing this algorithm and a lot of variants are proposed in the literature. Among them, key works include tutorial papers by Casella and George (1992) and Chib and Greenberg (1996), a monograph by Tanner (1996) and a long survey by Gelman and Rubin (1992), Geyer (1992) and Besag *et al.* (1995).

Here, before proceeding, it is important to note a major advantage of this technique. That is, MCMC can provide a more flexible solution than other methods to deal with the general Bayesian problems. As has been illustrated in the last chapter, if one is to use a standard Monte Carlo simulator such as direct sampling or ARS to compute the inference, it is required that we can find either an explicit solution for sampling posterior density or a proper source density which can closely approximate it. As a comparison, the goal of MCMC is, however, to construct on state space, Θ, a Markov chain for the parameter of interest, say, $\{\varphi^{(m)}\}$, so that its density can converge to the posterior $p(\varphi \mid y)$ after an initial transient period is discarded. For this particular algorithm, since the kernel to be evaluated no longer needs to be analytical, sampling random draws becomes easier.

Now, since all simulated values are going to form Markov chains, it is important to understand some properties of this particular stochastic process before we proceed further. In Appendix V, a detailed illustration of this issue has been provided. However, here only one thing needs to be re-emphasised. That is the convergence theorem “*under regularity conditions any Markov ergodic chain will converge to a stationary distribution after a sufficiently long run.*” Given this theorem, it then explains why draws, even if not appearing to be *i.i.d* but only showing Markovian properties, can still be used to approximate the distributional characteristics of statistical inference.[32] In the following, we use Metropolis Hasting algorithm and Gibbs

[32] The aim of using MCMC to calculate the Bayesian inference is to construct an ergodic Markov chain for

sampler as examples to illustrate the simulation process of MCMC and the emphasis is put onto a variant of the later technique, namely the Griddy Gibbs sampler.

4.2 Metropolis Hastings algorithm (MH)

Metropolis Hastings algorithm is an important MCMC technique. Although this simulator is not to be implemented in this research, we illustrate its aim, sampling process and variants here due to its similar importance in statistics to the Gibbs sampler. This algorithm is initially described in Hastings (1970) as a generalization of standard Metropolis algorithm. Its main purpose is to simulate a sequence of dependent realizations whose stationary distribution can be used to approximate the posterior density. More precisely, given the current state $\varphi^{(m-1)}$, it generates a Markov chain with the next state $\varphi^{(m)}$ chosen by considering a small change to $\varphi^{(m-1)}$ and accepting or rejecting this change based on the comparison result of a probability statement.

4.2.1 Sampling Process

To illustrate its sampling process in more detail, we now consider an example. If the posterior density is denoted by $p(\varphi \mid y)$ and current state is $\varphi^{(m-1)}$, to use MH algorithm to generate a new draw φ^* for $\varphi^{(m)}$, first we give an arbitrary jumping density (or proposal function) $q(\varphi^* \mid \varphi^{(m-1)}, y)$ and simulate a value, say φ^*, for $\varphi^{(m-1)}$ to jump to. Then, a transition kernel that determines whether to accept or reject this new candidate value is defined

$$p(\varphi^* \mid \varphi^{(m-1)}, y) = q(\varphi^* \mid \varphi^{(m-1)}, y) a(\varphi^* \mid \varphi^{(m-1)}) \tag{4.1}$$

so that random feature of $\varphi^{(m)}$ is jointly determined by jumping density and acceptance probability $a(\varphi^* \mid \varphi^{(m-1)})$. Here, $a(\varphi^* \mid \varphi^{(m-1)})$ is a probability statement determining whether $\varphi^{(m)}$ will jump to the new candidate value or remain at the current state. If the transition kernel makes a move from $\varphi^{(m-1)}$ to φ^* more likely than from φ^* to $\varphi^{(m-1)}$, that is $a(\varphi^* \mid \varphi^{(m-1)}) > a(\varphi^{(m-1)} \mid \varphi^*)$, MH algorithm will accept the new candidate φ^*. Otherwise, $\varphi^{(m)}$ will just be equal to the current state $\varphi^{(m-1)}$.

Given this criterion, now it is necessary to formalize a proper function for evaluating $a(\varphi^* \mid \varphi^{(m-1)})$. Concerning this task, first we rewrite the equation (4.1) to

sampling kernel of a parameter so that the stationary distribution of this parameter can approximate the posterior density of interest.

$$a(\varphi^* \mid \varphi^{(m-1)}) = \frac{p(\varphi^* \mid \varphi^{(m-1)}, y)}{q(\varphi^* \mid \varphi^{(m-1)}, y)} \tag{4.2}$$

and then apply the reversibility property (see Appendix V) of Markov chain to resolve $p(\varphi^* \mid \varphi^{(m-1)}, y)$ and derive another form of $a(\varphi^* \mid \varphi^{(m-1)})$ in (4.4). That is

$$p(\varphi^{(m-1)} \mid y) p(\varphi^* \mid \varphi^{(m-1)}, y) = p(\varphi^* \mid y) p(\varphi^{(m-1)} \mid \varphi^*, y) \qquad \text{or}$$

$$p(\varphi^* \mid \varphi^{(m-1)}, y) = \frac{p(\varphi^* \mid y) p(\varphi^{(m-1)} \mid \varphi^*, y)}{p(\varphi^{(m-1)} \mid y)} \tag{4.3}$$

$$a(\varphi^* \mid \varphi^{(m-1)}) = \frac{p(\varphi^* \mid y) p(\varphi^{(m-1)} \mid \varphi^*, y)}{p(\varphi^{(m-1)} \mid y) q(\varphi^* \mid \varphi^{(m-1)}, y)} \tag{4.4}$$

After considering a symmetric sample path and defining a reverse jump for (4.1), we rewrite the transition kernel from φ^* to $\varphi^{(m-1)}$ to

$$p(\varphi^{(m-1)} \mid \varphi^*, y) = q(\varphi^{(m-1)} \mid \varphi^*, y) a(\varphi^{(m-1)} \mid \varphi^*) \tag{4.5}$$

Now, by inserting (4.5) into (4.4), a new solution for $a(\varphi^* \mid \varphi^{(m-1)})$ can be derived

$$a(\varphi^* \mid \varphi^{(m-1)}) = \frac{p(\varphi^* \mid y) q(\varphi^{(m-1)} \mid \varphi^*, y) a(\varphi^{(m-1)} \mid \varphi^*)}{p(\varphi^{(m-1)} \mid y) q(\varphi^* \mid \varphi^{(m-1)}, y)} \tag{4.6}$$

After rearrangement, finally we get

$$D(\varphi^* \mid \varphi^{(m-1)}) = \frac{a(\varphi^* \mid \varphi^{(m-1)})}{a(\varphi^{(m-1)} \mid \varphi^*)} = \frac{p(\varphi^* \mid y) q(\varphi^{(m-1)} \mid \varphi^*, y)}{p(\varphi^{(m-1)} \mid y) q(\varphi^* \mid \varphi^{(m-1)}, y)} \tag{4.7}$$

$D(\cdot)$ here is a function for evaluating whether $a(\varphi^* \mid \varphi^{(m-1)}) > a(\varphi^{(m-1)} \mid \varphi^*)$. Once this value is obtained, we can use the result to determine the value of $\varphi^{(m)}$. To understand more clearly how this sampling process will work, we provide below its pseudocodes.

Metropolis Hastings algorithm

```
1. Draw $u$ from a uniform distribution $[0,\ 1]$
2. Draw $\varphi^*$ from $q(\varphi^* \mid \varphi^{(m-1)}, y)$
3. Calculate the acceptance probability $D(\varphi^* \mid \varphi^{(m-1)})$ for $\varphi^*$
       $D(\varphi^* \mid \varphi^{(m-1)}) = \min\left\{ \frac{p(\varphi^* \mid y) / q(\varphi^* \mid \varphi^{(m-1)}, y)}{p(\varphi^{(m-1)} \mid y) / q(\varphi^{(m-1)} \mid \varphi^*, y)}, 1 \right\}$
4. If $u \le D(\varphi^* \mid \varphi^{(m-1)})$, $\varphi^*$ is accepted, let $\varphi^{(m)} = \varphi^*$
5. Otherwise, $\varphi^*$ is rejected, $\varphi^{(m)} = \varphi^{(m-1)}$.
```

4.2.2 Variants of Metropolis Hastings algorithm

Above, we illustrated the sampling procedure of a standard Metropolis Hasting algorithm. In the literature, there are also many variants developed based on it. Since, for a Markov chain to be valid, it is only required that ergodicity condition for ensuring the convergence theorem be satisfied, variants of MH then can be easily proposed by replacing $q(\varphi^* \mid \varphi^{(m-1)})$. In the following, we describe four typical examples of these variants.

However, before proceeding, it is necessary to note a relationship between this density and convergence because $q(\varphi^* \mid \varphi^{(m-1)})$ now determines the (acceptance rate) efficiency of searching to be performed in the high probability region. Generally, it is desirable that the acceptance rate of a MH is set as high as possible. Thus, to generate the parameter value of next state, we do not need to simulate too many new draws and then reject them. However, Tanner (1996) described a situation where even a chain with a close-to-one acceptance rate may still converge very slowly. This is because the distance moved between new draws is very short. Thus, it may take the chain a fairly long time to forget its origin. From these illustrations, it is not difficult to see that convergence of a Markov chain is actually an empirical issue.

a. Metropolis algorithm

Now, we illustrate one of the simplest MH variants. That is the Metropolis algorithm of Metropolis *et al.* (1953). For this simulator, the authors replaced the reversibility condition assumed for jumping density $q(\varphi^* \mid \varphi^{(m-1)}, y)$ in equation (4.1) with a fixed symmetric function so that the transition $\varphi^{(m-1)} \to \varphi^*$ and its reverse $\varphi^* \to \varphi^{(m-1)}$ have the same probability $q(\varphi^* \mid \varphi^{(m-1)}, y) = q(\varphi^{(m-1)} \mid \varphi^*, y)$ and the acceptance probability $D(\cdot)$ is set to be

$$D(\varphi^* \mid \varphi^{(m-1)}) = \min\left\{ \frac{p(\varphi^* \mid y)}{p(\varphi^{(m-1)} \mid y)}, 1 \right\} \tag{4.8}$$

Here, it is obvious that, after assuming this symmetric function, Metropolis algorithm now becomes a limiting case of standard MH. In all cases except when significant asymmetry is observed in target density, Gleman *et al.* (1995) proved that the convergence induced by an ergodic Markov chain will always occur for a symmetric transition function as if the homogeneity of sampling process is kept changed. Thus, the posterior result generated by using this simulator, if the convergence of algorithm can be confirmed, is always valid (see Brooks and Robert, 1998, for proofs).

b. Independence Metropolis chain

Tierney (1994) proposed another MH variant, called Independence Metropolis chain. He let $q(\varphi^* \mid \varphi^{(m-1)}, y) = q(\varphi^*)$ so that sampling a new candidate φ^* from the jumping density

$q(\varphi^* \mid \varphi^{(m-1)}, y)$ is independent of the current state $\varphi^{(m-1)}$, and the acceptance probability $D(\cdot)$ is modified to

$$D(\varphi^* \mid \varphi^{(m-1)}) = \min\left\{ \frac{w\left(\varphi^*\right)}{w(\varphi^{(m-1)})}, 1 \right\} \tag{4.9}$$

where $w(\varphi) = p(\varphi \mid y) / p(\varphi)$.

Here, since the simulated samples are forming *i.i.d* sequence and acceptance and rejection of a new draw is determined by a probability statement, this technique is closely related to the ARS algorithms described in last chapter. However, note that their interpretations of the decision rules for φ^* are slightly different. For example, if a rejection occurs, ARS algorithm explains it by the simulator now placing low weight on a draw that is unlikely to be relevant to the density of interest. However, when Independence Metropolis chain is used, this rejection is then interpreted as the sampler assigning a very low probability of accepting φ^* as the new draw for $\varphi^{(m)}$. As for the flexibility of the algorithm, Independence Metropolis chain is usually considered the easiest MH algorithm to perform. However, sometimes its convergence rate could be extremely low.

c. Random walk Metropolis chain

Apart from the above two samplers, a more frequently used MH variant is the Random walk Metropolis chain. With this simulator, each φ^* is now drawn from a jumping density defined to be $q(\varphi^* \mid \varphi^{(m-1)}, y) = q(\varphi^{(m-1)} - \varphi^*)$ whose domain is close to the current state $\varphi^{(m-1)}$, and the search for new candidates is performed without any preference concerning the direction. Empirically, researchers usually let this density be hyperspherically multinomially distributed so that $q(\varphi^* \mid \varphi^{(m-1)}, y) = N(\varphi^{(m-1)}, s \cdot I_k)$ where I_k is a diagonal identical matrix; s is an adaptive factor used to maintain an acceptable jump. This is because, given this setting, the new candidate draws φ^* will be automatically locating around $\varphi^{(m-1)}$ and the probability of accepting new draws will decrease along with the span of exploration. Besides, to induce no directional preference while searching the parameter space, this density is frequently set to be mutually exclusive so that every direction of the movements can generate the same probability.[33]

[33] We assume independent multi-normal distribution here.

d. Forced walk algorithm

In spite of the Random walk chain, many researchers have also documented the use of another MH variant where directional preference can be included. This sampling technique is called Forced walk algorithm. Its jumping density is set to be multivariate Gaussian $\phi(\varphi^{(m-1)}, s \cdot V)$ where V denotes the observed covariance matrix. Since the directional preference of simulation can now be obtained from the density values of updated Gaussian, a new candidate draw φ^* for $\varphi^{(m)}$ can be simulated once the preferences originating from $\varphi^{(m-1)}$ are all calculated and averaged.

To use this simulator, it is important to note that a proper tuning for V is usually indispensable. If this covariance matrix is set too large, the jumping density could be too dispersive relative to the density of interest. Thus, more candidate draws need to be simulated to obtain one accepted sample since the probability of rejection will dramatically increase. Conversely, if V is set too small, the distance moved between different draws will probably become very short. Therefore, a much larger number of iterations are required to cover the whole parameter space and the convergence of the chain may become every slow. In practice, usually we can tune this algorithm by firstly running a series of sub-runs to increase the speed of convergence, and then periodically updating V according to the previous result so that the next simulation can adapt to the 'successful' searching direction (See Robert, 1996, for illustration of an example).

4.3 Gibbs Sampler

Apart from the MH algorithm, another popular MCMC technique that is also frequently used to simulate Markov chain is the Gibbs sampler of Geman and Geman (1984). This method was initially applied in statistical physics to analyze Gibbs distribution on lattices for image reconstruction. In 1990, Gelfand and Smith successfully demonstrated a much larger scope of potential for its uses in inference calculation. Since then, new interests has been continuously generated to develop this simulator. For example, Gibbs sampler combined with the data augmentation technique of Tanner and Wong (1987) has been proved very successful in treating latent variables in econometrics. As remarked by Geman and Geman (1984, p24), '…*this sampling method provides a much simpler way of drawing from a multivariate probability density based on the densities of parameter subsets conditional on all other parameters and data...*' In the following, we present a detailed illustration of this simulator's sampling process and several of its typical variants. Besides, some initial settings concerning its implementation are also briefly discussed. For a more comprehensive review, see Casella and George (1992). For a good survey, see Smith and Roberts (1993), Tanner (1996), Gilks *et al.* (1996) and Robert and Casella (1999).

Here, we first define the aim of this technique. Gibbs sampler, by its definition, is proposed to perform the high-dimensional stochastic simulation. Its basic idea, which is not difficult to conceptualised, is that if it is possible to partition the parameter set into several blocks and specify sampling kernel of each parameter as a density function conditioned on all other parameters, then, by cycling through these low-dimensional conditional statements, we can eventually reach the true joint distribution of interest (Gill, 2002). Note that although for Gibbs sampler the posterior updating may now involve multiple simulations, its conditional densities usually correspond to only one parameter each. Thus, the simulation task is simply to sample a series of dependent draws for a set of one-dimensional densities. Even if, in some special cases, we might be able to define a sampling kernel encompassing several different parameters, it is generally assumed that these parameters are highly correlated and their joint conditional density has an analytical form. Thus, as far as the computational cost is concerned, Gibbs sampler is then usually considered as a much cheaper solution for performing high-dimensional simulation than numerical integration. Besides, its advantage of conceptual simplicity and the ease of implementation are also quite obvious.

Now, it is worth noting an important assumption for performing this algorithm. Since the transition kernel in Gibbs sampler is formed by a set of conditional densities, to facilitate the simulation process, it is usually assumed that the probability statements of these conditional densities are articulated enough so that it is possible to draw *i.i.d* values directly from these densities. Although this assumption, as has been mentioned repeatedly, is too strong for general problems, and only in some illustrative cases may one find analytical sampling kernel for parameters in a financial model, the real contribution of Gibbs sampler is not constrained by this at all. This is because, even if there are several densities which are analytically intractable, dimensionalities of these densities are usually quite low; thus, numerical integration techniques which do not need much computational expense could still be used for sampling. It is the idea of reducing the dimensionality of the density to be simulated that really popularizes the application of Gibbs sampler. To see how the joint posterior density of Bayesian inference can be uniquely defined using a series of unidimensional distribution, we provide, in the following subsection, an example.

Suppose φ now denotes the parameter set of interest and can be partitioned into K blocks, that is $\varphi = \left\{\varphi_1, \varphi_2, \cdots, \varphi_K\right\}$. Here, we let $\varphi_{<(k)} = \left\{\varphi_1, \varphi_2, \cdots, \varphi_{k-1}\right\}$, $\varphi_{>(k)} = \left\{\varphi_{k+1}, \cdots, \varphi_K\right\}$ and $\varphi_{-(k)}$

be the parameter vector φ without the element *k,* $\varphi_{-(k)} = \left\{ \varphi_{<(k)}, \varphi_{>(k)} \right\}$. Meanwhile, we also define the posterior density of φ_k conditioned on recent values of all other parameters to be $p_k(\varphi_{(k)} \mid \varphi_{-(k)})$ and assume there exists an analytical sampling kernel $q_k(\varphi_{(k)} \mid \varphi_{-(k)})$ which specifically corresponds to it. Given these settings, task of Gibbs sampler is then to simulate from this sampling kernel. Note that φ_k here can be either uni- or multi- dimensional. If we only consider a single parameter *k,* $p_k(\varphi_{(k)} \mid \varphi_{-(k)})$ is called full conditional distribution, or just full conditional. Since, in Bayesian statistics, prior density of different parameters are generally assumed to be independent, this full conditional can be easily obtained after all parameters that do not relate to φ_k are absorbed in the joint posterior density of *φ*. For example, if we now consider a two-parameter model whose joint prior density is given by $p(\varphi_1, \varphi_2)$=$p(\varphi_1)p(\varphi_2)$, to define $p_1(\varphi_{(1)} \mid \varphi_{-(1)})$ we only need to eliminate all elements that do not depend on φ_1 in joint posterior density $p(\varphi_1, \varphi_2 \mid y)$ and absorb them in the normalization constant.

4.3.1 Sampling process of standard Gibbs sampler

Now, we illustrate the sampling process of standard Gibbs sampler. Consider the same posterior density $p(\varphi \mid y)$ (the stationary distribution to be approximated) as before. Our task is now to produce a Markov chain for each element in φ that can move toward this density after cycling through all full conditionals. Provided that the current state is $\varphi^{(m-1)} = (\varphi_1^{(m-1)}, \varphi_2^{(m-1)}, ..., \varphi_K^{(m-1)})$ and $q_k(\varphi_{(k)} \mid \varphi_{-(k)})$ is a simulating kernel of $p_k(\varphi_{(k)} \mid \varphi_{-(k)})$ for φ_k, to generate the next state of the chain $\varphi^{(m)}$, we proceed as follows:

Gibbs Sampler Algorithm

1. Draw $\varphi_1^{(m)}$ from $q_1(\varphi_1 \mid \varphi_2^{(m-1)}, \varphi_3^{(m-1)}, ..., \varphi_K^{(m-1)})$
2. Draw $\varphi_2^{(m)}$ from $q_2(\varphi_2 \mid \varphi_1^{(m)}, \varphi_3^{(m-1)}, ..., \varphi_K^{(m-1)})$
3. Draw $\varphi_3^{(m)}$ from $q_3(\varphi_3 \mid \varphi_1^{(m)}, \varphi_2^{(m)}, \varphi_4^{(m-1)} ..., \varphi_K^{(m-1)})$
4. $\vdots$
5. Draw $\varphi_K^{(m)}$ from $q_K(\varphi_K \mid \varphi_1^{(m)}, \varphi_2^{(m)}, ..., \varphi_{K-1}^{(m)}, \varphi_K^{(m-1)})$

Given the above procedure, now we can easily confirm that this simulator is indeed producing Markov chains that will converge to the posterior distribution. This is because all necessary conditions required in the convergence theorem are satisfied. For example, simulation of the next state is now only conditioned on the values of the current state. The sampling process is kept homogeneous with all consecutive probabilities independent of the current length of the chain. Liu, Wong and Kong (1991a, b) and Schervish and Carlin (1990) presented the regularity conditions under which a Gibbs sampler will converge. A more general case that leads to a geometric convergence rate is discussed in Roberts and Polson (1994).[34] Here, note that although we are updating only one parameter at each step, in practice it is applicable and desirable several parameters can be combined into the same group and updated together. As Roberts and Sahu (1999, p21) argued, '*...by blocking highly correlated parameters, the convergence rate of the Gibbs sampler might be improved...*' Besides, it is also worth mentioning that the Gibbs sampler is actually a special case of the aforementioned MH algorithm. To see this proof, consult Appendix VI for the details.

4.3.2 Variants of Gibbs sampler

In this section, we describe two variants of standard Gibbs procedure. One is Completion Gibbs sampler of Robert and Casella (1999). This other is Slice sampler of Higdon (1998). As for the Completion Gibbs sampler, if there is a function g that satisfies the condition $p(\varphi \mid y) = \int g(\varphi, z)dz$ and the full conditional distributions of $g(\varphi, z)$ are very easy to simulate, Robert and Casella (1999) suggested using this new function as the source for updating rather than simulating draws directly from the original posterior density.[35] Higdon (1998) proposed another generalization of standard Gibbs sampler by introducing some auxiliary uniformly distributed random variates. Suppose now the posterior density $p(\varphi \mid y)$ can be written as the multiplication of some positive functions, i.e. $p(\varphi \mid y) = \prod_{i=1}^{k} p_i(\varphi \mid y)$ and at $(m\text{-}1)^{th}$ iteration we can generate a uniform random variable u according to $u_i \sim U(0, p_i(\varphi^{(m-1)} \mid y))$. Then, for the next state, $\varphi^{(m)}$ is simulated from $\varphi^{(m)} \sim U(A^m)$, where $A^m = \{a; p_i(a \mid y) \geq u_i, i = 1,2...,k\}$. Here, since only uniformly distributed random draws will be simulated, this sampler is usually considered as a computationally very cheap way to perform the general Bayesian learning. For more details of how to implement this algorithm, see Damien *et al.* (1999).

4.3.3 Hybrid Gibbs-MH algorithm

Above, a necessary condition for implementing the Gibbs sampler is we can find either an analytical form for full conditional or full conditional itself is decomposable. However, empirically, a common situation is there could be one or more blocks of joint posterior density

[34] Converging at a geometric rate means variation distance moved between two samples drawn at consecutive time points decreases at a geometric rate.

[35] *z* here is an arbitrary variable.

not satisfying these assumptions; therefore, a more generalized simulator is needed to deal with these non-conjugate cases. In MCMC framework, Gilks and Wild (1992, 1993) proposed a solution called adaptive rejection method to simulate draws from a non-analytical log-concave density. Ritter and Tanner (1992) suggested using grid-based evaluation. Here, before we proceed to illustrate the Griddy Gibbs sampler, a hybrid approach that combines the MH algorithm and Gibbs sampler is described firstly because this method provides a naïve solution for sampling non-trivial densities.

As for this hybrid approach, it is usually called '*Metropolis within Gibbs method*'. Simply put, it is a simulator where MH algorithm is used to solve the non-conjugate blocks, whilst Gibbs sampler is used to evaluate the analytically tractable blocks. For example, if we now assume *b* is the only block in φ whose sampling kernel does not have an analytical expression, we then use MH algorithm to simulate a candidate value $\varphi_{(b)}^{*}$ for $\varphi_{(b)}^{(m)}$ from $\varphi_{(b)}^{*} \sim q(\varphi_{(b)}^{*} \mid \varphi_{<(b)}^{(m)}, \varphi_{>(b-1)}^{(m-1)})$ if the current state is $\varphi^{(m-1)}$ and use Gibbs sampler to simulate all remaining values in $\varphi_{-b}^{(m)}$. Here, for $\varphi_{(b)}^{*}$ its acceptance probability is computed by

$$D\left(\varphi_{(b)}^{*} \mid \varphi_{<(b)}^{(m)}, \varphi_{>(b-1)}^{(m-1)}\right) = \min\left(\frac{p(\varphi_{<(b)}^{(m)}, \varphi_{(b)}^{*}, \varphi_{>(b-1)}^{(m-1)}) / q(\varphi_{<(b)}^{(m)}, \varphi_{>(b-1)}^{(m-1)})}{p(\varphi_{<(b)}^{(m)}, \varphi_{>(b-1)}^{(m-1)}) / q(\varphi_{(b)}^{(m-1)} \mid \varphi_{<(b)}^{(m)}, \varphi_{(b)}^{*}, \phi_{>(b-1)}^{(m-1)})}, 1 \right) \tag{4.10}$$

and we decide whether to accept or reject this new draw after comparing this probability with a uniformly distributed random variate *U*. If $D\left(\varphi_{(b)}^{*} \mid \varphi_{<(b)}^{(m)}, \varphi_{>(b-1)}^{(m-1)}\right)$ is larger than *U*, then we say $\varphi_{(b)}^{*}$ is accepted. Otherwise, this draw will be rejected and $\varphi_{(b)}^{(m)}$ is traced back to the last state. That is $\varphi_{(b)}^{(m)} = \varphi_{(b)}^{(m-1)}$. To see the convergence result of this simulator, Geweke (2005) illustrated an example using a two-parameter model. For its application in finance, see Cappuccio, Lubian, Raggi (2004).

4.3.4 Griddy-Gibbs sampler

Not only using a hybrid approach, evaluation of a non-conjugate block can also be resolved by enhancing the standard Gibbs process with a Monte Carlo numerical integration technique. Since full conditionals reduced from joint posterior density are usually low-dimensional, using a deterministic integration rule to evaluate a non-analytical density over a grid of points is then economically feasible. This approach is initially proposed in Ritter and Tanner (1992) to estimate a non-linear regression model and a two-parameter Cox model. Briefly, its main aim is to approximate the *c.d.f* of a full conditional which is difficult to simulate by using a piecewise linear function; once the high mass is detected, a new random draw is then generated by inserting a uniformly distributed random variable to the inversion of that approximation. Since implementation of this algorithm is greedy on computational time, it is then usually called ‘Griddy Gibbs sampler’.

Following the standard sampling procedure described in Section 4.3.1, we now give an example for this simulator. Given that the current state of posterior simulation is $\varphi^{(m-1)} = \left(\varphi_1^{(m-1)}, \cdots \varphi_b^{(m-1)} ..., \varphi_K^{(m-1)}\right)$ and φ_b is a non-conjugate block whose simulating kernel $q_b(\varphi_b \mid \varphi_{(-b)}^{(m-1)})$ does not have an analytical expression, to use Griddy Gibbs sampler to generate a random draw for this block, we firstly select a grid of points $(\varphi_{b(1)}, \varphi_{b(2)}, \cdots, \varphi_{b(G)})$ for φ_b and then use the following steps to generate $\varphi_b^{(m)}$ of the next state.

Griddy Gibbs sampler

1. Insert the grid points $(\varphi_{b(1)}, \varphi_{b(2)}, \cdots, \varphi_{b(G)})$ to the sampling kernel $q_b(\varphi_b \mid \varphi_1^{(m-1)}, \cdots \varphi_{b-1}^{(m-1)}, \varphi_{b+1}^{(m-1)} ..., \varphi_K^{(m-1)})$ to calculate the density values of block *b*. That is $G_q = \left(q_{(1)}, q_{(2)}, \cdots, q_{(G)}\right)$.
2. Compute the *c.d.f* values of G_q by applying a deterministic integration rule to $\Phi_{(i)} = \int_{\varphi_{b(1)}}^{\varphi_{b(i)}} q_b(\varphi_b \mid \varphi_{(-b)}^{(m-1)}) d\varphi_b$ where $i = 2, \cdots G$ and derive $G_\Phi = \left(0, \Phi_{(2)}, \cdots, \Phi_{(G)}\right)$
3. Normalize G_Φ through the function $G_\Phi^{'} = \Phi_{(i)} / \Phi_{(G)}$ to make cumulative distribution values of $G_\Phi^{'}$ span over [0, 1].
4. Generate a uniformly distributed random variable $u \sim U[0,1]$ and insert it to the inversion of $G_\Phi^{'}\left(\varphi_b \mid \varphi_{(-b)}^{(m-1)}\right)$. And, after applying the numerical interpolation, we obtain a new draw for $\varphi_b^{(m)}$.

Above, if there are any blocks other than *b* which are also non-conjugate, we can adopt the same procedure to generate a new sample for their parameters. However, for those where an analytical sampling solution is obtainable, only direct sampling needs to be performed.

Since Griddy Gibbs sampler is now the only MCMC algorithm to be used for inference calculation in this thesis, a detailed illustration of several issues concerning its implementations is provided below. Concretely, we will discuss issues like how to choose a proper grid of points for parameter of interest, which integration and interpolation technique to use in simulation, and several advantages of this simulator.

4.3.4.1 Choice of grid points

First, concerning the selection of grid points, usually we can start by determining a theoretical bound for the parameter of interest and then draw either equally-spaced or variably-spaced points from the specified domain to form a grid. Here, for certain parameters, this boundary information can be obtained from restrictions imposed on them. For example, to ensure the covariance stationarity of a GARCH process, volatility persistence parameter β is often constrained to an interval [0, 1]. However, empirically, a more typical solution is to restrict the value of β to an even narrower space, say [0.5, 1]. This is because a large body of evidence has confirmed the strong volatility persistence for various finance time series, and in very few cases is estimated parameter value for β found less than 0.5. Since random draws in the low mass such as those in the range of [0, 0.5] now can be purposely avoided, an efficient search can be expected.

Once the upper and lower bounds are determined, the second step is to generate each point from the given interval. Generally, if no prior information is available, we can simply choose equally-spaced points from the selected domain to form a uniformly distributed prior. However, it is always preferable, either through some past experience or an expert's advice, to obtain some early knowledge of the posterior so that the grid can put more points (emphasis) on neighbourhoods of the high mass and fewer points near the low mass. By so doing, the efficiency of the algorithm can be improved a lot. However, unfortunately, in general Bayesian learning such prior knowledge is often not available. Hence, in countless cases it is still the equally-spaced points that are used the most.[36]

4.3.4.2 Integration rule and Interpolation technique

Apart from the selection of grid points, in the sampling process of the Griddy Gibbs sampler the deterministic integration rule applied in step 2 and the numerical interpolation technique used in step 4 are also two factors related to the posterior results. Usually, compared to the task of choosing grid points, it is much easier to choose these techniques because more objective election criteria can be adopted. For example, if one is asked to choose a series of good grid points, the decision is usually made subjectively. We might choose a large number of points to calculate the integral for a relatively simple function, but much fewer for a complicated one

[36] Theoretically, in Gibbs sampler it is also possible to use a variable grid. For example, when the simulation has just started, the performance of using grid point-based numerical integration to approximate a non-trivial density function could be quite poor; thus more points are needed to search the area where substantial volatility is present. However, when the sampler tends to be more stable and the approximation results improve, to obtain a random sample the number of points needed to be input for evaluation can be greatly reduced.

due to the computational cost concern. However, as for selecting a proper integration or interpolation technique, implications then can be easily obtained from the massive researches that have been performed in statistical literature.

First, concerning the numerical integration, Davis and Rabinowitz (1975) provided a detailed survey and comparison of various techniques. When Griddy Gibbs sampler is used, many authors suggested using a simple method such as trapezoidal rule to calculate the integral over a fixed grid of points. This is because using more complicated alternatives such as iterative Simpson algorithm, although good, is very likely to induce high computational cost. Since the full conditionals to be evaluated are already assumed to be complicated (or non-conjugate) functions, it is then desirable to use a relatively simple method so as to alleviate the overall computational cost.

Besides, this concern also applies when interpolation technique is chosen. That is to say, it is preferable to use simple linear function for interpolation, although high-order polynomials are also available for implementation. Clearly, to depict the relationship between adjacent points, using linear function is easier and cheaper. However, when convexity or concavity are present, quadratic functions are then probably a better choice since minor changes due to the second-order derivative can also be accounted. Here, although even more sophisticated techniques, such as splines analysis, for solving multidimensional interpolation are also possible, generally their implementations are not recommended for common empirical uses.

4.3.4.3 Advantages and implementation issues

Given the sampling process and technical settings illustrated above, it is now necessary to summarize some advantages of Griddy Gibbs sampler and illustrate why this simulator is preferable to other alternatives for solving general Bayesian problems. Apart from the conceptual simplicity which has been briefly discussed at the start of this subsection, one of the most important advantages of this grid-based simulator is its ease of implementation. As pointed out by Ritter and Tanner (1992, p172), "*…The Griddy Gibbs sampler in its simplest form generally can be implemented in only 30 to 50 lines of codes without including any subroutine that computes the posterior…*" Thus, for an experienced programmer, the main task is only to add an enhancement of density function to a highly modular form. Even if a very complicated full conditional is considered, the cost of coding will not increase substantially. This algorithm can be easily 'transplanted' to solve any statistical functions. However, when other simulators such as importance sampling or ARS are used, their codes are then often sample-specific and not re-useable which means that one has to rewrite the program all over again for each new application.

Besides that, another major advantage of using Griddy Gibbs sampler is it allows us to obtain a smooth estimation of marginal posterior density. Empirically, it means this algorithm can deal with a variety of statistical characteristics (or density shapes), e.g. skewness and high-peakedness. This is mainly because integration is now performed on a grid so that every direction in posterior density can be explored in detail. Moreover, it is also easy to incorporate a variance reduction technique into the sampling process so that the variance in estimation of moments of marginal posterior density can be reduced. This technique in MCMC is called 'conditioning'. For example, to estimate parameter *b*, we can use $\sum_{n=s+1}^{N} E[b \mid \varphi_{-b}^{(n)}, y, z^{(n+1)}]/(N-s)$ instead of $\sum_{n=s+1}^{N} b^{(n)}/(N-s)$ where s is the number of draws to be sampled for posterior to reach its equilibrium state.

Since the correlation mixture models to be proposed in the next chapter are going to assume a heteroskedastic (GARCH) specification, here it is also worth noting another implementation issue of this MCMC algorithm when it is implemented in a heterogeneous environment. As illustrated before, a necessary condition for performing Griddy Gibbs sampler is that a parameter set of interest φ is separable for each element. However, note that this condition is not satisfied in all cases. For example, Bauwens and Lubrano (1998) illustrated a case of a regression model whose innovation is modelled by GARCH-*t*. Say φ now consists of regression parameter λ and GARCH parameter θ: this model then can be specified as,

$$\begin{aligned} y_t &= x_t^{'}\lambda + u_t \\ u_t &\sim t_v(0, \Sigma_t) \end{aligned}$$

where u_t follows GARCH process. Here, since $u_t = (y_t - x_t^{'}\lambda)$, Σ_t is then a function of both λ and θ. And simulating GARCH parameter of the next state is not only determined by current information on θ but also by current information on λ. To illustrate it more clearly, posterior sampling kernel of $(\theta \mid \lambda)$ according to the sequential sampling procedure of Gibbs sampler is now

$$\kappa(\theta \mid \lambda) \sim f(h(\theta, \lambda), \theta) \tag{4.11}$$

Since $\kappa(\theta \mid \lambda)$ is no longer a sole function corresponding to θ (sampling kernel of θ is not explicitly related to its own), using Gibbs sampler is not appropriate here. This problem appears because regression parameter and GARCH parameters are now both presented in the same model. To circumvent this difficulty, we do not consider in our paper any regression term in the mean equation when correlation mixture model is proposed.

4.3.5 Data augmentation

As has been shown, the Gibbs sampler and its variants provide an easy way to resolve complex statistical inferences. However, their empirical potentials can be further developed if 'incomplete data theorem' is exploited. Tanner and Wong (1987, 1991) proposed a so-called *data augmentation* technique to provide such an improvement (see Carlin *et al.*1992, and Kim *et al.,* 1998, for examples). As a special case of Gibbs sampler but unlike Gibbs sampler, this MCMC algorithm provides a simple method to simulate unknown parameter values by augmenting the given information (observable data) with a series of latent variables and then iteratively improving the quality of these augmented quantities. From this aspect, it is clear that this technique is actually similar to the EM algorithm of Dempster *et al.* (1977). Both methods are based on the assumption of the existence of a complete information set. EM is valid when our task is to find a local maximum for model parameters. However, while the goal is to describe the complete posterior distribution, data augmentation then becomes a more appropriate resolution.

To illustrate the use of this simulator more clearly, consider now a typical state space model, a stochastic volatility model with observations *y*, unknown volatility *h* and the parameter set of interest φ. Suppose *h* and φ are now both unobservable and our task is to evaluate the posterior density $p\ (\varphi |\ y)$. To use data augmentation to calculate model inference, one first needs to define a predictive density $p\ (h|\ y)$ as an intermediate information processor. Usually, this density can be computed by integrating out latent variables from a joint density. If, for example, there exists a parameter set ϕ in φ that is related to the dynamic process of *h*, then after writing $p\ (h|\ y)$ into the following form

$$p(h\mid y)=\int_{\Phi}p(h\mid\phi,y)p(\phi\mid y)d\phi \tag{4.12}$$

a random draw of *h* can be simulated from the above predictive density. Since *h* and *y* now become observable, we can evaluate $p\ (\varphi |\ y)$ by just integrating out *h* from another joint density, that is,

$$p(\varphi\mid y)=\int_{h}p\left(\varphi\mid h,y\right)p\left(h\mid y\right)dh \tag{4.13}$$

where $p\left(\varphi\mid h,y\right)$ according to the Bayes theorem is proportional to $p(\varphi)p(y,h\mid\varphi)$, and an iterative algorithm for updating φ can then be constructed based on (4.13).

To obtain a more practical view, we provide the pseudo code of generating $\varphi^{(m)}$ using data augmentation technique and information at $\varphi^{(m-1)}$ in the following:

Data augmentation

1. Generate a set of N values of h from $p^{(m-1)}(\varphi \mid y)$
2. Update the parameter approximation using $p^{(m)}(\varphi \mid y) = \frac{1}{N} \sum_{n=1}^{N} p(\varphi \mid y, h_i)$
3. Simulate a value from $p^{(m)}(\varphi \mid y)$ for $\varphi^{(m)}$

Here, a major concern is how to chose *N*. Generally, the larger the *N* is, the better the approximation while the slower the convergence will be. Therefore, before each simulation starts, one always needs to make a proper choice of this number so that efficiency of the sampler and validity of posterior results can be soundly balanced. Besides, another feature worth noting here is that this sampling technique will reduce to standard Gibbs sampler if *N* is set to equal one.

4.4 Implementation issues of MCMC simulators

In practice, there are a lot of implementation issues concerning the use of MCMC algorithms. For example, before sampling starts one needs to choose a proper initial value for each parameter and a reasonable prior density. As new points are being drawn, appropriate tuning of the sampler is indispensable and one may also find it necessary to apply a variance reduction technique. After the sampling process has been iterated for a sufficiently long time, issues like whether the simulated chain has converged or how many more independent replications need to run could then also be raised. In the literature, the aforementioned issues have all been thoroughly studied. In the following, we only selectively discuss some of them that are related to the use of our Griddy Gibbs sampler. They include the selection of initial values, choosing burn-in period and some miscellaneous issues (See Neal, 1993, for a good survey of other implementational issues of MCMC).

4.4.1 Selection of Initial values

When an iterative method, say a Bayesian approach, is used to calculate the model inference, estimation always starts by selecting a proper initial value for each parameter. In the literature, a variety of techniques have been provided to perform this prior exploration of posterior distributions. Typical methods include the simulated tempering of Geyer and Thompson (1995), simulated annealing of Jennison (1993) and mode hunting of Gelman and Rubin (1992). All these methods can be utilized to suggest an appropriate initial value.

Generally, if prior information is available, that is, if we know roughly where the high mass will be located, initial values of a parameter can be easily chosen as just equal to the mode of prior density assumed. For example, we can, by either fitting an EM algorithm or a grid search, find this mode. Gelman and Rubin (1992) illustrated such an example. Besides, similar findings are also documented in Rubin and Wu (1997, p34) where the authors argued, *"…using EM algorithm and its variants is a wise step for solid computing involving the simulation of the posterior distribution. It gives a rough picture of the posterior distribution at a lower cost than the Gibbs sampler…."*

However, it is necessary to note that pre-estimation of sample data is not always easy. Using EM algorithm for example, to calculate its inference for a multi-dimensional problem, is often a time-consuming job. In such cases, to search for the initial values, one then probably has to rely on an arbitrary method. For instance, if a relevant domain for a parameter to be simulated can be determined, we can then select a random point in this domain as the starting value for φ. Although this method is theoretically feasible, using it could sometimes result in a very slow convergence rate because the algorithm may now be initialized by a point far from the high mass; thus it will take the chain longer than 'normal time' to finally converge. In practice, it would always be better if we could try different initial values for posterior sampling if the computational cost of generating multiple chains is not a major concern. This is because a diversified map can then be shown. Besides, one can also expect some valuable information about the non-convergence for the target chain, if these chains can be compared to each other.[37] Gelman and Rubin (1992) proposed a convergence diagnostic test based on this virtue. They argued the more dispersed the initial values are, the more sensible assessment the result will be. For a more detailed illustration of this issue, see Section 4.5.2.

4.4.2 Burn-in period

As just illustrated, since posterior sampling using MCMC technique is often initialized by some arbitrarily-selected starting points; realizations of Markov chain generated in the initial transition period will then unavoidably contain bias from these starting values and cannot be used as a valid sample from target distribution. Given this feature, it is then important to determine how long this transition period (or so-called burn-in period) would be because, even if the chain is now initialised by the mode of high mass, it may take it some time to forget its origin, and some further time to fully explore the posterior distribution.

[37] Note that the comparison result here can be used to assess the *non-convergence* but not the *convergence* of Markov chain. This is because, even if Markov chains are initialised by different starting values and congregate in the same region, all of the chains could only be seduced by the same local maxima and mix around in their own local region. Thus, it does not necessarily construct a representative for the stationary distribution.

Concerning this issue, although a variety of theoretical analyses have been performed, a general consensus has still not been reached. Usually, one still needs to rely on the result of a specific convergence diagnostic test to obtain some implications. For example, Raftery and Lewis (1996) proposed such a diagnostic test to address the issue of burn-in period. By setting up a new chain (not Markovian) parallel to the simulated chain, the authors accessed the convergence (number of iterations required for a chain to converge) in their research using quantile information. However, Robert and Casella (1999) argued that the unidimensional nature of the new chain assumed in their method does not account for the potential correlation between different primary chains; thus, Raftery and Lewis's test cannot be used to deal with the general Bayesian learning. To see more alternatives for accessing convergence, a detailed illustration is provided in Section 4.5.

4.4.3 Miscellaneous issues

Apart from the general issues just illustrated, here it is also worth noting some miscellaneous issues concerning MCMC. For example, using MCMC simulator is nothing but coding a program. Since a lot of conditions such as ergodicity need to be ensured when chains are generated, the programming codes are required to reflect these virtues. For example, to make the sampling sequence obtaining local property, the current state needs to be treated as the only input for simulation of the next state. To maintain the homogeneity, the posterior sampling process needs to be kept unchanged as the iterations proceed. Positiveness is guaranteed if any value in the parameter space can be randomly reached in later iterations whatever the initial values might be.

Given the above settings, although it may seem that using MCMC algorithm is now ready to produce accurately inferential results, different problems may still sometimes occur. For example, Gelman (1996) presented a discussion of these problems and listed three factors that are frequently overlooked. They are inappropriately specified model, error in programming Markov chains (stationary distribution of the chain may not be the target distribution) and a low convergence rate. Among them, the potential damage caused by the first factor is usually considered the most serious. This is because an incorrectly specified model may lead to an improper posterior density. Since a common result of this bias is that sampling kernels might not be integrable, even if a 'good-shaped' histogram is observed, the resulting posterior information would be unavoidably spurious. In such cases, the existence of a limiting distribution for parameter of interest then needs to be proved again. As for the slow convergence, early recognition of this problem for Markov chain is very difficult. Even for the most experienced researchers, anticipating problems like the chain getting stuck in a low mass for a long period of time is still a nearly impossible task.

4.5 Numerical accuracy and Convergence Diagnostic tests

In this section, we describe several diagnostic tests to be used in our later chapters for examining the convergence of simulated chains. For a parameter, if a sequence of its simulated values has been generated, an important thing to know is how well the empirical moments of these simulated samples can approximate the true parameter value. That is, we need to access whether the chain has converged and evaluate $\bar{\varphi}^{(m)} \to \bar{\varphi}$ when m, the length of chain, is sufficiently long.

In the literature, there are a lot of papers dedicated to proposing a valid test for examining the convergence and these tests are usually divided into two groups. One relies on the existence of an analytical kernel. The other however utilizes the output values from one or more replications of simulated chains. Since using the second method can provide a more problem-independent way to assess convergence, we illustrate several examples of it in the following subsections. However, before proceeding, it is important to note that the purpose of these tests is now not to detect the exact state from which a Markov chain starts to converge, but to find the evidence of failure of *non-convergence*.

4.5.1 Autocorrelation

Inspecting the sample path is one of the simplest ways to monitor the evolution of a Markov chain. To assess its convergence, we can rely on simple methods such as calculating its autocorrelation (or correlation) to see whether the target chain (or multiple chains) has converged. This statistic can tell how independent a simulated chain is in itself and of others. In the univariate context, usually the higher the autocorrelation, the slower the target chain would converge. Similarly, in the multivariate case, the higher the intercorrelation, the slower the multiple chains would be mixing. Although this test is now very easy to perform, frequently it needs to be used with other quantitative-based tests to explain the convergence result, because assessment is now made based solely on visual analysis.

4.5.2 Variance Ratio test

To propose a more objective criterion, Gelman and Rubin (1992) introduced a so-called *PSRF* test to examine whether subsamples generated from different starting values of a long chain are stemming from the same limiting distribution. To test the convergence, the authors used several independent shorter chains simulated from the same limiting distribution to replace the original long chain. They argued that, by so doing, a variety of information concerning the convergence can then be obtained. For instance, we can use the result of this test to discover how well the chain is mixing, to what extent the output from individual chains is indistinguishable and, most importantly, the sensitivity of posterior inference to different initial values.

As for its virtue, this test is now similar to performing an ANOVA test and its convergence can be assessed inferentially. Now, suppose that we have simulated *m* independent sequences of length *2n* for φ that begin with different starting values. To perform the test, first we calculate a quantity called variance between m sequence means using

$$V = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{\varphi}_i^{\cdot} - \bar{\varphi}_{\cdot}^{\cdot}) \quad (\mathit{Between\ Sequence\ Variance}) \tag{4.14}$$

$$where \qquad \bar{\varphi}_i^{\cdot} = \frac{1}{n}\sum_{t=n+1}^{2n}\varphi_i^t \ , \quad \bar{\varphi}_{\cdot}^{\cdot} = \frac{1}{m}\sum_{i=1}^{m}\bar{\varphi}_i^{\cdot} \tag{4.15}$$

where φ_i^t is the t^{th} realization of the chain generated from using i^{th} set of starting value for φ. Then, using a similar approach, we calculate the mean of *m* within-sequence variances s_i^2 according to

$$M = \frac{1}{m}\sum_{i=1}^{m}s_i^2 \quad (\mathit{Within\ Sequence\ Variance}) \tag{4.16}$$

$$where \qquad s_i^2 = \frac{1}{n-1}\sum_{t=n+1}^{2n}(\theta_i^t - \bar{\theta}_i^{\cdot}) \tag{4.17}$$

Once these two estimates are obtained, we now compare the between-sequence variance *V* with the within-sequence variance *M* through an approximating *t*-distribution with mean $\hat{\mu}$, variance

$$\hat{V} = \frac{n-1}{n}M + (1+\frac{1}{m})V \tag{4.18}$$

and degree of freedom

$$\hat{d} = \frac{2\hat{V}^2}{Var(\hat{V})} \tag{4.19}$$

and compute the potential scale reduction factor (PSRF) using

$$PSRF = \frac{\hat{d}+3}{\hat{d}+1}(\hat{V}/M) \tag{4.20}$$

Here, if *PSRF* is very large, it is suggested that the sampling sequence has not yet fully converged with the stationary distribution and the variance of simulated values can be further reduced. However, Gelman and Rubin (1992) argued that a value of $\sqrt{PSRF} \leq 1.2$ is often enough to claim the convergence.

Recently, based on the empirical interval lengths, Brooks and Gelman (1997) have developed an alternative variance ratio test. For each chain to be diagnosed, first they take the empirical $100(1-\alpha)\%$ interval (the $100(\alpha/2)\%$ and $100(1-\alpha/2)\%$ points) of n simulated draws to form m within-sequence interval length estimates. Then, from the entire set of observations that are obtained from all chains, they recalculate the empirical 100(1-α)% interval and generate a total-sequence interval length estimate. Finally, the new interval-based *PSRF* statistic is computed by

$$IPSRF = \frac{length\ \ of\ \ total-sequence\ \ \mathrm{int}erval}{average\ \ length\ \ of\ \ within-sequence\ \ \mathrm{int}erval} \tag{4.21}$$

Here, note that one of the main advantages of the above *PSRF*-type tests is their ease of implementation. Indeed, when sampling kernel is not too complex and the simulation can proceed without much grid evaluation, using these tests can help monitor the convergence of Markov chains periodically. Once a satisfying value is observed, the last *n* simulated values can then be treated as draws directly from the density of interest. However, when the sampling kernel becomes complicated, their implementations may then quickly become very burdensome since multiple chains now need to be simulated concurrently to assess the convergence.

4.5.3 Partial means test

Besides, since the Markov chain, once converged, is a stationary time series, we can also use some time series techniques such as spectral analysis to assess the chain's convergence. Geweke (1992) developed a so-called *Z* test based on this virtue. By exploiting the fact that the means of two subsamples of a stationary time series are the same, the author proposed a difference of means test on two subsamples which are respectively collected from some early era of the chain and some non-overlapping late era of the chain.

For instance, consider now the sampling sequence $\{\varphi^{(m)}; m = 1, 2, \cdots N\}$ with two subsamples: one is $\{\varphi_A = \varphi^{(m)}; m = 1 \cdots n_A\}$ and the other is $\{\varphi_B = \varphi^{(m)}; m = n_B \cdots N\}$ where $1 < n_A < n_B < N$. Geweke tested the null hypothesis of equal means by calculating

$$Z = \frac{(\overline{\varphi}_A - \overline{\varphi}_B)}{\sqrt{\overline{S}_{\varphi}^{A}(0)/n_A + \overline{S}_{\varphi}^{B}(0)/(N-n_B)}} \xrightarrow{d} \phi(0,1) \tag{4.22}$$

where $\overline{\varphi}_A$, $\overline{\varphi}_B$ are sample means of $\{\varphi_A\}$ and $\{\varphi_B\}$ and $\overline{S}_{\varphi}^{A}(0)$, $\overline{S}_{\varphi}^{B}(0)$ are their respective spectral density estimates (see Chatfield, 1996 for details of calculating spectral density function in a given window). Since this statistic is now to be asymptotically Gaussian distributed, values that are atypical of standard Gaussian are then interpreted as evidence for showing non-convergence.

4.6 Summary

As a major Bayesian method, Markov Chain Monte Carlo (MCMC) is the main topic illustrated in this chapter. Since using this technique can help a researcher to acquire inferential information on models even having very sophisticated specifications, it is then frequently applied in a variety of financial studies. Here, we provide a comprehensive overview of the aim and sampling process of this technique and illustrate two examples of it. One is the Metropolis Hasting algorithm. The other is Gibbs sampler. The emphasis is put onto a variant of the latter method, namely the Gribby Gibbs sampler. This simulator is important because from a Bayesian's perspective it can successfully reduce a multivariate simulation task to a series of multiple uni-dimensional jobs and sampling a sophisticated log-likelihood function becomes feasible. Besides, in this chapter we also describe some implementional issues concerning the MCMC. We answer questions like "how to choose a proper burning-period, initial value and prior density for each algorithm?" and "what are the factors that ultimately will be related to the convergence of simulated chains and how to access this convergence using different statistical tests". As for the second question, a detailed review with evidence is provided in Section 4.5.

Chapter 5

Correlation forecasting comparison in currency market

- A revisit of information efficiency derived from option market

Abstract

From this chapter, we begin to examine the performance of various correlation models using empirical data. Here, we use the foreign exchange market as an example to compare the forecasting performance of eleven existing models and special attention is paid to the implied correlation model whose forecast of the future calculation is generated from option prices rather than through a time-series tool. Since, in both theoretical and empirical aspects, an option contract is referred to as a derivative product which can convey forward-looking information through the embedded market expectation, we exploit this invaluable information source to utilize implied volatility collected from the OTC market to calculate implied correlation of two currency trios and compare the results to forecasts generated from a variety of competing models. After a series of comparison of the forecasting performance, our findings suggest there is no evidence of a consistently best performer in our forecasting pool. The relative accuracy of the generated forecasts in approximating realized correlation is very sensitive to the measures used to evaluate them. Therefore, we conclude that the correlation forecasting performance of these competing models is actually an empirical issue.

5.1 Introduction

Just like volatility, correlation is also a major input that needs to be accurately forecasted in finance. Recall that in Section 1.2 we have mentioned a variety of economic contents of this statistic and its importance in the daily financial applications such as asset allocation, risk management and derivative pricing. However, to understand this statistic properly, researchers went through a long path. For example, in the early days this coefficient was often considered as a static quantity in financial modelling. Its value is regarded as time-invariant if the sample period of interest is kept unchanged. However, after 1980s, benefiting from the gradual recognition of time-varying characteristics for volatility, financial researchers started to use a similar sampling process to model correlation. During that period, a typical method to capture the dynamic correlation is to through generalizing a univariate volatility model to a multivariate version so that time-varying characteristics of correlation can be obtain through an intermediate. However, it is necessary to note that not all methods developed based on this virtue can provide a dynamic out-of-sample forecast. For instance, correlations generated from using historical correlation model and EWMA are then often criticized as backward-looking because they assume that the future market will present exactly the same pattern as before. Although the implementation of these models is easy, no empirical fitting is required in their covariance generating process so that the calculated correlation is not actually dynamic.

Multivariate GARCH provides a solution to this problem. Bollerslev *et al.* (1988), by generalizing a univariate GARCH model to multivariate context, provided a typical example of using historical information to obtain correlation forecast. The VECH model, proposed in their paper, laid the foundation for calculating time-varying correlation through a multivariate conditional heteroskedastic framework. It assumed the covariance matrix follows the autoregressive process (see also BEKK model of Engle and Ng, 1995) so that this matrix, as a whole, could be modelled as a function of its own lagged term and past innovations. However, note that, although correlation is now allowed to be time-varying, estimation cost of this model and its variants usually rises at an exponential rate with the dimensionality (general VECH and BEKK specifications have a large parameter vector to estimate even for a bivariate case). Thus, the immediate cost of this implementation is that their empirical potentials are limited to a very narrow space and often can only be used to solve a system of very small size.

Besides, in the heteroskedastic framework conditional correlation can also be derived by decomposing the evolving process of covariance matrix into separate parts. Bollerslev (1990) suggested using different dynamic processes to model the individual volatility of each time series in a portfolio and the correlation among them. For example, in his study volatility is modelled by a series of independent univariate GARCH models whilst correlation is set to be constant. Meanwhile, a more flexible framework can be built if the correlation evolving process itself is allowed to be time-varying so that randomness in covariance matrix can be jointly determined by randomness in the volatility part and randomness in the correlation part. Recently, Engle (2002) provided such a refinement. Through his DCC model, the author used another independent GARCH to model correlation dynamics. Several even more generalized cases are also proposed in the literature. For instance, Sheppard (2002) extended Engle's work by introducing an asymmetric variable to the correlation evolving process; Pelletier (2004) made a contribution by incorporating a three-state regime switching model to Gaussian DCC.

As can be easily noticed, models mentioned above are unanimously utilizing a time series tool to calculate future correlation. However, this predication can also be made from a mechanism using option prices as information processor. Its result, compared to others, is usually considered capable of possessing a more naive view of how the future market will move because a direct mapping between option price and market-embedded expectation is given. Usually, to calculate this implied correlation, a triangular relationship between assets of interest needs to be identified first. Then, all three correlating assets are required to have option contracts specifically traded on them so that implied volatility data can be obtained. Although this model has the advantage of ease of implementation, its drawback is also clear in that implied correlation cannot be easily generalized to any future time unless some extrapolation techniques are used. This is because, for a given contract used to calculate this correlation, its maturity is now fixed. Thus, only when the forecast horizon of interest is set equal to this outstanding maturity will prediction generated from this model be considered theoretically valid. Besides, since correlation is now calculated using implied volatility as inputs, careful interpretation of the result, especially from a theoretical aspect, is needed because, except in some illustrative cases, financial asset returns usually will not be Gaussian-distributed either individually or jointly; thus, the condition for validating this market-embedded information itself is unsatisfied.

So far, nothing has been said about the forecasting performance of aforementioned correlation models. Now, to understand this issue, it is beneficial to start from the similar illustration of volatility models because volatility and correlation are two latent variables modeled, in most cases, using similar mechanisms. Thus, the forecast result of one has important implications for the other. Specifically, according to numerous literature contributing to the modeling of volatility process, a general consensus has still not yet been reached on a single model that can provide consistently the best forecast of the future realized volatility. Implied volatility, which is frequently confirmed as a conditionally biased estimator, outperforms other historical information-based forecasts in many empirical results (see Christensen and Prabhala, 1998, Fleming, 1998 and Blair, Poon and Taylor, 2001 for evidence of equity index option and Mayhew and Stivers, 2003, for evidence of individual stock options). However, its performance is not consistent all the time. For example, Kroner, Kneafsey and Claessens (1995) and Amin and Ng (1997) argued that forecasts generated from GARCH models may contain valuable information not presented in implied volatilities. Since it is very difficult to find a single best, many researchers turn to employing a combination of both historical and option information source to generate forecast. As confirmed by countless evidence, implementation of this strategy can provide a much-improved performance for volatility forecasting. Analogously, similar findings are also confirmed when time-varying correlation is predicted.

As mentioned earlier, since the empirical analysis of using implied correlation is very rare in financial literature, probably due to the difficulties of finding three triangularly related assets, the main aim of this chapter is then to fill the gap by extending the early works of Camp and Chang (1997) and Walter and Lopez (2000) to re-address the issue of correlation forecasting in foreign exchange market. Specifically, we calculate the realized correlation based on Anderson *et al.,* (2000) and compare the predictive accuracy and information contribution of eleven competing models.

Here, it is important to note three complements included in this research as contributions to the existing literature. First, to examine forecasting performance, we choose a variety of currency pairs, EUR/USD/GBP (or EU/US/UK) and EUR/USD/JPY (or EU/US/JP) for analysis due to the massive liquidity presented in their respective trading markets. For example, according to BIS's 2004 Triennial Central Bank Survey, currency pairs deviating from above trios altogether have the deepest spot and OTC market in the world. Trading volume of US/EU, US/UK and US/JP accounts for nearly 60% of daily volume in the global foreign exchange market (see Appendix VII). Meanwhile, implied volatility data used for calculating implied correlation is collected from a leading index that incorporates overall market expectation rather than from a single market participant, as in Camp and Chang (1997). Therefore, it is reasonable to expect that forecasts generated from such historical data would be theoretically more informative and efficient in terms of the incremental information they could contribute. Second, in this research a broad forecasting group with a total of eleven correlation models are used to predict future realized correlation. It includes implied correlation model, historical correlation model (with price history respectively set at 7, 22 and 65), EWMA, simplified univariate GARCH model (with Gaussian error and GED error), VECH, BEKK, CCC, and DCC. A combination of these forecasts comprises most of the market expectations extractable from spot and option market. Finally, in addition to comparing correlation forecasting performance among different currency pairs, we also launch cross-horizon forecasting performance comparison in this paper. Such investigation can be used to address issues like 'whether the forecasting performance of a specific model in the short run will possess a similar pattern when it is used in the long run'. If the answer is 'no', the resulting implication is then important for risk managers who tend to use the same correlation models to hedge risks on different maturities. For this reason, we calculate multi-horizon forecasts for each model in this chapter. And, concretely, forecasting horizons of interest are set to be one week, one month and three months respectively. Here, note that analysis of the first two horizons is essential for practical daily risk management. According to the Basle Committee on Banking Supervision rules (see Basle Committee on Banking Supervision, 1998, 2004), in order for investors to have a reasonable time to unwind a position, VaR estimates need to be re-calculated every 10 days (nearly a week). For fund mangers sensitive to market risk, one month (nearly 20 days) is usually a sufficiently long holding period for them to adjust their positions for rebalancing risk/return. Therefore, analyzing these two forecast-horizons can generate important implications for short-term risk management. Similarly, those correlations calculated for the next three months may then be useful for medium-term asset allocation strategy. In this paper, to perform cross-horizon forecast comparison, we use GFESM ranking test of Newbold, Harvey and Leybourne (1999).

Next, we proceed as follows. In Section 5.2 we review some of the literature concerning the use of implied correlation in different financial markets and various forecast evaluation methods. The emphasis here is put onto those with an economic loss function. Then, specifications of eleven competing models and their multi-step ahead correlation forecasting function are depicted in Section 5.3. In Section 5.4, we present three statistical methods to examine the optimality and information efficiency of correlation forecast. After illustrating the data and empirical results in section 5.5 and 5.6 respectively, we conclude in section 5.7.

5. 2 Literature reviews

Since the main attention in this chapter is paid to the implied correlation model, we present below an overview of the literature concerning the application of this option-driven information source to different financial markets. Besides this, several works that contribute to the evaluation of correlation forecasts under different economic loss functions are also summarized to highlight the practical use of these forecasts for trading in daily foreign exchange market and equity market.

a. Using implied correlation in different markets

To our best knowledge, the first paper to study implied correlation in foreign exchange market was written by Bodurtha and Shen (1994), where the authors matched the option data collected from PHLX (Philadelphia stock exchange) to calculate correlation of two exchange traded currency pairs USD/DEM and USD/JPY. By extending the univariate implied volatility estimation method of Whaley (1982) to bivariate cases, the authors computed implied correlation and compared the results to three historical information-based forecasts to determine individual information contributions. Since a high degree of autocorrelation was found, Stock-Watson's (1993) OLS procedure was used in their regression test to evaluate predictive accuracy. The results showed historical information and option-driven information were both very useful for predicting future correlation.

A similar investigation using exchange-traded option data was performed in Siegel (1997). Compared to previous studies, a larger sample including two currency trios was used in his research and missing values in implied volatility were input using monthly average of unconditional volatility so as to avoid interpolation. The author examined the implied correlation from a hedger's perspective. Concretely, Siegel calculated the actual risk reduction after a standardized exposure was proportionally hedged using ratios (correlations) generated by different econometric models, and his findings suggested the implied correlation model was statistically the 'best' in terms of the volatility that can be reduced.

In the above cases, a common feature is that implied correlation was calculated from implied volatility data obtained from a specific exchange. However, it is now well-understood that OTC markets can provide a more informative source than traditional exchange market to extract embedded information of option contract. Market sentiment exploited from this source is also more versatile. Besides this, other credits, such as derivative contract's constant maturity and currency option's exactly at-the-money strikes, also contribute to the effectiveness of its information. Based on these motivations, Campa and Chang (1997) then re-addressed the issue of correlation forecasting in currency market using OTC implied volatility data. After analysis, they found the forecast combination test implied that correlation could always incrementally improve the performance of other forecasts, and this informative superiority held even when forecast errors were weighted by realized volatility. A further step was taken by Walter and Lopez (2000) where a significantly different cross-trio performance of implied correlation was documented. For example, in one trio (USD/DEM/JPY) the authors showed that implied correlation was statistically useful in predicting realized correlation and the resulting estimate was partially optimal to the information used to generate them although these forecasts did not fully incorporate the information presented in historical prices. However, concerning the others (USD/DEM/CHF), the economic benefit of using implied correlation then diminished a great deal whilst forecasts themselves still remain statistically optimal. Thus, they concluded that forecasting performance of implied correlation was actually an empirical issue.

In addition to the three-currency trio, implied correlation can also be calculated in the equity market (often called implied beta in this case), if a specific condition is met. That is, we can construct an authentic portfolio whose constituents and the portfolio itself both have traded option contracts. Siegel (1995) performed such a study using three interactive exchanged traded options: one equity option, one equity index option and one option to exchange stock for shares of the market index to calculated implied correlation (beta) of an individual equity with respect to the whole market index. Skintzi and Refenes (2003) took a step further, by utilizing portfolio theory to calculate implied correlation of Dow Jones average index relative to all its constituents. In their research, a new measure of diversification was suggested and calculated through a so-called average implied correlation index. Note that this result has very important implications for practical asset allocation because overall market expectation of the future correlation (or diversification effect) in the US market can now be readily supervised and fund managers who are inclined to adopt a passive strategy by only tracking the stock index can simply rely on this benchmark to access their portfolio's risk-return profile.

b. Correlation forecast evaluation under economic loss functions

As acknowledged by countless practitioners, since the ultimate aim of developing correlation and volatility models is not just to fit coherently the past data but, more importantly, to forecast these latent variables so that resulting estimates can be input to a specific mechanism to generate profits (or test market inefficiency), it is then necessary to assess the target model not only under a statistical loss function but also under an economic loss function.

Usually, statistical loss function is the most common criterion applied in finance to determine the optimality of a forecast. However, frequently, the best model picked by using this method is sensitive to the loss function itself. Thus, it may appear that the chosen model differs along with the loss function used to evaluate them. To obtain a more practical view, it is then necessary to complement the existing statistical evaluation method using more economically-oriented loss functions. As for volatility forecasting, criteria like the trading profitability function of Engle *et al.,* (1993) and the probability loss function of Lopez (1999) have already been proposed and examined in the literature. However, concerning the correlation forecasting, comparatively little were done, although several works are still worth mentioning. For example, following the study of Siegel (1997), Brooks and Chong (2001) compared the correlation forecasting performance of eleven models, including time-series ones and an option-driven model, by computing optimal hedge ratio. Contrary to most of the findings in similar areas, they suggested that the option market was a poor information source from which to extract accurate hedge ratio and only EWMA in their samples prevailed. By extending Engle *et al.,* (1993) and Gibson and Boyer (1998), Chong (2004) re-examined the economic losses of different correlation models under an authentic trading profitability function. Among all the examples analyzed, univariate EGARCH was found to be the best in terms of wealth that could be accumulated. The author confirmed the weak form efficiency in currency market after transaction cost was taken into account. That is, the directional bets taken before transaction costs were charged can generate positive returns; however, when this cost was accounted, profits then immediately evaporated. Besides this, a similar investigation, emphasizing the VaR estimates, can also be found in literature (see Chong, 2005, for more details).

5.3 Correlation Forecasting Models

In this section, we describe eleven correlation models to be used in our later forecast generation and comparison. To ease the expression, we categorize these models into three groups. They are historical correlation models, conditional heteroskedastic models and implied correlation model. Some of them such as EWMA and multivariate GARCH models have already been briefly illustrated in chapter one. Now, a detailed description of their specification and statistical characteristics is provided below.

5.3.1 Historical correlation and EWMA

First, we describe two intuitively simple correlation models. One is the historical correlation model. The other is the exponential weighted moving average model, called EWMA or exponential smoother. Both models estimate and forecast correlation by exploiting historical information. As a result of their simple specifications, they have both gained substantial popularity in industrial uses.[38]

[38] Since their model specifications are so simple, correlation forecasts generated from them are usually called simple forecasts.

a. Historical correlation model

Concerning the first, consider now two currency pairs B/A and C/A. Conditional correlation forecasts of this model, made at time *t* with a forecast horizon of *T* days and a past history of *P* days, is calculated by

$$\rho(r_{B/A,}r_{C/A})_{t,T} = \frac{\sum_{i=1}^{P}(r_{B/A,t-i+1} - \bar{r}_{B/A})(r_{C/A,t-i+1} - \bar{r}_{C/A})}{\sqrt{\sum_{i=1}^{P}(r_{B/A,t-i+1} - \bar{r}_{B/A})^2}\sqrt{\sum_{i=1}^{P}(r_{C/A,t-i+1} - \bar{r}_{C/A})^2}} \qquad (5.1)$$

where $r_{B/A,t}$ and $r_{C/A,t}$ denote the conditional returns of B/A and C/A at *t*; $\bar{r}_{B/A}$ and $\bar{r}_{C/A}$ represent their corresponding sample means.

Here, since the forecast horizon *T* cannot be found in the right hand side of above equation, it is fair to say that the correlation to be generated would be independent of this horizon and the resulting forecast would present a flat-term structure. Applied in this chapter, correlation forecasts for the next week will then be equal to the one for the next one-month and the one for the next three-months once the length of past price history is determined. Given this feature, since the only parameter we can tune now is *P*, it is then preferred that the value of this variable can be set as long as possible. In this chapter, to make the recent observations the most relevant information to predict the future, we let the price histories of historical correlation models have the same length as the forecast horizons of interest. As illustrated earlier, since forecast horizons are now set at one-week, one-month and three-months, we respectively consider three historical correlation models here with *P* equalling 7 days, 22 days and 65 days. Thus these models are called HISTOR7, HISTOR22 and HISTOR65. While forecasting, we use rolling window to make sure the length of *P* is kept fixed as *t* evolves. For example, if today is *t*, correlation forecast generated by HISTOR65 for all future days made today is then based on the past observations from *t-65* to *t*. Analogously, with the forecasts made tomorrow *t+1* is calculated by using data from *t-64* to *t+1*.

b. EWMA

From equation (5.1), one can easily note that all samples included in the past price history are given the same importance. However, it is understood that observations taken far from the time when the forecast is made may have little impact in the whole sample. Thus, theoretically, these observations are supposed to be assigned less weight than those representing the recent history.

To make this amendment, JP Morgan proposed a solution. In the risk management tool *RiskMetrics*™ proposed by them, a decay factor λ is introduced to equation (5.1) to formulize the EWMA model. Through this refinement, a time-sensitive structure for modelling correlation dynamics is then presented with recent observations given greater importance than all earlier ones. For example, consider now the same currency pairs B/A and C/A as seen earlier: the correlation forecast with price history *P* using EWMA is now calculated by

$$\rho(r_{B/A}, r_{C/A})_{\lambda,P,t} = \frac{\sum_{i=0}^{P} \lambda^i (r_{B/A,t-i} - \bar{r}_{B/A})(r_{C/A,t-i} - \bar{r}_{C/A})}{\sqrt{\sum_{i=0}^{P} \lambda^i (r_{B/A,t-i} - \bar{r}_{B/A})^2} \sqrt{\sum_{i=0}^{P} \lambda^i (r_{C/A,t-i} - \bar{r}_{C/A})^2}} \tag{5.2}$$

Here, concerning equation (5.2), it is important to note that, although the allocation of importance for various observations is resolved, forecasts once generated still present a flat-term structure because *T* now once again is eliminated in the right hand side of the correlation generating process. Thus, for a fixed *P,* correlation forecasts generated by EWMA would be the same for all forecast horizons of interest. Besides, as before, there is no empirical fitting needed in this case (there is no parameter we need to estimate). Thus, we can simply use industrial standard to determine the value of decay factor, $\lambda = 0.94$, and set *P* equal to 1000 to ensure a long past-price history.

5.3.2 Conditional Heteroskedastic Models

Above, it has been shown that, for historical correlation models, although in-sample correlation can be modelled as a time-varying variable, its forecasts (out-of-sample correlation) are time invariant. Thus, the dynamic property of correlation is not captured due to the flat-term structure assumed in these models' mechanisms. In order to more flexibly model the correlation evolving process, a natural solution is then to utilize multivariate GARCH models. In the following, we describe the specification and property of three different types of GARCH models for computing time-varying correlation. To see their recursive functions for generating multi-step ahead forecast, an illustration is also provided.

5.3.2.1 Multivariate GARCH models

a. Diagonal VECH and Diagonal BEKK

First, in the multivariate context, we describe two typical heteroskedastic models for estimating time-varying covariance and, accordingly, time-varying correlation. One is VECH, the other is BEKK. Both models are earliest multivariate GARCH models proposed in literature to model covariance as a function of its own lagged terms and past innovations. Although BEKK is generally considered a refinement of VECH because, by imposing a series of quadratic terms on parameter values, positive definitiveness of resultant covariance matrix is ensured, both models are often criticized for their high estimation cost. This is because, for a portfolio with even a very small number of assets, estimation using these two specifications is often associated with a very large parameter set. Taking the most generalized form of an *N*-dimensional VECH model, for example, one needs to estimate a staggering amount of $(N^2+N)*(N^2+N+1)/2$ parameters. That is, for a bivariate case, 21 parameters need to be computed simultaneously through either maximizing the log-likelihood or Bayesian inference.[39]

To circumvent this numerical difficulty, various strategies are proposed. For example, Engle and Mezrich (1996), by forcing the model implied unconditional covariance to equal a pre-calculated sample average, suggested using the 'variance targeting' technique. In so doing, non-linear estimation of the interception parameter is then purposely avoided. More often, in order to achieve additional parsimony, restrictions on parameters are imposed directly on variance equation either through trimming the parameter matrix or just changing the whole parameterization. For instance, in VECH, by letting all parameter matrices be diagonal, the number of elements that need to be estimated is then reduced to $3(N^2+N)/2$.[40] Although the goal of reducing the cost is partially achieved, empirically, with o(N^2) parameters still needing to be estimated, this method clearly is not suitable for solving systems of medium and large size. Besides, as a price to pay, the flexibility of dynamics being modelled is also downgraded. Similarly, for BEKK, if this strategy is adopted, a substantial relief of numerical cost is also expected although the benefit is considered modest.

[39] To obtain a clear view of the massive parameters included in the VEC model, consider a bivariate innovation $\varepsilon_t = [\varepsilon_{1,t}, \varepsilon_{2,t}]^{'} \sim N(0, \Sigma_t)$ where $\Sigma_t = \begin{pmatrix} \sigma_{1,t} & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{2,t} \end{pmatrix}$, $\Sigma_{1,t}, \Sigma_{2,t}$ are the individual variance of two assets, $\Sigma_{12,t}$ is the covariance. Here, if ε_t is modelled by a full version bivariate VEC, then

$$\begin{pmatrix} \Sigma_{1,t} \\ \Sigma_{12,t} \\ \Sigma_{2,t} \end{pmatrix} = \begin{pmatrix} \varpi_{11} \\ \varpi_{12} \\ \varpi_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{pmatrix} + \begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix} \begin{pmatrix} \sigma_{1,t-1} \\ \sigma_{12,t-1} \\ \sigma_{2,t-1} \end{pmatrix}$$

[40] Consider the same innovation $[\varepsilon_{1,t}, \varepsilon_{2,t}]^{'}$ as above; the matrix form of the diagonal bivariate VEC can be specified as

$$\begin{pmatrix} \Sigma_{1,t} \\ \Sigma_{12,t} \\ \Sigma_{2,t} \end{pmatrix} = \begin{pmatrix} \varpi_{11} \\ \varpi_{12} \\ \varpi_{22} \end{pmatrix} + \begin{pmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{pmatrix} \begin{pmatrix} \varepsilon_{1,t-1}^2 \\ \varepsilon_{1,t}\varepsilon_{2,t-1} \\ \varepsilon_{2,t-1}^2 \end{pmatrix} + \begin{pmatrix} \beta_{11} & 0 & 0 \\ 0 & \beta_{22} & 0 \\ 0 & 0 & \beta_{33} \end{pmatrix} \begin{pmatrix} \sigma_{1,t-1} \\ \sigma_{12,t-1} \\ \sigma_{2,t-1} \end{pmatrix}$$

For the purpose of this Chapter, since our research aim is mainly on implied correlation, concerning its alternatives, models are then proposed in their most parsimonious forms to avoid numerical difficulties. Applied to this case, we then respectively choose diagonal VECH (1, 1) and diagonal BEKK (1, 1) for generating time-varying correlation. As for the specifications of these two models, we have presented them in Chapter one. Now, it only remains for us to stress that, in order to ensure the stationarity, we impose a non-linear restriction on *arch-* and *garch*-parameter of these models so that eigenvalues of their summation will lie within the unit circle.

b. Conditional Correlation models

Apart from adopting an autoregressive function to model covariance matrix like VECH and BEKK, one can also use, in the multivariate GARCH framework, methods developed by Bollerslev (1990) and extended by Engle and Sheppard (2002), to generate conditional correlation forecasts by separating the covariance matrix into a volatility part and a correlation part and then using a series of independent dynamic processes to model them.

To calculate this conditional correlation, consider now a *d*-asset portfolio whose vector of return and corresponding residuals are respectively denoted by r_t (*d*-dimensional) and ε_t, and its conditional correlation for two assets, say *i* and *j*, is calculated by

$$R = \rho_{ij} = \frac{E_{t-1}(\varepsilon_{i,t}\varepsilon_{j,t})}{\sqrt{E_{t-1}(\varepsilon_{i,t}^2)E_{t-1}(\varepsilon_{j,t}^2)}} = E_{t-1}(\varepsilon_{i,t}\varepsilon_{j,t}) \tag{5.3}$$

where $\forall i, j \in d$. Here, it is necessary to standardize (or normalize) these residuals (to let the means of ε_{it} and ε_{jt} equal zero and their variances equal one) so that the denominator of (5.3) equals one and can be absorbed.

Since r_i can also be expressed as

$$r_{i,t} = \Sigma_{i,t}^{\ 1/2}\varepsilon_{i,t} \quad r_{j,t} = \Sigma_{j,t}^{\ 1/2}\varepsilon_{j,t} \tag{5.4}$$

after transformation we can readily obtain

$$\varepsilon_t = D_t^{-1} r_t \tag{5.5}$$

where $D_t = diag(\sqrt{\Sigma_t})$ is a diagonal matrix with $\sqrt{\Sigma_t}$ on its i^{th} diagonal, denoting the univariate volatility of i^{th} time series.

Constant Conditional Correlation (CCC)

In Bollerslev (1990), each element in D is modelled using a univariate GARCH so that its variance-covariance matrix can be computed by $\Sigma_t = D_t R D_t$.[41] The CCC model, after (5.5) is inserted to (5.3), can be specified as,

$$\begin{gathered} r_t \middle| \Omega_{t-1} \sim N\left(0, \Sigma_t\right); \Sigma_t = D_t R D_t \\ CCC(1,1) \longrightarrow D_{i,t}^2 = \varpi_i + \alpha_i \circ r_{i,t-1} r_{t-1}^{'} + \beta_i \circ D_{i,t-1}^2 \\ R = E_{t-1}(\varepsilon_t \varepsilon_t^{'}) = D_t^{-1} \Sigma_t D_t^{-1} \end{gathered} \tag{5.6}$$

Here, to ensure the positive definitive of resultant covariance matrix, Bollerslev (1990) utilized the full rankness of parameter matrix. Since R is now assumed to be constant, this quantity can be readily calculated by $R = \sum_{i=1}^{n} (\varepsilon_i \varepsilon_i^{'}) / n$ once innovations of all returns have been standardized by computed GARCH volatility. The forecasts, once generated, will again be the same for all forecast horizon of interest since the correlation is now assumed to be constant.

Dynamic Conditional Correlation (DCC)

In CCC, randomness in the covariance matrix solely depends on the randomness in individual volatilities. However, a more flexible model can be obtained if condition correlation, just like univariate volatility, is also allowed to be time-varying and generated from a dynamic process. Engle and Sheppard (2002) provided such a solution by using another univariate GARCH, independent of those used for modelling D, to model the correlation evolving process.

Concretely, the authors used standardized residual generated from (5.5) as input to estimate a univariate GARCH so that an authentic covariance matrix Q can be fitted. Since time-varying property is purposely incorporated into this new covariance matrix, a simple transformation of Q can help retrieve the conditional correlation, now also as a time-varying quantity. Given this virtue, specification of DCC model then can be written as

$$DCC(1,1) \longrightarrow \begin{aligned} & r_t \mid \Omega_{t-1} \sim N\left(0, D_t R_t D_t\right); \qquad \varepsilon_t = D_t^{-1} r_t \\ & D_{i,t}^2 = \varpi_i + \alpha_i \circ r_{i,t-1} r_{t-1}^{'} + \beta_i \circ D_{i,t-1}^2 \\ & Q_t = \bar{Q} \circ (1 - \eta - \varsigma) + \eta \circ \varepsilon_{t-1} \varepsilon_{t-1}^{'} + \varsigma \circ Q_{t-1}; \\ & R_t = diag\left\{Q_t\right\}^{-1/2} Q_t diag\left\{Q_t\right\}^{-1/2} \end{aligned} \tag{5.7}$$

Here, $\circ$ denotes the Hadamard product of two identically-sized matrices. Parameterization for $D_{i,t}$ is set equal to those illustrated in (5.6). However, as for $Q_{t,}$, since this variable is now

[41] For a bivariate time series, the separation of volatility part and correlation part in a variance-covariance matrix, given a conditional correlation model, can be described as

$$\Sigma_t = D_t R D_t \longrightarrow \begin{pmatrix} \sigma_{1,t} & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{2,t} \end{pmatrix} = \begin{pmatrix} \sqrt{\sigma_{1,t}} & 0 \\ 0 & \sqrt{\sigma_{2,t}} \end{pmatrix} \begin{pmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{pmatrix} \begin{pmatrix} \sqrt{\sigma_{1,t}} & 0 \\ 0 & \sqrt{\sigma_{2,t}} \end{pmatrix}$$

intrinsically related to Σ_t, we have a variety of choices for modelling it. For example, an exponential smoother can be applied so that each element $q_{ij,t}$ in this authentic matrix can be calculated using $q_{ij,t} = \bar{\rho}_{ij}(1-\lambda)(\varepsilon_{i,t-1}\varepsilon_{j,t-1}) + \lambda q_{ij,t-1}$. However, a more frequently-used case, just as that given in (5.7), is to fit Q using another unidimensional GARCH so that the resulting structure can be interpreted as a GARCH-in-GARCH. $\bar{Q}$ in (5.7) then represents the unconditional (sample) covariance of standardized residuals.

Here, several things need to be stated concerning this model before we proceed further. First, with regard to its estimation, usually a two-step procedure will be adopted to maximize the log-likelihood function. That is, we start from estimating the GARCH parameters governing the volatility evolving process to estimating similar parameters used to model the correlation process.[42] Although, in the optimization step, the target log-likelihood function will be separated, consistency and unbiasedness of the resulting maximum likelihood estimators are asymptotically ensured (See Newey and McFadden, 1994, for evidence). Second, the positive definitiveness of covariance matrix can be guaranteed by imposing a proper parameterization. For example, one can use Cholesky decomposition to reparameterize Σ_t (see Tsay 2002). The advantage of using this approach is that it requires no constraints for the positive definitiveness of covariance matrix.[43] However, the drawback is that the interpretation of the resultant parameter after covariance transformation will then become a difficult task. Besides, we can also, by squaring the parameter vectors like specification proposed in Hafner and Franses (2003), achieve the same goal.[44] Finally, it is also necessary to mention that a significant success of DCC is its massive reduction of associated parameters for estimation to only *N*, that is, the same as the number of assets included in a target portfolio. Thus, fitting a large covariance matrix becomes economically feasible even for institutional investors who may

[42] Log-likelihood function of CCC and DCC can be decomposed into volatility part and correlation part through the form of $L(\theta,\psi) = L_{Volatility}(\theta) + L_{Correlation}(\psi \mid \theta)$, where θ,ψ represents the volatility and correlation parameters respectively. Usually, the univariate volatility log-likelihood function will be maximized first, followed by the function concerning the correlation parameters.

[43] Cholesky decomposition of Σ_t can be written as

$$\Sigma_t = \begin{pmatrix} \sigma_{1,t} & \sigma_{12,t} \\ \sigma_{12,t} & \sigma_{2,t} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ \sigma_{12,t}/\sigma_{1,t} & 1 \end{pmatrix} \begin{pmatrix} \sigma_{1,t} & 0 \\ 0 & \sigma_{2,t} - \sigma_{12,t}/\sigma_{1,t} \end{pmatrix} \begin{pmatrix} 1 & \sigma_{12,t}/\sigma_{1,t} \\ 0 & 1 \end{pmatrix}$$

[44] Positive definitiveness of covariance matrix $\Sigma_t = D_t R_t D_t$ can be ensured once positive definitiveness of R_t is ensured. To achieve the goal, we can use either $\rho^*_{ij,t} = \rho_{ij,t}\big/\sqrt{1+\rho_{ij,t}}$ or $\rho^*_{ij,t} = \exp(\rho_{ij,t}) - 1\big/\exp(\rho_{ij,t}) + 1$ to transform the estimated correlation coefficient $\rho_{ij,t}$ in

$$R_t = \begin{pmatrix} 1 & \rho_{ij,t} \\ \rho_{ij,t} & 1 \end{pmatrix}$$

so that $\left|\rho^*_{ij,t}\right| < 1$ and positive definitiveness of R_t then can be ensured.

hold hundreds of assets at any one time. Given this clear advantage over other multivariate GARCH models, like VECH and BEKK, a recent and growing body of work is now dedicated to proposing DCC variants. For example, Tse *et al*., (2002) introduced a weighting function to covariance dynamics. A Markovian regime-switching structure is imposed by Pelletier (2004) to enhance the correlation dynamics.

5.3.2.2 Simplified univariate GARCH

Indeed, the multivariate model provides a naïve solution to model covariance dynamics so that time-varying correlation can be extracted from it. However, this is not to say that only a multivariate structure can be utilized to calculate association measure. Recently, Harris *et al.,* (2004) proposed a new method for generating correlation dynamics using only multiple univariate models. In their method, four univariate GARCH models are estimated to calculate the time-varying correlation of bivariate return. In the univariate context, although the number of models to be fitted now increases, their total estimation cost, when compared to that of a multivariate GARCH, is still lower because the number of parameters increases only on a linear rate with dimensionality.

Now, consider two standardized residuals $\varepsilon_{1,t}, \varepsilon_{2,t}$ and their corresponding standard deviation $\sigma_{1,t}$ and $\sigma_{2,t}$. To calculate time-varying correlation, according to Harris *et al.,* (2004), first, it is necessary to construct two new innovations representing the summation and the subtraction of the original series, $\xi_{+,t} = \varepsilon_{1,t} + \varepsilon_{2,t}$ and $\xi_{-,t} = \varepsilon_{1,t} - \varepsilon_{2,t}$, and give their conditional variance equations respectively by $\sigma_{+,t} = \sigma_{1,t}^2 + \sigma_{2,t}^2 + 2\sigma_{12,t}$ and $\sigma_{-,t} = \sigma_{1,t}^2 + \sigma_{2,t}^2 - 2\sigma_{12,t}$. Then, by adding up $\sigma_{+,t}$ and $\sigma_{-.t}$ so that covariance of $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ is equal to $\sigma_{12,t} = \left(\sigma_{+,t}^2 - \sigma_{-,t}^2\right)/4$, time-varying correlation can be readily computed by $\sigma_{12,t} / \sigma_{1,t}\sigma_{2,t}$.

Here, to obtain the time-varying estimates for $\sigma_{12,t}$ and $\sigma_{1,t}, \sigma_{2,t}$, we need to fit four times a standard univariate GARCH respectively to $\varepsilon_{1,t}, \varepsilon_{2,t}, \xi_{+,t}$ and $\xi_{-,t}$, since financial return is often characterized by significant evidence of fat tails. Apart from using a standard GARCH (with Gaussian error), it is desirable for us to also incorporate this feature into the modelling of univariate volatility. For this purpose, we then consider the use of a Generalized Error Distribution (also known as exponential power distribution) with a univariate GARCH. This density was initially proposed in Subbotin (1923) and later developed by Johnson (see Johnson, Kotz, and Balakrishnan, 1995, for overview) to account for the leptokurtosis. Since its distributional form is so flexible that a variety of standard densities such as Gaussian, Laplace, Weibull and Pareto can be nested, in many cases it is also applied to various financial situations (see Nelson, 1991, for its application in fitting stock index return and Hsieh, 1989, for its application in fitting foreign exchange returns).[45]

5.3.2.3 GARCH correlation forecasting

Since a prime interest of this chapter is to obtain correlation forecast, it is now necessary to proceed further to present recursive functions of GARCH models for generating multi-step-ahead covariance forecasts so that correlation over a future period can be calculated. [46] Here, to generate these forecasts we use the same rolling window as those illustrated in Section 5.3.1. Besides, since correlation evolving process, based on heteroskedastic models, is assumed to be step-dependent, we use the temporal aggregation rule to calculate their horizon forecast.

a. Traditional GARCH forecasting

First, concerning the use of Diagonal-VECH, Diagonal-BEKK and simplified univariate GARCH models (Normal/GED), since their variance-covariance matrix $\Sigma_{ij,t}$ can be written by

$$\Sigma_{ij,t} = \varpi + \alpha(\varepsilon_{t-1}\varepsilon_{t-1}^{'}) + \beta\Sigma_{ij,t-1} \tag{5.8}$$

to generate *K*-step-ahead forecast of $\Sigma_{ij,t}$, we only need to calculate

$$\Sigma_{ij,t+K} = \omega\sum_{k=0}^{K-2}\left(\alpha+\beta\right)^{k} + \left(\alpha+\beta\right)^{K-1}\Sigma_{ij,t+1} \qquad \text{for } K \geq 2 \tag{5.9}$$

However, to obtain the forecast over the whole horizon *T*, it is then necessary to aggregate all *K*-step-ahead variance-covariance matrix forecasts included in this horizon and then divide the result by volatility forecast of *i* and *j* over the same horizon so that the resulting $\hat{\rho}(\varepsilon_i,\varepsilon_j)_{t,T}$ can be expressed as

$$\hat{\rho}(\varepsilon_i,\varepsilon_j)_{t,T} = \frac{\Sigma_{ij,T}}{\sqrt{\Sigma_{i,T}\Sigma_{j,T}}} \tag{5.10}$$

[45] Note that, in spite of fat tails, other stylized features of return distribution such as leverage effect of innovations will not be examined in this paper. This is because foreign exchange market is usually not characterized by pronounced asymmetry, which is frequently presented in equity market (See Camp and Chang, 1997 for evidence).

[46] Here, the forecast of interest is the correlation over the next *T* days, or, say, the horizon *T*. It is a different concept from the multi-step-ahead forecast that indicates the correlation of a future specific day such as the correlation forecast for day *t+7* made on day *t*. Usually, practitioners are only interested in the horizon forecast, because it meets their needs much better.

where $\Sigma_{ij,T}$ denotes the aggregated covariance forecast estimated at time t (omitted) for horizon T and $\Sigma_{ij,T} = \sum_{K=1}^{T} \Sigma_{ij,t+K}$.

b. DCC Correlation Forecasting

If DCC model is used, then there are two approaches available to solve its recursive forecasting function foreword, through which we can obtain the multi-step-ahead correlation forecast. First, in equation (5.7), since Q is now modelled by

$$Q_t = \bar{Q}\circ(1-\eta-\varsigma)+\eta\circ\varepsilon_{t-1}\varepsilon_{t-1}^{'}+\varsigma\circ Q_{t-1} \quad \text{where } E[\varepsilon_{t-1}\varepsilon_{t-1}^{'}]=R_{t-1} \qquad (5.11)$$

we can make the approximation $E[\varepsilon_{t-1}\varepsilon'_{t-1}] \approx Q_{t-1}$ directly so that derivation of K-step-ahead covariance forecast Q_{t+K} is similar to the process assumed in equation (5.9) and R_{t+K} can be computed analytically by $R_{t+K} = diag\{Q_{t+K}\}^{-1/2} Q_{t+K} diag\{Q_{t+K}\}^{-1/2}$. Besides, we can also let $\bar{Q} \approx \bar{R}$ and $E[Q_{t-1}] \approx R_{t-1}$ so that updating of covariance matrix is no longer required in each step. K-step-ahead correlation forecast can be readily computed by

$$R_{t+K} = (1-\eta-\varsigma)\bar{R}\sum_{k=0}^{K-2}(\eta+\varsigma)^{k} + (\eta+\varsigma)^{K-1}R_{t+1} \quad \text{for } K \geq 2 \qquad (5.12)$$

Here, note that both approaches can be used to generate correlation forecast of a future date. However, after testing the prediction bias, Engle and Sheppard (2001) confirmed the second method could provide a slightly better performance than the first although neither of them can significantly outperform the other. To exploit this result, we thus use equation (5.12) to generate multi-step-ahead correlation forecast in this paper.

Here, to calculate $\hat{\rho}(\varepsilon_i, \varepsilon_j)_{t,T}$, since variance forecasts $\Sigma_{i,T}, \Sigma_{j,T}$ over the horizon T can be readily obtained, according to (5.10) we only need to compute

$$\Sigma_{ij,T} \equiv \sum_{K=1}^{T}\Sigma_{ij,t+K} = \sum_{K=1}^{T} D_{t+K}R_{t+K}D_{t+K} \qquad (5.13)$$

5.2.3 Implied correlation

As seen above, volatility and correlation forecasts are all generated from a time series model using historical returns as input. However, empirically, it has been repeatedly argued that option price is also an efficient information source which can be exploited to predict these latent variables, since a direct mapping is now provided. For example, through either a stochastic volatility model or the Black Shores model, implied volatility can be computed to forecast future realized volatility.[47] Although, at least on a statistical level, we can argue that these models are not justified due to the stringent conditions assumed, empirically, massive evidence has been reported on the information superiority of, say, implied volatility over other competing forecasts. Thus, it is fair to conclude that implied volatility though might produce biased estimates for realized volatility can still reveal at least to a certain extent the true market expectation, no matter which model is used here for mapping. Given this feature, it is then interesting to see whether the implied correlation, calculated from implied volatility, will possess the same property.

To calculate implied correlation, first it is necessary to ensure that we can identify three assets that have a triangular relationship and they all have option contracts traded on them. A typical example can be illustrated through a three-currency trio. Concretely, consider now a sample trio A/B/C where B/C can be regarded as a portfolio of B/A and C/A. Given this authentic portfolio, conditional variance of B/C at time *t* according to Markweiz's portfolio theory then can be calculated by

$$\Sigma_{B/C,t} = \Sigma_{B/A,t} + \Sigma_{C/A,t} - 2\rho_{(B/A,C/A),t}\sqrt{\Sigma_{B/A,t}\Sigma_{C/A,t}} \tag{5.14}$$

using variance of B/A and C/A at the same time. Since only univariate volatility now needs to be estimated, implied correlation of B/A and C/A with forecast horizon *T* then can be readily computed by

$$\rho_{IC}(r_{B/A}, r_{C/A})_{t,T} = \frac{\Sigma_{IV(B/A),t,T} + \Sigma_{IV(C/A),t,T} - \Sigma_{IV(B/C),t,T}}{2\sqrt{\Sigma_{IV(B/C),t,T} \cdot \Sigma_{IV(C/A),t,T}}} \tag{5.15}$$

after unconditional variance Σ in (5.15) is replaced with implied volatility Σ_{IV}.[48] Here, since forecast horizon has already been implied in outstanding maturities of each option contract, we do not need to use the recursive function as that required in GARCH models to generate the multi-step-ahead correlation forecast.

[47] Most stochastic volatility (*SV*) models assume the volatility follows a similar stochastic process as asset returns. For example, consider a derivative asset *f* with a price that depends on some security prices *S* and instantaneous variance $V = \sigma^2$. Then, a typical SV model can be written as,

$$dS = \phi S dt + \sigma S d\varpi$$
$$dV = \mu V dt + \xi V dz$$

where the wiener process *dz* and *dw* can be either independent or dependent with correlation ρ. Note here that, to assume a specific stochastic process for the latent variable to be modelled is a very stringent assumption in finance. This is because it has restricted the sample paths of the resulting estimates to follow a specific pattern. However, as a comparison, a more flexible substitute for this assumption is to assume a distribution rather than a stochastic process for the underlying.

[48] Implied volatility here is calculated from Garman-Kohlhagen option pricing model.

However, it is still important to note one thing before we proceed. That is, in equation (5.14) and equation (5.15) a clear distinction needs to be made as to which set of assets (A/B, A/C) or (C/A, B/A) construct the authentic portfolio for calculating the variance of B/C. As can be seen, usually there are two ways simultaneously available to express the same cross-products in the foreign exchange market. Although the conditional volatility of A/B would be surely related to the volatility of B/A, in few cases will they be identical.[49] Thus, the resulting correlation derived from B/A and C/A pair is supposed to be different from that calculated from A/B and A/C pair. In this paper, to circumvent this potential confusion, we let only the intermediate currency stay at the denominator of the cross-product. Therefore, B/C for our cases only corresponds to B/A and C/A.

5.4 Realized Correlation and Forecast Evaluation

Now it is necessary to state how to calculate the realized correlation that various forecasts, once generated, can be compared with and the evaluation methods to access these forecasts.

5.4.1 Realized Correlation

Concerning this topic, it is then beneficial to highlight some similar researches performed for calculating the realized volatility. Since volatility and correlation are both unobservable in financial markets, to benchmark their forecasts, some auxiliary assumptions then need to be made to explaining on how the ex-post values are to be computed.

a. Calculating Realized Volatility

As for volatility, most of the early research work used squared daily returns to approximate the realized volatility of the same frequency (see Day and Lewis, 1992; Jorion, 1995, for example).[50] This method is intuitively simple to use, although the resulting estimates are often found noisy. In order to produce a more precise value, Anderson *et al.,* (2000) suggested using the summation of higher frequency (intra-day) returns so that 'realized volatility' can be approximated by

$$\tilde{\Sigma}_t \approx r_t^{2*} \equiv \sum_{f=1}^{F} r^2_{(t-1+f/F)} \qquad (5.16)$$

[49] Consider now an example. Given the price today p_t and yesterday p_{t-1} of B/C, its corresponding return is then calculated by $(p_t-p_{t-1})/p_{t-1}$. However, as for C/B, its return is computed by $(1/p_t-1/p_{t-1})/(1/p_{t-1})$ which after transformation is equal to $(p_{t-1}-p_t)/p_t$. Clearly, the volatility estimates for these two series will not be equal.

[50] Consider a random variable r_t which satisfies $r_t = \Sigma_t^{1/2}\varepsilon_t$. Σ_t represents the time-varying volatility; ε_t denotes an unspecified stochastic process with mean zero and constant variance. Now, if we add expectation to both sides of $r_t = \Sigma_t^{1/2}\varepsilon_t$, it is easy to obtain $E_t(\Sigma_t) \approx E_t(R_t)^2$. Thus, realized volatility can be just approximated by the squared daily return.

where F denotes the sampling frequency of intra-day data.

Given equation (5.16), it is natural to expect that this sampling frequency could be set as large as possible so that more information could be incorporated to calculate each realized volatility estimate. However, here there is an empirical problem. Usually, we may be short of a sufficiently long span of such intra-day data if F is set too large. Besides, even if this data is now available, many intervals with few or no trade can be found, leading to either a missing observation or a zero return. In this case, market microstructure may then take immediate effect by introducing unexpected bias, and the improved accuracy just obtained for approximating realized volatility can be easily offset. In addition, there are other open questions still being debated on the use of these intra-day data. For example, it might be asked 'which frequency of data is really high enough to make the approximation of realized volatility both accurate and cheap enough'? Obviously, this is a decision concerning the trade-off between accuracy and economic cost. Since no general consensus has been reached on this particular issue, it is often considered as an empirical question. For instance, Andersen *et al*., (2000) used 5 minutes intraday data to estimate daily realized volatility, while 10 and 30 minutes data are chosen in Granger et al., (2003) and Koopman *et al*'s (2000) respective works.

b. Calculating Realized Correlation

In literature, two methods can be used to calculate realized correlation. One is to use a forward-looking historical correlation model (see Walter and Lopez, 2000). Consider now the same sample pairs B/A and C/A as before; their realized correlation estimated at time t with a forecast horizon of T days, is calculated by,

$$\tilde{\rho}_{RC}\left(r_{B/A}, r_{C/A}\right)_{t,T} = \frac{\sum_{i=1}^{T}(r_{B/A,t+i} - \overline{r}_{B/A})(r_{C/A,t+i} - \overline{r}_{C/A})}{\sqrt{\sum_{i=1}^{T}(r_{B/A,t+i} - \overline{r}_{B/A})^2}\sqrt{\sum_{i=1}^{T}(r_{C/A,t+i} - \overline{r}_{C/A})^2}} \tag{5.17}$$

where $\overline{r}_{B/A}$ and $\overline{r}_{C/A}$ denotes the sample means of the conditional returns.

Besides, based on (5.16) we can also, by exploiting a result from Anderson *et al.,* (2000), approximate the realized correlation. Since our aim here is to analyze the correlation forecasts over the next one-week, next one-month and next three-months, for all three horizons we can use daily observations as the high frequency resource to calculate realized volatility of lower frequency (weekly, monthly and quarterly). Thus, $\tilde{\rho}_{RC}\left(r_{B/A}, r_{C/A}\right)_{t,T}$ can be steadily computed by

$$\tilde{\rho}_{RC}\left(r_{B/A}, r_{C/A}\right)_{t,T} = \frac{\tilde{\Sigma}_{B/A,t,T} + \tilde{\Sigma}_{C/A,t,T} - \tilde{\Sigma}_{B/C,t,T}}{2(\tilde{\Sigma}_{B/C,t,T} \cdot \tilde{\Sigma}_{C/A,t,T})^{1/2}} \tag{5.18}$$

after Σ_{IV} in (5.15) is replaced with $\tilde{\Sigma}_t$ in (5.16). $\tilde{\Sigma}_{(B/A),t,T}$ here denotes the realized volatility of B/A estimated at *t* with forecast horizon *T*. It can be calculated by firstly taking the daily return of B/A, squaring them and then summing them over the relevant (one-week, one-month and three-month) horizons.[51]

Given equation (5.17) and (5.18), it is easy to note that both equations are now utilizing daily observations to approximate realized correlation over a future period. However, the theoretical foundations they are based on are slightly different. The first uses forward-looking historical correlation model; thus an approximation can be made only after returns of future days are known today. However, by exploiting the portfolio theory, the second method uses the past returns to approximate the realized correlation of today. Although both methods assume innovations are to be multivariate Gaussian-distributed, empirically only the second has been examined in literature for approximating 'true correlation'. Therefore, "which one is the better", we believe, is still an empirical question worth further study. As Anderson *et al.,* (2000, p21) pointed out, "*…it is not necessarily the case these two measures will give the same model rankings, let alone the same values of the error measures….*" To circumvent the potential bias, in this paper we only use the second approach to compute realized correlation.

5.4.2 Forecast Evaluation

Once realized correlations have been computed, a major task is then to evaluate forecasts generated from various models with respect to this benchmark using some specific criteria. Here, we carry out three statistical assessments. First, partial optimality of individual forecast is examined for all competing models. Then, cross-pair and cross-horizon forecasting performance are investigated. Finally, forecast combination is also studied along with the analysis of incremental information contributed by each correlation model.

a.1 Partial Optimality test

First, we examine the partial optimality of each correlation forecast. Theoretically, if a forecast is partially optimal, the distance of this forecast to its true value (also called forecast errors)

[51] It is important to note that, in the real OTC market, the maturity of option contract is determined by the realized calendar day. Therefore, these forecast horizons should be empirically different depending on the exact month we are investigating. However, to ease the computation, we omit this variation and assign all horizons a fixed time period. Thus, one week equals 7 days, one month and three months respectively correspond to 22 days and 65 days. Thus, the forecast horizon *T* in all the above equations is now either 7 or 22 or 65. And the *F* in (2.16) also has the same values.

should be unpredictable with respect to the information set used to generate them.[52] To examine this feature, two methods can usually be adopted. One is to perform Mincer and Zarnowitz's (1969) regression test. For example, as applied in this paper, when the realized correlation of two innovations, say $\tilde{\rho}_{RC}(\varepsilon_i\varepsilon_j)_{t,T}$, has been calculated, to examine the optimality of forecasts generated by n^{th} correlation model, we only need to regress $\tilde{\rho}_{RC}(\varepsilon_i\varepsilon_j)_{t,T}$ on $\hat{\rho}_n(\varepsilon_i\varepsilon_j)_{t,T}$

$$\tilde{\rho}_{RC}(\varepsilon_i\varepsilon_j)_{t,T} = a + \beta\hat{\rho}_n(\varepsilon_i\varepsilon_j)_{t,T} + \xi_t \tag{5.19}$$

and test whether $(\alpha,\beta)=(0,1)$.[53] If the null cannot be rejected, it equals to saying n^{th} correlation model has partially optimally exploited the given information set and there is no further information extractable from past information set Ω_t to generate a better forecast than $\hat{\rho}_n(\varepsilon_i\varepsilon_j)_{t,T}$.

Besides, we can also use the sign test of Campell and Dufour (1991, 1995) to perform the same task. The advantage of using this method is that it can release the normality assumption required in the previous regression test so that optimality can be examined in a distribution-free environment. Hence we only need to calculate one statistic to test the null. That is

$$S_{\perp} = \sum_{t=1}^{T} I_{+}(e_{n,t,T}\hat{\rho}_n(\varepsilon_i\varepsilon_j)_{t,T}) \tag{5.20}$$

where $e_{n,t,T} = \hat{\rho}_n(\varepsilon_i\varepsilon_j)_{t,T} - \tilde{\rho}_{RC}(\varepsilon_i\varepsilon_j)_{t,T}$ denotes the forecast error generated by n^{th} correlation model for forecast horizon T; I_+ equals one if $e_{n,t,T}\hat{\rho}_n(\varepsilon_i\varepsilon_j)_{t,T} = 0$ and zero otherwise; and $e_{n,t,T}\hat{\rho}_n(\varepsilon_i\varepsilon_j)_{t,T}$ represents an orthogonal function of forecast error with respect to past information set since $\hat{\rho}_n(\varepsilon_i\varepsilon_j)_{t,T}$ now can be regarded as a reflection of Ω_t .[54] The motivation for proposing this test here is that, if, for example, a forecast is optimal, its forecast error would then be orthogonal to Ω_t . Thus, for our cases, if n^{th} correlation model is being examined, the null that needs to be tested is then either $\text{cov}[e_{n,t,T}\hat{\rho}_{n,t,T}] = 0$ or $E_t[e_{n,t,T}\hat{\rho}_{n,t,T}] = 0$.

[52] Information used to generate the correlation forecasts is never based on a whole information set. Only a subset of the whole has been utilized. This is because all scientific models are exploiting only an incomplete information set, thus we can only term the resulting optimality of forecasts as partial optimal.

[53] To account for the heteroskedasticity and autocorrelation that may appear during the regression, we use Newey and West's correction methods to adjust the regression process.

[54] Sometimes, this test is also referred to as the rational expectation test in some literature (see Brown and Maital, 1981, and Diebold and Lopez, 1996, for detailed illustration).

a.2 Comparison among competing forecasts

In this research, apart from the partial optimality, correlation forecasts generated from different models are also compared using three statistical loss functions: MFE (mean forecast error), MAE (mean absolute error) and MSE (mean square error). Specifically, MFE, once calculated, is used to perform an unbiasedness test by regressing the forecast error on a constant. If the coefficient of this constant is found to be insignificantly different from zero, then we say the forecast is unbiased with respect to the true correlation. However, when MAE and MSE are used, forecast errors are then penalized differently, but symmetrically for each model. Since a quadratic function is used in MSE, large forecast errors are weighted more heavily compared to MAE in which only absolute term for forecast error is used.

b. Cross-horizon Comparison (GMSFEM test)

The evaluation method, illustrated above, is usually applied to answer a question like 'for a given forecast horizon, is the correlation generated by one model more informative than and superior to others in approximating realized correlation?' However, since multiple horizons are investigated in this paper, it is then also interesting to see whether this superior performance, if confirmed, is consistent when different forecast horizons are analyzed. To assess the cross-horizon forecasting performance between different models, Newbold, Harvey and Leybourne (1999) proposed the Generalized Mean Square Forecast Error Matrix test, or GMSFEM. Specifically, first we calculate the vector of forecast errors of, say, model A and model B, for all forecast horizons up to *T*, that is $E(A)^{'}_{t,T} = (e^{A}_{t+1}, e^{A}_{t+2}, ..., e^{A}_{t+T})$ and $E(B)^{'}_{t,T} = (e^{B}_{t+1}, e^{B}_{t+2}, ..., e^{B}_{t+T})$. Then, their second moments are respectively computed using $\Phi_{AT} = E_t[E(A)_{t,T} E(A)^{'}_{t,T}]$ and $\Phi_{BT} = E_t[E(B)_{t,T} E(B)^{'}_{t,T}]$. If either the condition $d^{'}\Phi_{AT}d \leq d^{'}\Phi_{BT}d$ or $d^{'}(\Phi_{AT} - \Phi_{BT})d \leq 0$ is now satisfied for at least one vector $d^{'} = (d_1, d_2, ..., d_T), \forall d \neq 0$, then we say the forecast generated by model A dominates model B across horizon from one step ahead to *T* step ahead. Simply put, if the eigenvalue of $(\Phi_{AT} - \Phi_{BT})$ is all non-positive with at least one negative, then model A dominates B; however, if the eigenvalue is all non-negative with at least one positive, then the latter model is preferred. Indeterminacy will be encountered when both positive and negative appear in the same set. In this paper, since three forecast horizons are analyzed, the forecasts error to be examined for n^{th} correlation model is then $E(n)^{'}_{t,T} = (e^{n}_{1Week}, e^{n}_{1Month}, e^{n}_{3Month})$.

c. Encompassing test

Besides, in this paper, since the correlation forecasts are generated either through a time series model using historical return as input or through option prices, another interesting topic worth investigating is ‘whether the combination of these different information sources will lead to an improved forecasting accuracy’. Here, we use the encompassing test to examine this forecast combination after information aggregation. To ease the expression, as illustrated already, we divide different models into three groups: simple historical correlation models, GARCH-type models and implied correlation model. For the first two groups, only the forecasts which have demonstrated the highest explanation power in the previous partial optimality test will be incorporated to the current regression. Thus, this test is formed as

$$\tilde{\rho}_{RC}(\varepsilon_i\varepsilon_j)_{t,T} = \alpha + \beta_1\hat{\rho}_1(\varepsilon_i\varepsilon_j)_{t,T} + \beta_2\rho_2(\varepsilon_i\varepsilon_j)_{t,T} + \beta_3\rho_3(\varepsilon_i\varepsilon_j)_{t,T} + \xi_t \tag{5.21}$$

where $\hat{\rho}_1(\varepsilon_i\varepsilon_j)_{t,T}$, $\hat{\rho}_2(\varepsilon_i\varepsilon_j)_{t,T}$ are correlation forecasts of the ‘best’ performing models selected from simple historical correlation group and GARCH family with maximal R^2 in the partial optimality regression test; $\hat{\rho}_3(\varepsilon_i\varepsilon_j)_{t,T}$ represents the implied correlation forecast for realized correlation $\tilde{\rho}_{RC}(\varepsilon_i\varepsilon_j)_{t,T}$. Here, to examine the encompassing effect, we perform three hypothesis tests on coefficient of constant and three independent variables. Specifically, first we test whether $(\alpha,\beta_1,\beta_2,\beta_3)=(0,0,0,1)$. If this null cannot be rejected, implied correlation model is then said to be able to forecast encompass GARCH-series model and historical correlation model.[55] Second, we test whether β_2 as an individual parameter is insignificantly different from zero to see the forecasting performance of correlations generated by using GARCH-series model. Finally, we also examine correlation generated by time series tools as a group; thus the null to be tested is $\beta_1 = \beta_2 = 0$.

5.5 Data and Empirical results

5.5.1 Spot returns and option data

In order to examine the forecasting performance of eleven correlation models, in this chapter we analyze the daily return of two currency trios: EU/US/UK (GBP trio) and EU/US/JP (Yen trio). A total of six currency pairs are derived from these trios and we collected their data from *DataStream* with a span of six years starting from 1999/1/1 to 2005/5/31. After eliminating the official holidays such as Christmas and Easter, a total of 1621 observations are obtained in our sample for each pair. Since GARCH model is to be used for forecasting future correlation, we let the first 1000 observations be the in-sample set to ensure the asymptotic property of its

[55] Here, it is important to note that finding no evidence of encompassing is usually not a surprise. This is because the correct mapping to the ‘true correlation’ is still under investigation; thus we cannot only rely on the incomplete information to find a single ‘better-than-all-others’ model.

estimation. Therefore, the remaining 621, starting from 2002/11/4, are used to obtain correlation forecasts.

As for the daily implied volatility data, we collected them from BBA-Reuters FX option volatility index with quotes for one-week, one-month and three-month respectively.[56] Calculation of these implied volatilities is based on an ATM forward straddle pricing model and the data is generated from 2001/10/1 to 2005/5/31. After removing the official holidays, we find some missing data in our resultant sample.[57] To fill this information gap, a linear interpolation technique suggested in Dennis *el al*. (2005) is then applied.

Here, before proceeding further, another thing needs to be noted. Early researches into similar area, such as Camp and Chang (1998), Walter and Lopez (2000) and Chong (2001), in their samples unanimously used the volatility data collected from a single market participant to calculate implied correlation in FX market. However, as for our cases, the BBA option volatility is actually an index averaging daily quotes obtained from 12 different market participants. Since a broader group is now incorporated, it is reasonable to expect that using this information source can provide a more extensive and integrated market view to accurately forecast the future correlation.

5.5.2 Empirical Results

5.5.2.1 Summary statistics

a. Implied correlation

Table 5.1 Panel A presents the descriptive statistics of implied correlation for two currency trios and Figure 1 (panel A and panel B) shows their corresponding time series plots. For pairs in both trios, it is now evident that implied correlations present different types of dynamics although, for the same pair, the multi-horizon performances are rather similar. For instance, the sample means of implied correlations in EUR/USD/JPY trio, for all three forecast horizons, are around 0.5, whilst those of EUR/USD/GBP trio range from 0.17 to 0.74. Negative skewness is observed in most of the cases except for currency pair (USD/EUR and JPY/EUR). And kurtosis estimates show these correlations have thinner tails than Gaussian. Meanwhile, for the time series plot, a clear pattern is that implied correlation tends to be more stable as the forecast horizon becomes longer. For example, standard deviation of implied correlation

[56] *BBA-Reuters FX Option Volatility Index* was officially co-launched by *British Bank Association* (BBA) and *Reuters* on December 31st 1997. The initial motivation of quoting these data is to improve the market transparency by enhancing the quality and accessibility of independent valuations. Quotes on 13 currency pairs have been generated on a daily fixing since August 2001.

[57] In the out-of-sample set, there are a total of 26 missing data in the implied volatility. As cited from the BBA, "if there are fewer than 5 rates received by the contributors, then the benchmark will not be published."

between EUR/USD and JPY/USD for the next three months, $\rho_{IC}(\varepsilon_{EU/US}, \varepsilon_{JP/.US})_{t,3m}$, is now less than half of the same statistic calculated for the same correlation over the next week. This result is as expected because long-term correlation is usually believed to be more persistent than short-term correlation.

(Insert Table 5.1 Panel A and Figure 1 panel A and Panel B)

b. Realized correlation

Similar statistics are also calculated for realized correlation, and the results are presented in Table 5.1 Panel B. Here, note that, for the same pair, sample means of most realized correlations are very close to the means reported for implied correlation, although their conditional second moments differ a lot. It is clear that realized correlation is now following a dynamic process much more volatile than implied correlation. For example, *s.t.d* estimates of, say, $\rho(\varepsilon_{EU/US}, \varepsilon_{JP/US})_{t,3m}$ has risen from the previous value of 0.05 (for implied correlation) to the current value of 0.12 (for realized correlation) and this feature becomes even clearer when short-term correlation such as realized correlations over the next week are analyzed e.g., volatility of $\rho(\varepsilon_{EU/US}, \varepsilon_{JP/US})_{t,1w}$ is approximately 0.33.

For different forecast horizons, the same as before, long-term correlation appears more stable than short-term correlation. Evidence for this argument can be found in Figure 2 where kernel density plots of various realized correlations are provided. As can be seen, for both trios although the density shapes of three-month realized correlation when compared to that of the one-week correlation are now about the same, their evolving processes appear more central to the means with relatively higher peaks. This result is no surprising because most researches contributing to understand the correlation evolving process have already found this coefficient very stable if the time frame for analysis is set sufficient large. Therefore, it is usually expected that long term correlation would be much easier to forecast than short term correlation and a large distance between these two estimates may leads to trading opportunity. For example, for an experienced trader who is specialized in long-short pairs trading, if he only wants to take advantage of the market inefficiency but not results based on the fundamental changes, a common strategy is then to long a stock A and simultaneously short another related stock B. Here, to what extent these two stocks are related to each other can be explained and quantified using a specific correlation model. However, one thing needs to noted is, to make profit, these two positions are usually required to be taken at the time when short term correlation is significantly different from long term correlation or when the price difference of two stocks reach an abnormal level, suggesting that the pegged relationship between two stocks now may have been temporary broken. Since correlation itself tends to show mean-reverting characteristic, it is then expected that after some period of time short term correlation will approach the long term correlation again and this price difference will return to the normal state.

Besides, with respect to their density plots, there is another interesting finding worth noting here. That is, characteristics of mixture distribution (multi-modality) are observed and it is especially the case for long term correlation. For example, if we look at the density plot of correlation forecast over the next three months between USD/EUR and GBP/EUR, it can be easily seen that two modes are now simultaneously appearing in one conditional distribution and their values are far apart. This feature has important implications for financial researchers and fund managers because it reveals the fact that the market is now forming diverged opinions on how future correlation will move. One group of the investors is now maintaining their traditional view that the correlation will stay at 0.5, the same as previously, even three months later. However, another group of people are then expecting this correlation to rise to 0.7. As a researcher, to identify this sign of divergence as soon as possible is very important and beneficially. Correct interpretation of this feature will leads to more accurate understanding of the correlation evolving process. Meanwhile, another things needs to be noted is even if this divergence of market opinion occur, usually this feature is more easily to be reflected in the plot of long term correlation than the short term correlation. This is because, if investors are now only asked to forecast the correlation, say, for the next one or two days, and given that there is no significant evidence of asymmetric information, then it is reasonable to say that there will not be much difference among their expectations for the future correlation. Put it in another way, even if investors now do have the diverged expectation, it is very difficult for this feature to be sufficiently exploited in the short term and flexibly reflected in a distributional form if our forecast of interest is only the correlation of a few days later. Therefore, in density of short term correlation, usually one can only observe one peak along with negative skewness. However, this is not to say this asymmetry (negative skewness) then cannot be generated from a mixture distribution, because, as has been proved, a proper mixing strategy could also lead to unimodality. Therefore, it is implied that, by incorporating another correlation dynamics to the current framework, the realized correlation, especially for those concerning a long forecast horizon, can be more flexibly modeled and accurately forecast. For a more detailed illustration of how to develop such a new framework to capture ‘correlation mixture’, we dedicate all the remaining chapters to this topic. However, for now, we proceed only by focusing on the task of evaluating various correlation forecasts generated from existing time series tools and implied correlation model.

(Insert Table 5.1 Panel B and Figure 2 Panel A and Panel B)

5.5.2.2 Statistical evaluation of forecast error

After eleven correlation models are fitted using empirical data, statistical evaluation results of their out-of-sample forecasting performance are reported in Table 5.2 panel A and panel B. Here, for both trios a common finding is the predicative accuracy of these forecasts in approximating their corresponding realized correlation is now found very sensitive to the statistical loss function used to evaluate them. Only a historical correlation model can consistently produce unbiased estimates for realized correlation, although the forecasting performance of implied correlation, especially in the JPY trio, is also worth mentioning here.

a. Evaluation based on MFE

MFE results show that sophisticated forecasts generated from multivariate GARCH models and implied correlation model are more inclined to introduce biases than historical correlation models. The conditional means of their forecast errors are frequently found to be significantly different from zero. For instance, when the realized correlation $\rho(\varepsilon_{US/EU}\varepsilon_{UK/EU})$ is to be forecast, for all three horizons, forecast errors generated using GARCH models, either univariate or multivariate, are found to be conditionally biased. Although implied correlations here can provide a slightly better performance, in half of a total of 18 cases the expectations of their forecast errors are also confirmed as significantly different from zero.

To provide a plausible explanation for such a massive number of biased estimates, it is worthwhile to start from their model misspecifications and the stringent conditions assumed in their mechanisms for generating forecasts (see section 5.3.4). For example, conditional bias of implied correlation is not surprising because the implied volatility, from which these correlation forecasts are calculated, is already frequently found to be biased (see Jorion, 1995). Unjustified assumptions such as constant volatility and normal distribution are usually penalized as the potential reasons for causing its bias. However, recently, researchers have suggested other possibilities. For example, after studying the sample selection bias in S&P500 index option, Engle and Rosenberg (2000) attributed the conditional bias found in implied volatility to the testing procedure. Similarly, Christensen *et al.,* (2001) argued that the overlapping observations may also be a problem.

Here, an interesting finding is that when some specific correlation such as $\rho(\varepsilon_{EU/UK}\varepsilon_{US/UK})$ becomes the target of forecasting, MFEs of GARCH models tend to be smaller for short-horizon forecasts, suggesting that this model may be more useful in producing unbiased forecasts for short-term correlation than for long-term correlation. Although empirical research on volatility forecasting has already found many analogous results, i.e., the one-step-ahead GARCH volatility is usually found to be more accurate than multi-step-ahead forecast since the latter is closer to 'static' unconditional volatility; similar evidence on correlation forecasting is not very consistent.

On average, under MFE it is the simple forecasts generated from either a historical correlation model or EWMA that can most frequently produce the best. This result is as expected and has been confirmed by other researchers as well. For example, as Walter and Lopez (2000, p33) illustrated, "*… the simple correlation forecasts always approximate the unconditional correlation of the series by using a sub-sample of the available data… thus, the small MFE of their forecasts are not surprising ...*"

(Insert Table 5.2 panel A and panel B)

b. Evaluation based on MAE and MSE

When MAE and MSE results are analyzed, a different picture is presented. Sophisticated forecasts, especially the implied correlation, now present a much closer relationship than other alternatives to realized correlation. In 16 out of 18 cases, the forecasts derived from the option prices successfully generate the lowest MSE, and in 13 out of 18 cases they generate the lowest MAE. Besides, the performance of GARCH-based forecast in approximating realized correlation has also improved a great deal with more evidence showing only small biases in its resulting estimates. Moreover, it is noticeable here that MSEs and MAEs tend to be lower for long-term correlation forecasts than for short-term correlation forecasts. For example, in Panel A the MSE of $\rho(\varepsilon_{US/EU}\varepsilon_{UK/EU})_{t,1w}$ generated from DCC is 0.1104, whilst the same estimates for one-month and three-month forecasts are only 0.042 and 0.022. This result suggests that correlation forecasts tend to be more accurate when they are used in a long forecast horizon, possibly reflecting the reversion of the dynamics to unconditional correlation.[58]

(Insert Table 5.2 Panel A and Panel B)

Besides, for comparing predictive accuracy, in this chapter we also perform Diebold Mariano test to discriminate models which have generated similar MAE values and similar MSE values. Specifically, since for each correlated pairs we have identified a best model under these two

[58] Correlation, just like volatility, is usually modeled as a mean reverting process. In the short run, its evolving process may present a volatile style along with some jumps. However, in the long run the reversion of its sample paths to the unconditional mean is usually evident. Sometimes, even if some structural changes are observed, this reversion pattern will still be sustained with mean adjusted to a new level.

loss functions, the new test is then performed to examine whether any other models can provide a statistically equal performance to the best model. Concerning this result, now it can be easily seen from Table 5.2 that the advantage of implied correlation for forecasting long-term correlation is actually very evident. In no cases, forecasts generated by other methods in two trios can provide a statistically equal standing as implied correlation model. However, the thing does change a little bit when short-term correlation becomes the target of forecasting. For example, in the JPY trio if we are going to predict the correlation between EU/JP and US/JP for the next week, the best model is then EWMA and in no case this model can be statistically outperformed by others under MSE and MAE. Besides, another thing needs to be noted is multivariate GARCH models sometimes obtain similar prediction power to implied correlation model. However, its performance is not consistent as forecasts generated from derivative markets.

5.5.2.3 Forecast Optimality results

a. Partial optimality regression results

Now, we proceed to illustrate the partial optimality regression results for two currency trios in Table 5.3 Panel A and Panel B. First, concerning the simple historical correlation models, their forecasting performances in two trios are rather similar. Individual hypothesis (either a=0 or b=1) and joint hypothesis (a=0 and b=1) for partial optimality are consistently violated.[59] EWMA gives on average the best performance among simple forecasts with 12 out of 18 highest R^2 derived from it. This result is not surprising because the past price history it includes is already known to be the longest of all. Although they can produce unbiased forecasts, historical correlation models perform badly when they are used to explain the variation in realized correlation. The maximum R^2 these models can generate in GBP trio are less than 0.05.

With regard to the sophisticated forecasts, a mixed picture is presented concerning their performances in two trios. As for the GARCH models, joint hypothesis of optimality is rejected in all cases. Of 108 GARCH forecasts examined in GBP trio, only 7 null hypotheses of individual optimality (either a=0 or b=1) are not rejected. And, interestingly, this evidence unanimously supports the superiority of CCC model where correlation is modelled as a constant rather than a dynamic. In the Yen trio, the usefulness of BEKK and DCC models is confirmed several times although, as before, in the majority of cases null of optimality is again rejected either individually or jointly. As far as the explanation power is concerned, there is no

[59] Simple historical correlation models include His-7, His-22 and His-65, which are historical correlation models respectively using past 7 day, past 22 days and past 65 days observations as input, and EWMA model.

evidence to show that GARCH model can provide an improved performance compared to the simple forecasts just illustrated. However, a slightly better performance can be confirmed when implied correlations are regressed on realized correlation. For example, in GBP trio half of the highest R^2 is obtained using its forecasts and, in 4 out of 18 cases examined, both individual and joint hypothesis for partial optimality is accepted. However, we should note that, in most cases, the explanatory power of this model is still not as high as we expected.

Concerning this issue, two things need to be noted. First, in this research low R^2 is very easy to generate when the task is to forecast short-term correlations. This result is as expected because the correlation forecasts usually tend to be more accurate for the longer horizons. Second, correlation estimated on two assets whose liquidity is not strong enough could also lead to low R^2 in forecasting realized correlation. For example, in the GBP trio, EUR/GBP and USD/GBP are two currency pairs much less traded than EUR/USD in the FX market; thus it is reasonable to expect that the realized correlation of $\rho(\varepsilon_{EU/UK}\varepsilon_{US/UK})$ will be harder to predict than either $\rho(\varepsilon_{EU/US}\varepsilon_{UK/US})$ or $\rho(\varepsilon_{UK/EU}\varepsilon_{US/EU})$. As can be seen from Table 5.3 Panel A, all models used to forecast $\rho(\varepsilon_{EU/UK}\varepsilon_{US/UK})$ can only generate R^2s lower than 0.10. After all, compared to the similar research studies performed for volatility forecasting, the low explanation power found in our cases for forecasting realized correlation is not totally surprising. For example, the R^2 for the regression of realized volatility on a constant and implied volatility, according to Jorion (1995), ranges from only 0.02 to 0.05. And this value also hardly exceeded 0.10 in Guo (1996) for forecasting USD/JPY volatility.

(Insert Table 5.3 panel A and B)

b. Sign test result

As a complement to the partial optimality regression, a non-parametric two-sided sign test is also performed in this paper to examine whether the covariance of forecast error and forecast itself has zero expected value.[60] We report its result in Table 5.4. For both trios, the evidence of partial optimality is now much more pronounced than previously. Not only are sophisticated forecasts observed with more cases of accepting the null hypothesis $E_t[e_{t,T}\hat{\rho}_{t,T}]=0$, the simple forecasts have also shown much improvement in exploiting the past information. For

[60] Here, the purpose of this test is to examine whether the forecast error is orthogonal to the past information. Therefore, the function to be analyzed is $e_{t,T}\hat{\rho}_{t,T}$, and we examine it by testing whether its expected value is zero. One thing to note here is that we use median instead of mean to approximate this expected value. This is because the forecast error generated in this paper frequently presents different degrees of asymmetry, and in such cases, median is usually regarded as a more reliable statistic than mean to describe the whole distribution through a single estimate. Thus, the sign test performed here is to examine whether the median of $e_{t,T}\hat{\rho}_{t,T}$ is zero.

example, of the 18 cases examined in two trios, only 3 forecasts generated from the HIS-22 model fail to pass the test.[61] This result is a little bit surprising because it is significantly different from those reported in Table 5.4 where none of the simple forecasts is found to be partially optimal. However, the explanation is not too difficult. For example, we can attribute this discrepancy to the more stringent conditions required in the regression test to confirm optimality than those required in the sign tests.

For the GARCH and implied correlation model, the improved performance when compared to the previous result is also very significant. In 11 out of 18 cases, the forecasts generated from the option prices are proved to be partially optimal. Besides, the simplified univariate Normal-GARCH, GED-GARCH and DCC models are also frequently able to fulfil the orthogonal condition. However, as noted, their performances are not uniform across the currency trios examined and across the forecast horizons of interest.

(Insert Table 5.4)

5.5.2.4 Encompassing Regression results

In Table 5.5, we report the encompassing regression results for two currency trios. To improve the forecasting accuracy, we combine the forecasts generated from three different correlation-modelling groups. They are implied correlation, simple historical correlation and GARCH-based correlation. To circumvent the potential multicollinearity among forecasts that are derived from similar modeling structures, in the latter two groups only those which have demonstrated the highest R^2 in the previous partial optimality results are incorporated into the current regression. Therefore, the realized correlation is now regressed on a constant and three different forecasts.

(Insert Table 5.5)

Since a forecast combination technique is adopted, R^2 improves a lot for both trios, suggesting that aggregation of historical information and option-driven information can create a more accurate correlation forecast. The usefulness of implied correlations in forecasting realized correlations is found not only significant but also consistent. However, when time series forecasts are analyzed, their cross-trio and cross-horizon performance is not uniform. Although GARCH-based forecasts, in some cases, are found containing valuable information, coefficients of their forecasts in regression are either insignificantly different from zero or negative. Besides, this situation also applies to simple historical correlation when the Yen trio

[61] HIS-22 refers to the historical correlation model where the past history (*P*) equals 22 .

is investigated. To illustrate these results with more detailed evidences, next we summarize several typical features found in Table 5.5 and present them in the following.

a. GBP trio

First, for GBP trio, the implied correlation is found contributing to the forecast of realized correlation on a consistent basis. In all nine cases examined, the coefficients of their forecasts are all positive and significantly different from zero. Besides, while three special correlations $\rho(\varepsilon_{US/EU}, \varepsilon_{UK/EU})_{t,1w}$, $\rho(\varepsilon_{US/EU}, \varepsilon_{UK/EU})_{t,1m}$ and $\rho(\varepsilon_{EU/US}, \varepsilon_{UK/US})_{t,1w}$ are predicted, implied correlation forecasting encompasses all other historical information-based estimates. According to the Wald test result presented at the bottom of Panel A, we cannot reject the null hypothesis that the regression coefficients of time series forecasts are all zero, suggesting that implied correlation now fully incorporates all information extractable from the time series data. However, as for the others, correlation forecasts generated from GARCH models or simple historical correlation models convey important information that is not presented in option prices.

Here, it is important to note one thing. Among the encompassing evidence we have just reported, the forecast horizons of interest are all relatively short (two one-week forecasts and one one-month forecast). Therefore, it is fair to say that the implied correlations maybe more useful in predicting short-term correlation. For example, as we may usually expect, news such as immediate interest rate changes and long-term currency reform are empirically much easier to be predicted and reflected in option prices with short maturity rather than using trend-focused time series model. This is because the former is more sensitive to the temporary changes in market expectation, whilst the latter more focuses on the value-tracking given that the time period analyzed is long enough. However, this is not to say implied correlation can then always explain the variation of realized correlation very well. As can be seen from Figure 3, Panel A, the dynamics followed by short-term realized correlation such as $\rho(\varepsilon_{EU/US}, \varepsilon_{UK/US})_{t,1w}$ are very bumpy and volatile. As a comparison, the implied correlation used to forecast it seems much more stable. Therefore, we can only draw the conclusion that the implied correlation in some cases may be a better forecast than historical information-based estimates for predicting short-term realized correlation. However, how well it really performs is still an empirical question.

b. Yen trio

With respect to the Yen trio, GARCH-based forecast now improves a lot in terms of the information contribution it can make to predict the realized correlation. All coefficients of its forecasts are significantly different from zero, suggesting that they contain information not presented in other forecasts. However, we should note here that these coefficients are sometime found negative, making the explanation a very difficult task. As for those generated from other historical information-based models, similar findings are also observed. Not only are coefficients often found non-positive, evidence of insignificance is also found several times, suggesting that these forecasts conveys no incremental information for forecasting future realized correlation.

Although the regression results do not create a uniform picture for the usefulness of simple forecasts, the performance of implied correlation is consistent for both trios. Of the nine cases reported in Table 5.5, Panel B, only one coefficient of implied correlation fails to reject the null hypothesis that is significantly different from zero. Although the Wald test results now suggest no evidence for encompassing, it is not a surprising result here since the information contribution conveyed by competing GARCH-based forecasts has already improved a lot.

c. Other features

Meanwhile, it is also worth noting some interesting findings presented in Table 5.5. For example, a typical feature here is to favour the CCC model among GARCH variants. As can be confirmed from Panel A and Panel B, this model has been selected seven times as the representative of GARCH models to generate a forecast for realized correlation. Although its advantage in generating higher explanatory power than other GARCH-variants is only marginal, on average this model is still statistically the best. Besides, we also find that the realized correlation can be more accurately predicted in the long term. After forecast combination, R^2 of encompassing regressions increase a great deal compared to those documented in Table 5.4. Predicting long-term correlation (correlation over the next one month and next three months) makes it easier to generate a higher quality of fit than predicting short-term correlation (correlation over the next week). The only exception is for $\rho(\varepsilon_{US/EU}\varepsilon_{JP/EU})$ where the regression on one-month correlation yields 0.305 R^2 whilst three-month correlation only generates 0.241 R^2. To obtain a clearer view, we present in Figure.3, Panel A and Panel B, the time series plot of realized correlation of $\rho(\varepsilon_{EU/US}\varepsilon_{UK/US})$ and $\rho(\varepsilon_{EU/US}\varepsilon_{EU/JP})$ and various forecasts used to predict them.

(Insert Figure 3 Panel A and Panel B)

5.5.2.5 GMSFEM Test results

a. Three-horizon comparison

Apart from the cross-trio evaluation, comparing various models' cross-horizon performance is also an interest of this paper. To perform this task, we firstly calculate the GMSFEM ranking statistics for correlation forecasts of all three horizons (one week, one month and three month). Surprisingly, of all 224 comparisons analyzed, only one case displays the evidence of consistent cross-horizon performance. Even more surprisingly, this case is not generated by implied correlation but from the GARCH family. BEKK model in the forecasting of realized correlation $\rho(\varepsilon_{EU/UK}\varepsilon_{US/UK})$ shows a consistent out-performance over GED estimates across all three horizons of interest. The eigenvalues derived for this comparison are all positive (0.034163, 0.6837, 13.155). However, as for others, mixed sign results are then generated, suggesting that there is insufficient evidence to determine the superiority of one over another consistently.

Although this result is a little bit unexpected, it is not totally inexplicable. For example, as can be confirmed from Figure 3, short-term realized correlation follows a dynamic process that is much more volatile than long-term correlation. Given this feature, one then might want to argue that the best models used to depict these two processes should be intrinsically different. Since the models that perform well in predicting short-term correlation are now probably not the ones which perform well in forecasting long-term correlation, the potentially inconsistent cross-horizon forecasting performance is then not surprising for the correlation dynamics.

b. Two-horizon comparison

In order to launch a further analysis, we combine the correlation forecasts over the next one-month and the next three-month into a new category and re-perform the GMSFEM test. In Table 5.6, we report its result. Clearly, the evidence of cross-horizon out-performance is now much more pronounced. For the two trios examined, there are a total of 39 cases confirming the domination of one forecast over the other, across two horizons. Simple historical forecasts are the poorest among all competing forecasts. In no cases forecasts of this group are found capable of dominating others. However, for GARCH models, significant out-performances are then consistently found both within and across the forecasting groups. For example, in the GBP trio DCC forecasts present three cases of out-performance over historical correlation. When univariate GARCH-Normal and GARCH-GED are examined, they are then dominated by other multivariate GARCH variants such as BEKK, VECH in $\rho(\varepsilon_{EU/UK}\varepsilon_{US/UK})$. In respect of the Yen trio, similar evidence is also documented. Besides, it is especially worth noting the overwhelming domination of implied correlation over all other historical information-based forecasts when realized correlation $\rho(\varepsilon_{EU/US}\varepsilon_{EU/JP})$ is predicted. Here, the eigenvalues derived from ten comparisons are all found positive, suggesting that implied correlation is now favoured under GMSFEM criterion over all its alternatives across two horizons. However, such significant preference is only observed once in all cases.

(Insert Table 5.6)

5.6 Summary

In this chapter, we examine the forecasting performance of eleven correlation models in predicting realized correlation. After contributing to the current literature in three aspects, our findings suggest that the best model to forecast future correlation is very sensitive to the loss functions used to evaluate them. Implied correlation can convey valuable information on a consistent basis but its cross-horizon performance is not uniform. GARCH-based forecasts sometimes contain incremental information not included in the option prices. However, its advantage of capturing the time-varying characteristics of correlation dynamics is not fully confirmed in our research because the most favoured model among GARCH variants is actually the CCC which assumes correlation to be fixed. This is probably because the level and direction of realized correlation change just too markedly in our samples. After performing the encompassing test, we find that the combination of historical information source and option-derived information source can produce a more accurate correlation forecast than any single technique in terms of improved explanatory power. And it is easier to accurately forecast the long-term correlation than the short-term correlation.

Meanwhile, it is also worth noting another interesting finding of this paper. The kernel density estimate of the realized correlation is frequently found to be showing multi-modality, suggesting that, by adopting a mixture modelling technique, the flexibility of capturing various characteristics presented in the correlation dynamics can be extended to a further degree. This application may also contribute to the generation of a more accurate forecast than other traditional tools in predicting realized correlation. To fully exploit this implication, we now devote the next part of this thesis to the development of two conditional heteroskedastic correlation mixture models. We start by presenting some elementary information concerning the mixture distribution (models) and numerical algorithms which can be used to estimate them.

Table 5.1 Summary statistics of implied- and realized- correlation for two currency trios

Panel A: Implied Correlation

EU/US/UK (GBP trio)

	Mean	Std	Skewness	Kurtosis	Min	Max
$\rho_{IC}(\xi_{US/EU}\xi_{UK/EU})_{t,1w}$	0.5124	0.1204	-0.0642	1.9760	0.2485	0.7496
$\rho_{IC}(\xi_{US/EU}\xi_{UK.EU})_{t,1m}$	0.5174	0.0940	-0.0447	1.9102	0.3056	0.6926
$\rho_{IC}(\xi_{US/EU}\xi_{UK/EU})_{t,3m}$	0.5352	0.0889	-0.4904	2.3706	0.3253	0.7059
$\rho_{IC}(\xi_{EU/US}\xi_{UK/US})_{t,1w}$	0.7417	0.0919	-1.0011	3.3921	0.4293	0.8939
$\rho_{IC}(\xi_{EU/US}\xi_{UK/US})_{t,1m}$	0.7433	0.0671	-1.1077	3.5453	0.5287	0.8555
$\rho_{IC}(\xi_{EU/US}\xi_{UK/US})_{t,3m}$	0.7424	0.0473	-0.9093	3.0645	0.6093	0.8235
$\rho_{IC}(\xi_{EU/UK}\xi_{US/UK})_{t,1w}$	0.1793	0.1270	-0.2582	3.2272	-0.2312	0.5376
$\rho_{IC}(\xi_{EU/UK}\xi_{US/UK})_{t,1m}$	0.1784	0.0934	-0.1247	2.8723	-0.0904	0.3983
$\rho_{IC}(\xi_{EU/UK}\xi_{US/UK})_{t,3m}$	0.1620	0.0916	-0.2057	2.4738	-0.0945	0.3348

EU/US/JP (Yen trio)

	Mean	Std	Skewness	Kurtosis	Min	Max
$\rho_{IC}(\xi_{EU/US}\xi_{JP/US})_{t,1w}$	0.5318	0.1190	-0.6657	2.9687	0.1383	0.7268
$\rho_{IC}(\xi_{EU/US}\xi_{JP/US})_{t,1m}$	0.5290	0.0810	-0.5750	2.9393	0.2653	0.6759
$\rho_{IC}(\xi_{EU/US}\xi_{JP/US})_{t,3m}$	0.5313	0.0532	-0.5095	2.8310	0.3744	0.6352
$\rho_{IC}(\xi_{US/EU}\xi_{JP/EU})_{t,1w}$	0.5028	0.1272	0.3250	2.4708	0.1618	0.8395
$\rho_{IC}(\xi_{US/EU}\xi_{JP/EU})_{t,1m}$	0.5214	0.0930	0.4129	2.2162	0.3156	0.7687
$\rho_{IC}(\xi_{US/EU}\xi_{JP/EU})_{t,3m}$	0.5477	0.0628	0.3667	2.3144	0.3637	0.7150
$\rho_{IC}(\xi_{US/JP}\xi_{EU/JP})_{t,1w}$	0.4488	0.1264	-0.5427	2.7281	0.1018	0.7147
$\rho_{IC}(\xi_{US/JP}\xi_{EU/JP})_{t,1m}$	0.4399	0.0964	-0.5006	2.7538	0.1835	0.6695
$\rho_{IC}(\xi_{US/JP}\xi_{EU/JP})_{t,3m}$	0.4138	0.0704	-0.5165	2.8215	0.2165	0.5840

Panel B. :Realized Correlation

EU/US/UK (GBP trio)

	Mean	Std	Skewness	Kurtosis	Min	Max
$\rho_{RC}(\xi_{US/EU}\xi_{UK/EU})_{t,1w}$	0.4666	0.3096	-0.7194	3.1318	-0.6197	0.9541
$\rho_{RC}(\xi_{US/EU}\xi_{UK.EU})_{t,1m}$	0.4837	0.1822	-0.1973	2.8154	-0.2608	0.8717
$\rho_{RC}(\xi_{US/EU}\xi_{UK/EU})_{t,3m}$	0.4818	0.1207	0.3797	2.7661	0.2082	0.7800
$\rho_{RC}(\xi_{EU/US}\xi_{UK/US})_{t,1w}$	0.7218	0.2093	-1.5360	5.7906	-0.2882	0.9977
$\rho_{RC}(\xi_{EU/US}\xi_{UK/US})_{t,1m}$	0.7308	0.1177	-0.7047	2.9818	0.3226	0.9399
$\rho_{RC}(\xi_{EU/US}\xi_{UK/US})_{t,3m}$	0.7396	0.0794	-0.1603	2.1328	0.5361	0.8816
$\rho_{RC}(\xi_{EU/UK}\xi_{US/UK})_{t,1w}$	0.1877	0.3693	-0.3556	2.5043	-0.8368	0.8996
$\rho_{RE}(\xi_{EU/UK}\xi_{US/UK})_{t,1m}$	0.2111	0.2002	-0.1461	2.5875	-0.3321	0.7007
$\rho_{RC}(\xi_{EU/UK}\xi_{US/UK})_{t,3m}$	0.2186	0.0985	-0.2658	2.3152	-0.0469	0.4334

EU/US/JP (Yen trio)

	Mean	Std	Skewness	Kurtosis	Min	Max
$\rho_{RC}(\xi_{EU/US}\xi_{JP/US})_{t,1w}$	0.4985	0.3292	-0.9967	3.5931	-0.7173	0.9858
$\rho_{RC}(\xi_{EU/US}\xi_{JP/US})_{t,1m}$	0.5180	0.1920	-0.6890	4.0693	-0.2524	0.8831
$\rho_{RC}(\xi_{EU/US}\xi_{JP/US})_{t,3m}$	0.5246	0.1285	-0.5677	3.1021	0.1333	0.7446
$\rho_{RC}(\xi_{US/EU}\xi_{JP/EU})_{t,1w}$	0.4784	0.3517	-1.0299	3.8019	-0.8944	0.9869
$\rho_{RC}(\xi_{US/EU}\xi_{JP/EU})_{t,1m}$	0.4951	0.2324	-0.7038	3.4955	-0.4030	0.9366
$\rho_{RC}(\xi_{US/EU}\xi_{JP/EU})_{t,3m}$	0.5109	0.1592	-0.2915	2.8865	0.0951	0.8600
$\rho_{RC}(\xi_{US/JP}\xi_{EU/JP})_{t,1w}$	0.4236	0.3288	-0.7938	3.5145	-0.8229	0.9582
$\rho_{RC}(\xi_{US/JP}\xi_{EU/JP})_{t,1m}$	0.4420	0.2155	-0.6316	3.0184	-0.2972	0.8626
$\rho_{RC}(\xi_{US/JP}\xi_{EU/JP})_{t,3m}$	0.4441	0.1234	-0.6200	2.9045	0.0507	0.6677

This table presents six summary statistics of out-of-sample implied correlation and out-of-sample realized correlation for two currency trios. The reported statistics include mean, standard deviation, skewness, kurtosis, minimum value and maximum value. Implied correlation is represented by $\rho_{IC}(\xi_{A/C}\xi_{B/C})_{t,T}$, while realized correlation is denoted by $\rho_{RC}(\xi_{A/C}\xi_{B/C})_{t,T}$. Here, three forecast horizons are analyzed (one-week, one-month and three-month). And the sample starts from Nov 4th 2002 to May 31st 2005 with totally 621 observations.

Table 5.2 Panel A. Evaluation of correlation forecasts in EU/US/UK trio under MFE, MAE and MSE

		US/EU & UK/EU				EU/US & UK/US				EU/UK & US/UK			
		$\rho(\xi_{US/EU}\xi_{UK/EU})_{t,1w}$				$\rho(\xi_{EU/US}\xi_{UK/US})_{t,1w}$				$\rho(\xi_{EU/UK}\xi_{US/UK})_{t,1w}$			
		MFE		**MAE**	**MSE**	**MFE**		**MAE**	**MSE**	**MFE**		**MAE**	**MSE**
One week	HIS-7	0.009		0.327’	0.174’	-0.002		0.214’	0.085’	0.005		0.434’	0.284’
	HIS-22	-0.009		0.275’	0.120’	-0.008		0.171’	0.053’	-0.031		0.336’	0.177’
	HIS-65	-0.011		0.253’	0.101’	-0.014		0.157’	0.046’	-0.042		0.307’	0.147’
	EWMA	-0.010		0.253’	0.101’	-0.010		0.162’	0.047’	-0.029		0.319’	0.157’
	Normal	-0.062	**	0.253’	0.105’	-0.005		0.164’	0.047’	-0.011		0.324’	0.159’
	GED	-0.047	**	0.246	0.099	0.023	*	0.169’	0.048’	0.027		0.321’	0.156’
	CCC	-0.172	**	0.263’	0.124’	0.040	**	0.167’	0.045	0.082	**	0.317’	0.145’
	VECH	-0.088	**	0.251’	0.109’	0.013		0.162’	0.045	0.057	**	0.315’	0.147’
	BEKK	-0.095	**	0.248’	0.108’	-0.074	**	0.183’	0.062’	0.017		0.312’	0.147’
	DCC	-0.097	**	0.251’	0.110’	-0.004		0.157	0.043	0.040		0.311’	0.144’
	Implied	-0.045	**	0.237	0.094	-0.020		0.153	0.045	0.008		0.303	0.140
		$\rho(\xi_{US/EU}\xi_{UK/EU})_{t,1m}$				$\rho(\xi_{EU/US}\xi_{UK/US})_{t,1m}$				$\rho(\xi_{EU/UK}\xi_{US/UK})_{t,1m}$			
One Month	HIS-7	0.026		0.279’	0.127’	0.008		0.176’	0.059’	0.028		0.372’	0.206’
	HIS-22	0.007		0.197’	0.063’	0.001		0.117’	0.023’	-0.008		0.253’	0.094’
	HIS-65	0.005		0.155’	0.039’	-0.005		0.102’	0.016’	-0.019		0.188’	0.052’
	EWMA	0.006		0.165’	0.044’	-0.001		0.107’	0.018’	-0.005		0.219	0.071’
	NORMAL	-0.054	**	0.160’	0.039’	0.007		0.103’	0.016’	0.014		0.200’	0.061’
	GED	-0.042	**	0.160’	0.038’	0.032	**	0.107’	0.017’	0.048	**	0.201’	0.063’
	CCC	-0.156	**	0.186’	0.056’	0.050	**	0.105’	0.015’	0.106	**	0.187	0.053’
	VECH	-0.072	**	0.163’	0.044’	0.022	**	0.102’	0.015’	0.080	**	0.197’	0.056’
	BEKK	-0.079	**	0.159’	0.042’	-0.065	**	0.146’	0.033’	0.041	**	0.196’	0.055’
	DCC	-0.081	**	0.160’	0.043’	0.006		0.097’	0.014’	0.063	**	0.188	0.051
	Implied	-0.034	**	0.143	0.032	-0.012		0.086	0.012	0.033	**	0.184	0.052
		$\rho(\xi_{US/EU}\xi_{UK/EU})_{t,3m}$				$\rho(\xi_{EU/US}\xi_{UK/US})_{t,3m}$				$\rho(\xi_{EU/UK}\xi_{US/UK})_{t,3m}$			
Three Month	HIS-7	0.024		0.266’	0.115’	0.016		0.164’	0.054’	0.036	*	0.344’	0.173’
	HIS-22	0.006		0.157’	0.042’	0.010		0.103’	0.018’	0.000		0.182’	0.050’
	HIS-65	0.004		0.118’	0.025’	0.004		0.082’	0.010’	-0.011		0.128’	0.023’
	EWMA	0.004		0.127’	0.027’	0.008		0.090’	0.013’	0.002		0.142’	0.033’
	NORMAL	-0.075	**	0.116’	0.021’	0.020	**	0.073’	0.008’	0.024	**	0.117’	0.021’
	GED	-0.068	**	0.122’	0.023’	0.040	**	0.074’	0.008’	0.050	**	0.130’	0.025’
	CCC	-0.157	**	0.168’	0.037’	0.058	**	0.075’	0.009’	0.113	**	0.130’	0.023’
	VECH	-0.073	**	0.123’	0.025’	0.031	**	0.072’	0.008’	0.088	**	0.118’	0.021’
	BEKK	-0.081	**	0.120’	0.023’	-0.056	**	0.123’	0.024’	0.049	**	0.115’	0.020’
	DCC	-0.083	**	0.121’	0.022’	0.014	**	0.074’	0.008’	0.071	**	0.110	0.018
	Implied	-0.054	**	0.104	0.016	-0.003		0.056	0.004	0.057	**	0.122’	0.021’

This panel presents the evaluation results of correlation forecast in GBP trio under three statistical loss functions. The unbiasedness test is performed by regressing forecast error on a constant with standard errors corrected for heteroskedasticity and autocorrelation by adopting Newey and West (1987)'s procedure.** and * indicates the resultant error are significantly different from zero at 99% and 95% level. Underlined numbers are those having the lowest absolute value in the group of forecasts which evaluated by either MFE, MSE or MAE. It represents the 'best' model under these loss functions. For example, when correlation forecasts of $\rho(\xi_{US/EU}\xi_{UK/EU})_{t,1w}$ is evaluated by MFE(mean forecast error), the best model is then HIS-22. Here, His-22 denotes the historical correlation models using returns of past 22 days to calculate future correlation. Besides, for comparing predictive accuracy, we also perform Diebold Mariano test on MSE and MAE results here and use ‘ to represents the cases where the null of equal predictive accuracy of a forecast is rejected at 5% level, when it is compared to the best performing model in its group.

Table 5.2 Panel B. Evaluation of correlation forecasts in EU/US/JP trio under MFE, MAE and MSE

		EU/US & JP/US				US/EU & JP/EU				EU/JP & US/JP			
		$\rho(\xi_{EU/US}\xi_{JP/US})_{t,1w}$				$\rho(\xi_{US/EU}\xi_{JP/EU})_{t,1w}$				$\rho(\xi_{EU/JP}\xi_{US/JP})_{t,1w}$			
		MFE		**MAE**	**MSE**	**MFE**		**MAE**	**MSE**	**MFE**		**MAE**	**MSE**
One week	HIS-7	-0.0032		0.3276’	0.1878’	0.0078		0.3331	0.1906	0.0064		0.3106	0.1603
	HIS-22	-0.021		0.2735’	0.1279’	-0.0163		0.2667	0.1237	-0.0161		0.2641	0.1108
	HIS-65	-0.0233		0.2584’	0.1119’	-0.0139		0.2671	0.1204	-0.041	**	0.2637	0.1137
	EWMA	-0.0218		0.2591’	0.1142’	-0.0222		0.2511	0.1105	-0.0205		0.244	0.097
	Normal	0.0285	*	0.2653’	0.1076	-0.0031		0.2707	0.1191	-0.038	*	0.2613	0.1106
	GED	0.0325	*	0.2621	0.1039	-0.0127		0.2653	0.1156	-0.0328		0.2596	0.1078
	CCC	0.1702	**	0.3116’	0.1307’	-0.1104	**	0.2625	0.1297	-0.1241	**	0.269	0.1239
	VECH	0.0355		0.2679	0.1076	0.0511	**	0.307	0.1583	0.0382		0.2925	0.1395
	BEKK	-0.015		0.2584	0.1091	-0.0524	**	0.2555	0.1143	-0.0945	**	0.2528	0.1097
	DCC	-0.0052		0.2642	0.1132’	-0.059	**	0.2569	0.1157	-0.0697	**	0.2487	0.1047
	Implied	-0.0332	*	0.2487	0.1081	-0.0245		0.2506	0.1049	-0.0253		0.2568	0.1059
		$\rho(\xi_{EU/US}\xi_{JP/US})_{t,1m}$				$\rho(\xi_{US/EU}\xi_{JP/EU})_{t,1m}$				$\rho(\xi_{EU/JP}\xi_{US/JP})_{t,1m}$			
One Month	HIS-7	0.0162		0.2868’	0.1426’	0.0246		0.295’	0.1441’	0.0249		0.3111’	0.1532’
	HIS-22	-0.0015		0.1972’	0.064’	0.0005		0.2096’	0.0682’	0.0024		0.2277’	0.0811’
	HIS-65	-0.0038		0.1606’	0.0396’	0.0029		0.1931’	0.0608’	-0.0225	*	0.2099’	0.0683’
	EWMA	-0.0023		0.1699’	0.0473’	-0.0054		0.1817’	0.0523’	-0.002		0.1986’	0.0623’
	Normal	0.0558	**	0.1543’	0.0368’	0.0133		0.1929’	0.0549’	-0.0189	*	0.195’	0.0575’
	GED	0.0557	**	0.1529’	0.0368’	-0.002		0.1895’	0.055’	-0.02	*	0.1856’	0.0537’
	CCC	0.1897	**	0.2228’	0.0684’	-0.0936	**	0.178’	0.0575’	-0.1056	**	0.1856’	0.0595’
	VECH	0.0551	**	0.1564	0.0388	0.068	**	0.2411’	0.0921’	0.0568	**	0.2256’	0.0862’
	BEKK	0.005		0.1524	0.0371	-0.0356	**	0.1725’	0.0488’	-0.076	**	0.1856’	0.0581’
	DCC	0.0173	*	0.1643’	0.0414’	-0.0424	**	0.171’	0.0483’	-0.0514	**	0.1813’	0.0548’
	Implied	-0.011		0.1498	0.0337	-0.0263	**	0.1544	0.04	0.0021		0.1645	0.0429
		$\rho(\xi_{EU/US}\xi_{JP/US})_{t,3m}$				$\rho(\xi_{US/EU}\xi_{JP/EU})_{t,3m}$				$\rho(\xi_{EU/JP}\xi_{US/JP})_{t,3m}$			
Three Month	HIS-7	0.0229		0.2672’	0.1222’	0.0404	*	0.2948’	0.1431’	0.0269		0.301’	0.1404’
	HIS-22	0.0051		0.1538’	0.0399’	0.0163		0.2037’	0.0622’	0.0044		0.2134’	0.0663’
	HIS-65	0.0028		0.1162’	0.0213’	0.0187		0.1776’	0.0439’	-0.0204		0.1783’	0.0473’
	EWMA	0.0043		0.1301’	0.0279’	0.0104		0.1696’	0.0438’	0.0001		0.1833’	0.048’
	Normal	0.0792	**	0.1183’	0.0188’	0.0267	**	0.1663’	0.0374’	-0.0124		0.1433’	0.0298’
	GED	0.0695	**	0.1172’	0.0197’	-0.0011		0.1607’	0.0359’	-0.0266	**	0.1416’	0.0303’
	CCC	0.1963	**	0.2003’	0.0518’	-0.0778	**	0.1294’	0.0274’	-0.1035	**	0.1272’	0.03’
	VECH	0.0618	**	0.1136’	0.0191’	0.0838	**	0.2305’	0.0859’	0.0589	**	0.1817’	0.0533’
	BEKK	0.0118		0.1058’	0.0177’	-0.0197	**	0.1409’	0.0293’	-0.0739	**	0.1462’	0.0347’
	DCC	0.024	**	0.1142’	0.0209’	-0.0266	**	0.134’	0.027’	-0.0493	**	0.1426’	0.0318’
	Implied	-0.0066		0.0952	0.014	-0.0368	**	0.1146	0.0228	0.0303	**	0.1085	0.0163

This panel presents the evaluation results of correlation forecast in JPY trio under three statistical loss functions. The unbiasedness test is performed by regressing forecast error on a constant with standard errors corrected for heteroskedasticity and autocorrelation by adopting Newey and West (1987)'s procedure.** and * indicates the resultant error are significantly different from zero at 99% and 95% level. Underlined numbers are those having the lowest absolute value in the group of forecasts which evaluated by either MFE, MSE or MAE. It represents the 'best' model under these loss functions. For example, For example, when correlation forecasts of ρ(ξEU/USξJP/US)t,1w is evaluated by MFE(mean forecast error), the best model is then HIS-7. Here, His-7 denotes the historical correlation models using returns of past 7 days to calculate future correlation. Besides, for comparing predictive accuracy, we also perform Diebold Mariano test on MSE and MAE results here and use ‘ to represents the cases where the null of equal predictive accuracy of a forecast is rejected at 5% level, when it is compared to the best performing model in its group.

Table 5.3 Panel A. Partial optimal regression results of EU/US/UK trio for three forecast horizons

		US/EU & UK/EU					EU/US & UK/US					EU/UK & US/UK				
		$\rho(\xi_{US/EU}\xi_{UK/EU})_{t,1w}$					$\rho(\xi_{EU/US}\xi_{UK/US})_{t,1w}$					$\rho(\xi_{EU/UK}\xi_{US/UK})_{t,1w}$				
		a		b		R^2	a		b		R^2	a		b		R^2
One week	HIS-7	0.3971	**	0.1517	"	0.0256	0.6583	**	0.0878	"	0.0072	0.1801	**	0.0414	"	0.0004
	HIS-22	0.3802	**	0.1814	"	0.0113	0.5794	**	0.1951	"	0.011	0.181	**	0.0307	"	0.0063
	HIS-65	0.3142	**	0.3186	"	0.0142	0.4577	**	0.3591	"	0.0148	0.1741	**	0.0593	"	0.0054
	EWMA	0.2819	**	0.3867	"	0.0347	0.4601	**	0.3579	"	0.0251	0.1728	**	0.0689	"	0.0045
	Normal	0.3654	**	0.1913	"	0.0109	0.6814	**	0.2056	"	0.0106	0.2678	**	0.4035	"	0.0134
	GED	0.2311	**	0.4581	"	0.034	0.551	**	0.2445	"	0.0176	0.1887	**	-0.0062	"	0.0102
	CCC	-0.0426		0.7966	‘	0.0097	0.2076		0.7547		0.0224	0.5783	**	0.7342	"	0.0183
	VECH	0.3718	**	0.1709	"	0.011	0.4944	**	0.3206	"	0.0108	0.2948	**	0.8184	"	0.0131
	BEKK	0.2993	**	0.2975	"	0.017	0.9445	**	0.2799	"	0.0128	0.2487	**	0.358	"	0.0104
	DCC	0.3163	**	0.2663	"	0.0168	0.2762	*	0.6144	"	0.0268	0.3251	**	0.928	"	0.0119
	Implied	0.1423	*	0.6328	"	0.059	0.378	**	0.4636	"	0.0399	0.1212	**	0.3706	"	0.0146
		$\rho(\xi_{US/EU}\xi_{UK/EU})_{t,1m}$					$\rho(\xi_{EU/US}\xi_{UK/US})_{t,1m}$					$\rho(\xi_{EU/UK}\xi_{US/UK})_{t,1m}$				
One month	HIS-7	0.4423	**	0.0904	"	0.0263	0.6964	**	0.0476	"	0.0066	0.2134	**	0.0128	"	0.001
	HIS-22	0.4313	**	0.11	"	0.0121	0.5957	**	0.1852	"	0.0345	0.2416	**	0.1392	"	0.0188
	HIS-65	0.3363	**	0.3084	"	0.0411	0.4987	**	0.3156	"	0.0385	0.2224	**	0.0494	"	0.0105
	EWMA	0.3552	**	0.2691	"	0.0492	0.5204	**	0.2878	"	0.0529	0.2476	**	0.1687	"	0.0149
	Normal	0.3236	**	0.2978	"	0.0191	0.5632	**	0.2314	"	0.012	0.282	**	0.3589	"	0.0364
	GED	0.3122	**	0.3266	"	0.0231	0.4938	**	0.3395	"	0.0461	0.223	**	-0.073	"	0.0115
	CCC	-0.165		0.8321		0.0754	0.1222		0.7333		0.0606	0.6205	**	0.7943	"	0.0729
	VECH	0.2873	**	0.3508	"	0.0233	0.4546	**	0.3897	"	0.0986	0.3432	**	1.0109	"	0.0751
	BEKK	0.2665	**	0.3863	"	0.0402	0.3861	**	0.4466	"	0.1143	0.3085	**	0.5737	"	0.0496
	DCC	0.2756	**	0.3687	"	0.0447	0.3804	**	0.4832	"	0.0949	0.3832	**	1.1636	"	0.071
	Implied	0.1719	**	0.6025	"	0.0967	0.2268	**	0.6781	"	0.1483	0.2351	**	0.2446	"	0.0233
		$\rho(\xi_{US/EU}\xi_{UK/EU})_{t,3m}$					$\rho(\xi_{EU/US}\xi_{UK/US})_{t,3m}$					$\rho(\xi_{EU/UK}\xi_{US/UK})_{t,3m}$				
Three Month	HIS-7	0.4557	**	0.057	"	0.0237	0.7163	**	0.0322	"	0.0067	0.219	**	-0.002	"	0.0016
	HIS-22	0.4145	**	0.1414	"	0.0501	0.6701	**	0.0952	"	0.0193	0.2129	**	0.0258	"	0.0129
	HIS-65	0.403	**	0.1647	"	0.0262	0.5966	**	0.1946	"	0.0318	0.2523	**	0.1467	"	0.0209
	EWMA	0.3696	**	0.2349	"	0.0867	0.6331	**	0.1458	"	0.0291	0.2177	**	0.0043	"	0.0157
	Normal	0.2605	**	0.3976	"	0.0556	0.4921	**	0.3442	"	0.0485	0.2268	**	0.0422	"	0.0254
	GED	0.3785	**	0.1878	"	0.0815	0.3938	**	0.4952	"	0.1623	0.2109	**	0.0456	"	0.0151
	CCC	-0.3271	**	0.8942	‘	0.2005	0.042		0.7024		0.1776	0.3946	**	0.8341	"	0.0553
	VECH	0.2849	**	0.3516	"	0.1553	0.4331	**	0.4325	"	0.1141	0.2391	**	0.1567	"	0.046
	BEKK	0.252	**	0.4087	"	0.1052	0.3636	**	0.4073	"	0.21	0.2394	**	0.1228	"	0.0391
	DCC	0.2232	**	0.4581	"	0.1616	0.632	**	0.3484	"	0.1034	0.2605	**	0.2835	"	0.0162
	Implied	0.1708	**	0.5811	"	0.18	0.0228		0.9656		0.326 #	0.2148	**	0.3233	"	0.0415

This panel presents the partial optimal results of realized correlation in GBP trio regressed by the forecasts generated from 11 correlation models. Here, *a* denotes the coefficient of constant in the regression; b denotes the coefficient of the dependent variable. R^2 represents the goodness-of-fit. ** indicates the hypothesis of zero constant in the regression (a=0) is rejected at 99% level. * represents rejection at 95% level. “ indicates the hypothesis of coefficient of independent variable equaling to one (b=1) is rejected at 99% level. ' represents rejection at 95% level. # indicates the joint hypotheses test of both a=0 and b=1 cannot be rejected at 99% level after regression. The underlined number indicates the model that has the highest explanation power.

Table 5.3 Panel B. Partial optimal regression results of EU/US/JP trio for three forecast horizons

		EU/US & JP/US					US/EU & JP/EU					EU/JP & US/JP				
		$\rho(\xi_{EU/US}\xi_{JP/US})_{t,1w}$					$\rho(\xi_{US/EU}\xi_{JP/EU})_{t,1w}$					$\rho(\xi_{EU/JP}\xi_{US/JP})_{t,1w}$				
		a		b		**R^2**	a		b		**R^2**	a		b		**R^2**
	HIS-7	0.4127	**	0.171	"	0.0312	0.3556	**	0.2608	"	0.0757	0.3029	**	0.2892	"	0.0946
	HIS-22	0.3702	**	0.247	"	0.0199	0.231	**	0.5	"	0.1112	0.2156	**	0.4728	"	0.0966
	HIS-65	0.2896	**	0.4004	"	0.0222	0.2047	**	0.5559	"	0.0722	0.2454	**	0.3836	"	0.0223
	EWMA	0.2929	**	0.3952	"	0.0366	0.1369	**	0.6821	"	0.1378	0.1141	**	0.6968	"	0.1288
	Normal	0.2229	**	0.5865	‘	0.0247	0.1832	**	0.6131	"	0.0572	0.2256	**	0.429	"	0.0126
One week	GED	0.1093		0.8353		0.0494	0.1473	**	0.6742	"	0.0836	0.164	*	0.5688	"	0.0244
	CCC	0.2024	**	0.9019		0.0591	0.1487	**	0.601	"	0.086	0.214	*	0.3827	‘	0.0173
	VECH	0.2058	*	0.6322	‘	0.0243	0.386	**	0.3621	"	0.0199	0.3902	**	0.4934	"	0.025
	BEKK	0.2599	**	0.4648	"	0.0181	-0.0559		1.0066		0.0952	-0.0682		0.9492		0.0654
	DCC	0.359	**	0.2786	"	0.0261	-0.0641		1.0094		0.0898	-0.0625		0.9854		0.0759
	Implied	0.2107	**	0.5413	"	0.0367	-0.0723		1.0951		0.1553#	0.1611	**	0.5847	"	0.0489
		$\rho(\xi_{EU/US}\xi_{JP/US})_{t,1m}$					$\rho(\xi_{US/EU}\xi_{JP/EU})_{t,1m}$					$\rho(\xi_{EU/JP}\xi_{US/JP})_{t,1m}$				
	HIS-7	a		b		R^2	a		b		R^2	a		b		R^2
	HIS-22	0.4858	**	0.0641	"	0.0119	0.4103	**	0.1804	"	0.083	0.4121	**	0.0717	"	0.0122
	HIS-65	0.4434	**	0.1436	"	0.0198	0.3114	**	0.3714	"	0.1408	0.3832	**	0.1334	"	0.0167
	EWMA	0.3019	**	0.4142	"	0.0732	0.3062	**	0.3838	"	0.0789	0.4931	**	-0.101	"	0.0104
	Normal	0.3614	**	0.3008	"	0.0633	0.2334	**	0.5229	"	0.186	0.342	**	0.2253	"	0.0301
One month	GED	0.184	**	0.7225	"	0.0995	0.2643	**	0.4791	"	0.0782	0.5211	**	-0.1715	"	0.0324
	CCC	0.1812	**	0.6584	"	0.0972	0.2587	**	0.4756	"	0.0913	0.3952	**	0.1015	"	0.0105
	VECH	0.2638	**	0.7043	‘	0.1291	-1.084	**	0.5315	"	0.1578	0.3852	**	0.1038	"	0.0181
	BEKK	0.2506	**	0.5777	"	0.062	0.4075	**	0.5032	"	0.1028	0.4436	**	0.1406	"	0.0143
	DCC	0.2668	**	0.4897	"	0.0625	0.0835		0.7757	‘	0.1291	0.3744	**	0.1306	"	0.0203
	Implied	0.3618	**	0.3119	"	0.0287	0.045		0.8374		0.1423	0.3645	**	0.1571	"	0.0273
		0.124	*	0.7448	‘	0.0971	-0.2093	**	1.3511	"	0.2904	0.1387	**	0.6895	"	0.0935
	HIS-7	$\rho(\xi_{EU/US}\xi_{JP/US})_{t,3m}$					$\rho(\xi_{US/EU}\xi_{JP/EU})_{t,3m}$					$\rho(\xi_{EU/JP}\xi_{US/JP})_{t,3m}$				
	HIS-22	a		b		R^2	a		b		R^2	a		b		R^2
	HIS-65	0.4918	**	0.0654	"	0.0299	0.4707	**	0.0854	"	0.0388	0.445	**	0.0208	"	0.018
	EWMA	0.4245	**	0.1927	"	0.0843	0.4265	**	0.1707	"	0.0625	0.4612	**	-0.1389	"	0.0341
	Normal	0.3421	**	0.3498	"	0.1177	0.4174	**	0.1899	"	0.0404	0.6312	**	0.4028	"	0.1858
Three Month	GED	0.3767	**	0.2843	"	0.128	0.3852	**	0.2511	"	0.0906	0.4734	**	0.1659	"	0.0668
	CCC	0.1453	**	0.8517	‘	0.2439	0.4216	**	0.1845	"	0.0227	0.6438	**	0.1275	"	0.0957
	VECH	0.2362	**	0.6338	"	0.1472	0.3879	**	0.2403	"	0.0448	0.5837	**	0.2966	"	0.0507
	BEKK	0.2937	**	0.7033	"	0.2408	-0.7607	**	0.7912	"	0.1886	0.6242	**	0.3289	"	0.0159
	DCC	0.2529	**	0.587	"	0.1451	0.4984	**	0.2932	"	0.0846	0.4345	**	0.2502	"	0.0704
	Implied	0.2954	**	0.4471	"	0.1182	0.3269	**	0.3682	"	0.054	0.6399	**	0.378	"	0.0738
		0.358	**	0.3329	"	0.0755	0.2658	**	0.416	"	0.0893	0.6143	**	0.345	"	0.0642
		0.0213		0.9474		0.1522#	-0.0344		0.9756		0.1525	0.2447	**	0.4819	"	0.1141

This panel presents the partial optimal results of realized correlation in JPY trio regressed by the forecasts generated from 11 correlation models. Here, *a* denotes the coefficient of constant in the regression; b denotes the coefficient of the dependent variable. R^2 represents the goodness-of-fit. ** indicates the hypothesis of zero constant in the regression (a=0) is rejected at 99% level. * represents rejection of null at 95% level. " indicates the hypothesis of coefficient of independent variable equaling to one (b=1) is rejected at 99% level. ' represents rejection at 95% level. # indicates the joint hypotheses test of both a=0 and b=1 cannot be rejected at 99% level after regression. The underlined number indicates the model that has the highest explanation power.

Table 5.4 Sign test result of correlation forecasts with respect to their corresponding information set

EU/US/UK Trio

	$\rho(\xi_{US/EU}\xi_{UK/EU})$						$\rho(\xi_{EU/US}\xi_{UK/US})$						$\rho(\xi_{EU/UK}\xi_{US/UK})$				
	1 w		**1 m**		**3 m**		**1 w**		**1 m**		**3 m**		**1 w**		**1 m**		**3 m**
HIS-7	0.0000		0.0000		0.0000		0.0649	*	0.0003		0.0649	*	0.0000		0.0000		0.0000
HIS-22	0.2282	*	0.5208	*	0.2286	*	0.0102	**	0.7482	*	0.0102	**	0.0000		0.0000		0.0000
HIS-65	0.1486	*	0.1725	*	0.1486	*	0.0000		0.0030		0.0000		0.4222	*	0.5743	*	0.4222
EWMA	0.0448	**	0.6301	*	0.6301	*	0.0002		0.0448		0.0448	**	0.0013		0.0246	**	0.0246
Normal	0.7482	*	0.0000		0.0000		0.0000		0.0010		0.0010		0.9760	*	0.9780	*	0.3355
GED	0.9760	*	0.0007		0.0007		0.0000		0.0000		0.0000		0.5208	*	0.0246	**	0.8097
CCC	0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000
VECH	0.0199	**	0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000
BEKK	0.0161	**	0.0000		0.0000		0.0246		0.0199	**	0.0199		0.1273	*	0.0003		0.0003
DCC	0.0013	**	0.0000		0.0000		0.0000		0.0002		0.0002		0.0013		0.0000		0.0000
Implied	0.9360	*	0.0000		0.0000		0.0001		0.4222	*	0.4222	*	0.5208	*	0.0919	*	0.3773

EU/US/JP Trio

	$\rho(\xi_{EU/US}\xi_{JP/US})$						$\rho(\xi_{US/EU}\xi_{JP/EU})$						$\rho(\xi_{EU/JP}\xi_{US/JP})$				
	1 w		**1 m**		**3 m**		**1 w**		**1 m**		**3 m**		**1 w**		**1 m**		**3 m**
HIS-7	0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000
HIS-22	0.1486	*	0.0102	**	0.1486	*	0.1085	*	0.0302	**	0.1085	*	0.8097	*	0.5743	*	0.8097
HIS-65	0.0002		0.0246	**	0.0002		0.0023		0.0000		0.0023		0.2968	*	0.1725	*	0.2968
EWMA	0.0064		0.5743	*	0.5743	*	0.0030		0.4701	*	0.4701	*	0.2612	*	0.9360	*	0.9360
Normal	0.0000		0.0000		0.0000		0.0010		0.0000		0.0000		0.5743	*	0.0541	*	0.0541
GED	0.0000		0.0000		0.0000		0.0081		0.0161		0.0161		0.1486	*	0.7482	*	0.7482
CCC	0.0000		0.0000		0.0000		0.1486	*	0.0000		0.0000		0.0000		0.0000		0.0000
VECH	0.0000		0.0000		0.0000		0.0007		0.0628		0.0128	**	0.0246	**	0.0030		0.0030
BEKK	0.0000		0.0081		0.0081		0.0775	*	0.8097	*	0.8097	*	0.0199	**	0.0000		0.0000
DCC	0.0000		0.0000		0.0000		0.5743	*	0.5208	*	0.5408	*	0.2612	*	0.0128	**	0.0128
Implied	0.0001		0.8725	*	0.8725	*	0.0001		0.8725	*	0.8925	*	0.2612	*	0.0007		0.0007

This table presents the p-value of sign test performed to examine the partial optimality of the correlation forecast with respect to their corresponding information set. ** here represents the hypothesis of zero median cannot be rejected at 99 confidence level. It is an indication of partial optimality for forecast being analyzed. * indicates the hypothesis of zero median cannot be rejected at 95 confidence level. 1 w, 1 m and 3 m represents one week, one month and three month forecast horizons respectively. 0.000 denotes a very small value.

Table 5.5 Encompassing Regression result for two currency trios and three forecast horizons

Panel A **EU/US/UK Trio**

	$\rho(\xi_{US/EU}\xi_{UK/EU})$						$\rho(\xi_{EU/US}\xi_{UK/US})$						$\rho(\xi_{EU/UK}\xi_{US/UK})$					
	1 w		**1 m**		**3 m**		**1 w**		**1 m**		**3 m**		**1 w**		**1 m**		**3 m**	
Constant	0.16		0.14		-0.3	**	0.35	*	0.59	**	0.34	**	0.59	**	0.36	**	0.48	**
HIS-7	–		–		–		–		–		–		–		–		–	
HIS-22	–		–		–		–		–		–		0.30		–		–	
HIS-65	–		–		–		–		–		-0..30	**	–		–	–	-0.20	**
EWMA	0.13		0.05		0.02		0.19		0.11		–		–		0.18	**	–	
Normal	–		–		–		–		–		–		–		–		–	
GED	-0.10		–		–		–		–		–		–		–		–	
CCC	–		0.05		0.87	**	–	–	–		–		–		–		0.51	**
VECH	–		–		–	–	–		–		–		–		–		–	
BEKK	–		–		–	–	–		-0.30	**	-0.30	**	–		0.33	**	–	
DCC	–		–		–		-0.10		–		–		-0.6	**	–		–	
Implied	0.55	**	0.53	**	0.33	**	0.38	**	0.42	**	1.14	**	0.66	**	0.22	*	0.18	**
R^2	0.06		0.10		0.25		0.05		0.19		0.46		0.04		0.10		0.12	
CHSQ (GARCH =0)	-0.3		0.31		7.46	**	-0.2		5.73	**	8.65	**	3.32	*	7.18	**	-7.9	**
CHSQ (Implied = 0)	4.89	**	3.78	*	4.98	**	3.5	*	4.77	**	16.5	**	3.19	*	2.37	*	4.01	*
CHSQ (Others = 0)	0.66		0.55		56.9	**	2.83		33.1	**	145	**	38.6	**	65.5	**	81.9	**

Panal B **EU/US/JP Trio**

	$\rho(\xi_{EU/US}\xi_{JP/US})$						$\rho(\xi_{US/EU}\xi_{JP/EU})$						$\rho(\xi_{EU/JP}\xi_{US/JP})$					
	1 w		**1 m**		**3 m**		**1 w**		**1 m**		**3 m**		**1 w**		**1 m**		**3 m**	
Constant	-0.00		-0.10		0.04		0.34	**	-0.70	**	-0.80	**	0.29	**	0.35	**	0.35	**
HIS-7	–		–		–		–		–		–		–		–		–	
HIS-22	–		–		–		–		–		–		–		–		–	
HIS-65	–		-0.00		–		–		–		–		–		–		-0.60	**
EWMA	-0.00		–		-0.00		1.09	**	-0.00		-0.10		0.94	**	0.20	**	–	
Normal	–		–		0..75	**	–		–		–		–		-0.70	**	-0.20	**
GED	–		–		–		–		–		–		–		–		–	
CCC	0.83	**	0.72	**	–		–		0.95	**	1.75	**	–		–		–	
VECH	–		–		–		–		–		–		–		–		–	
BEKK	–		–		–		-1.60	**	–		–		–		–		–	
DCC	–		–		–		–		–		–		-0.6	*	–		–	
Implied	0.47	**	0.67	**	0.3	*	0.84	**	1.17	**	0.56	**	0.04		0.71	**	1.08	**
R^2	0.09		0.21		0.25		0.19		0.31		0.24		0.14		0.15		0.48	
CHSQ (GARCH =0)	5.17	**	8.38	**	7.8	**	-4.3	*	3.38	*	8.4	**	-2.1		-6	**	-3.4	*
CHSQ (Implied = 0)	3.3	*	6.79	**	2.69		5.28	**	8.6	**	3.93	*	0.03		6.99	**	18.7	**
CHSQ (Others = 0)	32.2	**	91.6	**	83.2	**	29.1	**	34.1	**	67.4	**	74.6	**	49.6	**	316	**

This table presents the encompassing result. Realized correlations in two trios are respectively regressed on a constant and three correlation forecasts generated from implied correlation model, one GARCH models and one historical correlation models. For the latter two, the models which have shown highest R2 in previous partial optimality regressions are selected as a representative here. The standard errors in regression are corrected for heteroskedasticity and autocorrelation using Newey and West (1987) procedure. The bottom rows of the panel contain the Wald test results for the null hypothesis in parentheses. CHSQ(GARCH = 0) tests the null that coefficient of GARCH-based forecast in regression equals zero; CHSQ (Other = 0) tests the null that coefficients of forecasts generated using time series tools (GARCH and historical correlation) in regression are both equal to zero. Here, ** indicates statistical significance at 99 percent level; * indicates the statistical significance at 95 percent level. 1 w, 1 m and 3 m represents one-week, one-month and three-month forecast horizon respectively.

Table 5.6 Panel A. GMSFEM test (cross horizon forecast ranking) results for EU/US/UK trio

$\rho(\xi_{US/EU}\xi_{UK/EU})$

	HISTOR7		HISTOR22		HISTOR65		EWMA		NORMAL		GED		CCC		VECH		BEKK		DCC	
HIS-22	-0.1	86																		
HIS-65	-0	111	25	-0.2																
EWMA	-0	106	21	-0.1	-4.7	0.2														
Normal	-0.3	113	27	-0.3	-0.7	2.6	-0.3	6.7												
GED	-0.4	113	28	-0.5	-0.4	2.3	7	-0.4	1.1	-0.9										
CCC	-0.1	93	7.6	-0.1	-18	0.1	-13	0	-19	0.3	-20	0.5								
VECH	-0.1	107	22	-0	-4	0.5	-0.2	1.3	-5.6	0.3	-6.1	0.6	14	-0						
BEKK	-0	110	25	-0	-1.6	1.1	-0.1	4.2	-2.6	0.3	-3.2	0.8	17	-0	3.1	0				
DCC	-0.1	110	24	0	-2.4	1.4	-0.3	3.9	-3.2	0.4	-4	1.1	-0	17	-0.1	2.7	-0.8	0.3		
Implied	-1	121	35	-1	-1	10	15	-0.9	8.1	-0.7	-0.6	7.9	28	-0.9	14	-0.9	11	-1	11	-1

$\rho(\xi_{EU/US}\xi_{UK/US})$

	HISTOR7		HISTOR22		HISTOR65		EWMA		NORMAL		GED		CCC		VECH		BEKK		DCC	
HIS-22	44	-0																		
HIS-65	0	54	-0	9.6																
EWMA	51	0	0	6.6	0.1	-3.1														
Normal	-0.1	56	-0.2	11	-0.3	1.8	-0.4	4.8												
GED	-0.2	55	-0.3	10	-0.6	1.2	-0.5	4.1	-0.9	0										
CCC	-0	55	-0	11	-0.1	1	-0.1	4.1	0.3	-0.8	0.8	-0.3								
VECH	-0	56	-0.1	12	-0.1	2.1	-0.1	5.2	0.7	-0.1	1.6	0	-0	1.1						
BEKK	-0	35	-10	0.2	-19	0.1	-17	0.1	-21	0.1	-20	0.2	-21	0	-22	0				
DCC	0	56	-0	12	2.3	0	-0	5.3	1.2	-0.3	2	-0.2	1.3	-0.1	0.5	-0.3	22	-0.1		
Implied	-0.3	60	-0.3	16	-0.3	5.9	-0.4	9	4.3	-0.1	5.2	-0.1	-0.3	4.9	-0.3	3.8	26	-0.2	-0.4	3.7

$\rho(\xi_{EU/UK}\xi_{US/UK})$

	HISTOR7		HISTOR22		HISTOR65		EWMA		NORMAL		GED		CCC		VECH		BEKK		DCC	
HIS-22	-0.1	146																		
HIS-65	189	-0	44	-0.6																
EWMA	-0	171	25	-0.1	-19	0.5														
Normal	-0.5	185	39	-0.5	-6	1.5	-0.4	14												
GED	-0.4	181	35	-0.6	-8.8	0.4	10	-0.4	0.1	-3.9										
CCC	188	-0	42	-0.5	-1.2	0.1	17	-0.4	4.9	-1.5	7.6	-0.3								
VECH	0	188	42	-0.2	-2.7	1.1	17	-0.1	3.4	-0.3	6.6	0.3	-1.6	1.3						
BEKK	0	189	43	-0.2	-2.2	1.9	18	-0	4.2	0.1	7.7	0.4	-1.4	2.2	-0.1	1.3				
DCC	193	0	46	-0.3	-0.5	3.8	22	-0.1	7.8	-0.1	11	0.3	-0.3	4.7	4.8	0	3.7	-0.1		
Implied	191	-0.7	45	-1.1	-1	1.9	20	-1	6.6	-1.1	9.9	-0.6	-0.8	2.9	3.4	-0.9	2.4	-1.2	0.5	-1.9

This table presents the GMSFEM test results for medium-term correlation forecasts. Here, by medium-term, we mean the correlation forecast over the next 'one-month' and correlation forecast over the next 'three month'. To determine whether a model forecast outperforms another for both horizons, we calculate the forecast error of various models first, and then use the distance between the autocovariance of these resulting forecast error to form a function to be evaluated. Above, we present the eigenvalues calculated from this function with column against row. The column model will dominate the corresponding row model if two eigenvalues in the same set are both non-positive and at least one is negative. Vice verse, row model dominates when two eigen-values are both nonnegative and at least one is positive. Indeterminacy comes when mixed sign is presented

Table 5.6 Panel B. GMSFEM test (cross horizon forecast ranking) results for EU/US/JP trio

$\rho(\xi_{EU/US}\xi_{JP/US})$

	HISTOR7		HISTOR22		HISTOR65		EWMA		NORMAL		GED		CCC		VECH		BEKK		DCC	
HIS-22	-0	100																		
HIS-65	127	0	27	-0.1																
EWMA	118	0	18	-0.1	-8.9	0														
Normal	130	-0.2	30	-0.3	3.5	-0.2	12	-0.2												
GED	129	-0	30	-0.2	2.8	-0.1	12	-0.1	0.2	-0.8										
CCC	90	-0	0.5	-11	0	-37	0	-28	0.2	-40	0	-40								
VECH	128	0	29	-0.1	-0.1	2	0	11	-1.7	0.4	-1.4	0.5	-0	39						
BEKK	130	0	31	-0.1	-0	3.9	13	0	-0.2	0.7	-0.3	1.4	-0	41	1.9	0				
DCC	0	126	26	-0	-1.3	0.4	-0	8	-4.5	0.3	-4	0.3	-0	36	-2.8	0	-4.7	0		
Implied	135	-0.3	35	-0.4	-0.4	8.5	-0.3	17	-0.2	5.1	-0.4	5.8	-0.4	45	-0.3	6.6	-0.3	4.7	9.4	-0.3

$\rho(\xi_{US/EU}\xi_{JP/EU})$

	HISTOR7		HISTOR22		HISTOR65		EWMA		NORMAL		GED		CCC		VECH		BEKK		DCC	
HIS-22	-0	97																		
HIS-65	-0.2	114	-0.7	17																
EWMA	-0.1	119	-0	21	6.4	-1.1														
Normal	-0	121	-0.3	24	0.2	7.5	-1.8	4.1												
GED	0	122	-0.4	25	0.2	8.4	-2	5.2	-0.2	1.1										
CCC	-0.7	126	-1.9	30	-1.2	14	-4.1	11	-2.3	7	-2.2	5.9								
VECH	-0	68	-30	0	0.3	-46	0	-51	0.1	-53	0	-54	0.9	-59						
BEKK	-0.2	130	-0.5	33	-0	17	-0.9	12	-0.2	9.1	-0.3	8.2	6	-1.8	-0.3	62				
DCC	-0.3	132	-0.6	35	-0.1	18	-1.1	14	-0.3	11	-0.3	10	7	-1.1	-0.3	64	-0.2	1.9		
Implied	0.4	139	0.3	42	0.6	26	0.2	21	18	0.4	17	0.3	14	0.6	0.4	71	9	0.5	7.5	0.3

$\rho(\xi_{EU/JP}\xi_{US/JP})$

	HISTOR7		HISTOR22		HISTOR65		EWMA		NORMAL		GED		CCC		VECH		BEKK		DCC	
HIS-22	-0	91																		
HIS-65	-0.1	111	-0.2	20																
EWMA	0	114	23	-0	4.3	-1														
Normal	-0.3	128	-0.6	38	-0.4	18	-1.3	16												
GED	0.1	130	-0	39	0.1	20	-0.3	17	2.6	-0.5										
CCC	-0.2	127	-0.6	37	-0.4	17	-1.5	15	-1.6	0.3	-4.1	0.7								
VECH	-0.4	96	-3.6	8.8	-16	0.9	-20	1.7	-32	0.2	-34	0.1	-31	0.1						
BEKK	-0.1	125	-0.2	34	-0	14	-0.7	12	0.6	-4	-0.2	-5.3	1.2	-3.2	29	-0.3				
DCC	-0.1	129	-0.2	38	-0	18	-0.5	15	1.7	-1.2	-0.1	-1.5	3.1	-1.3	33	-0.3	3.9	-0		
Implied	-0.5	146	-0.6	55	-0.4	35	-0.8	33	18	-0.2	-0.6	16	19	-0.4	50	-0.4	-0.4	21	-0.4	17

This table presents the GMSFEM test results for medium-term correlation forecasts. Here, by medium-term, we mean the correlation forecast over the next 'one-month' and correlation forecast over the next 'three month'. To determine whether a model forecast outperforms another for both horizons, we calculate the forecast error of various models first, and then use the distance between the autocovariance of these resulting forecast error to form a function to be evaluated. Above, we present the eigenvalues calculated from this function with column against row. The column model will dominate the corresponding row model if two eigenvalues in the same set are both non-positive and at least one is negative. Vice verse, row model dominates when two eigen-values are both nonnegative and at least one is positive. Indeterminacy comes when mixed sign is presented.

Figure 1 Panal A. Time series plot of implied correlation among the EU/US/GB trio

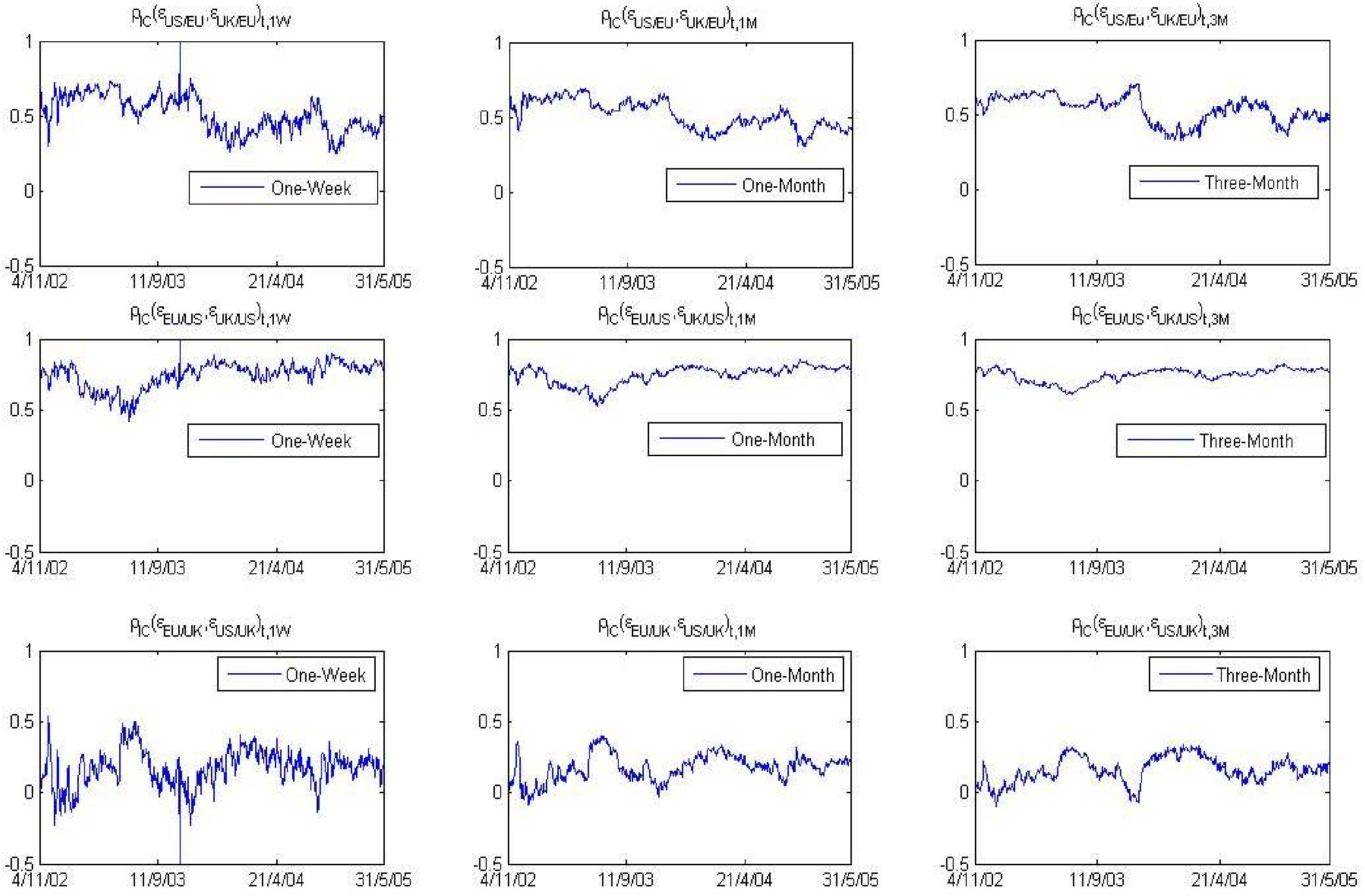


Figure 1, Panel B. Time series plot of implied correlation among the EU/US/JP trio

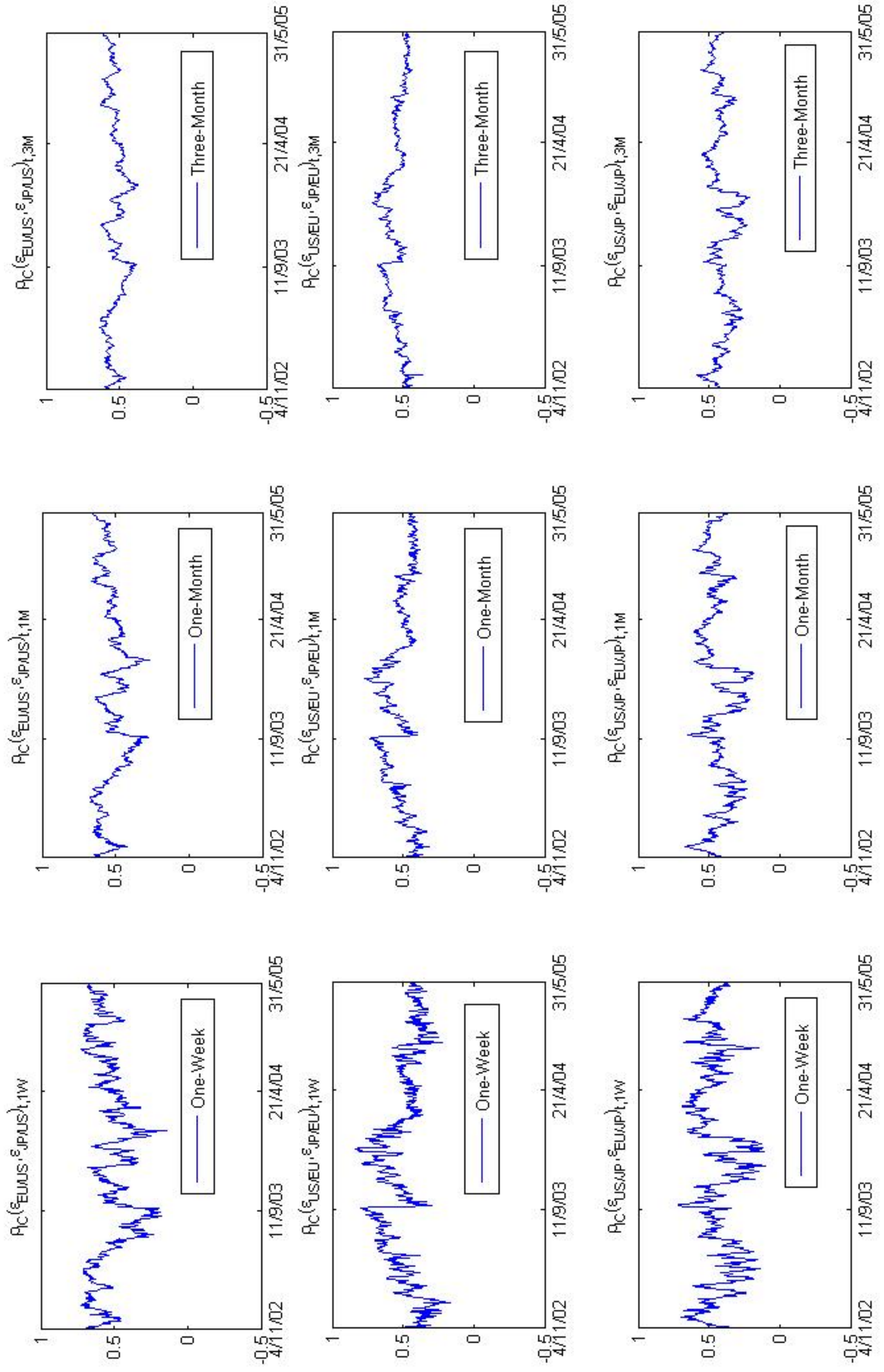


Figure 2, Panal A, Gaussian Kernal distribution of realized correaltion among EU/US/UK trio

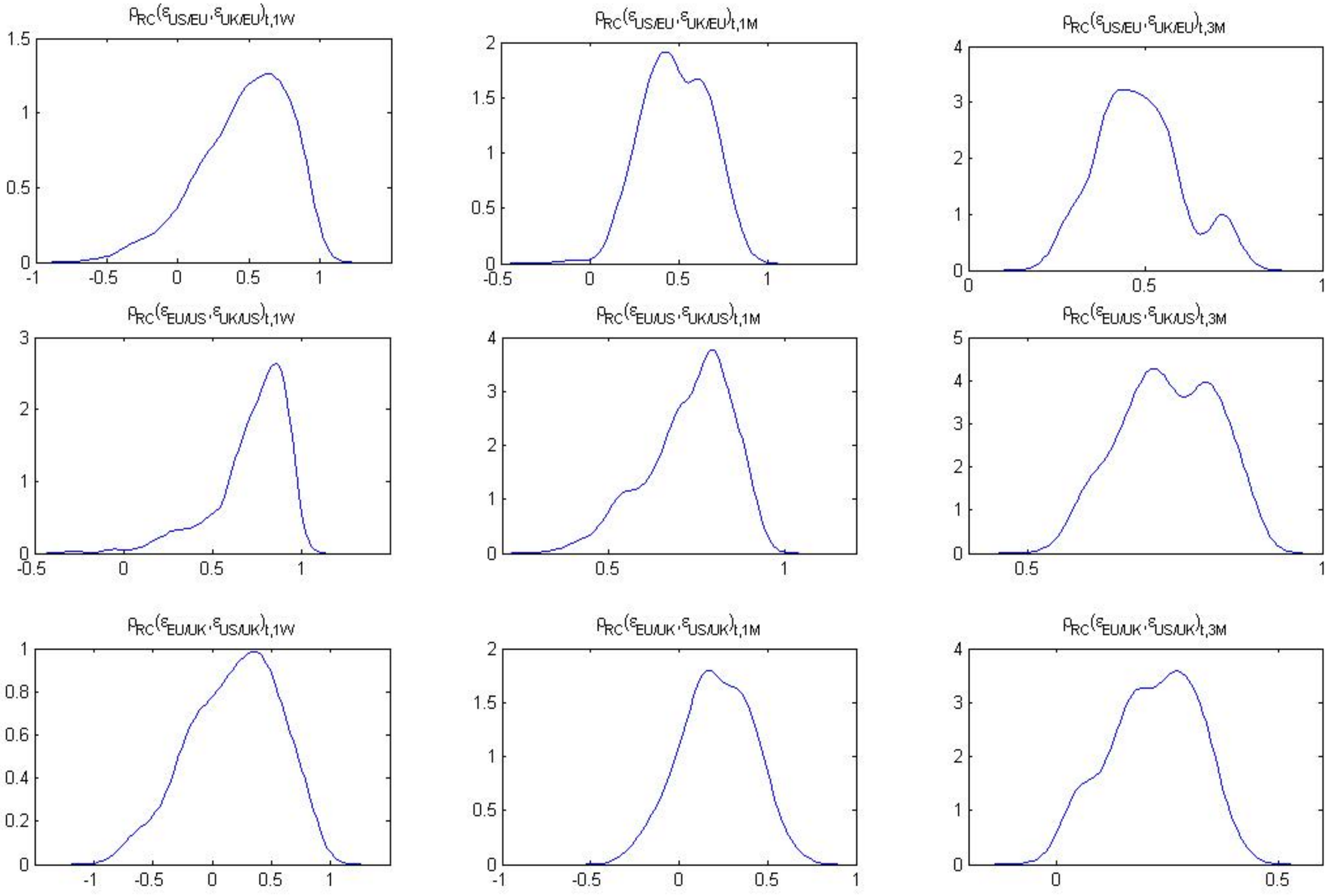


Figure 2, Panal B. Gaussian Kernel density estimates plot of Realized correlations in the EU/US/JP trio

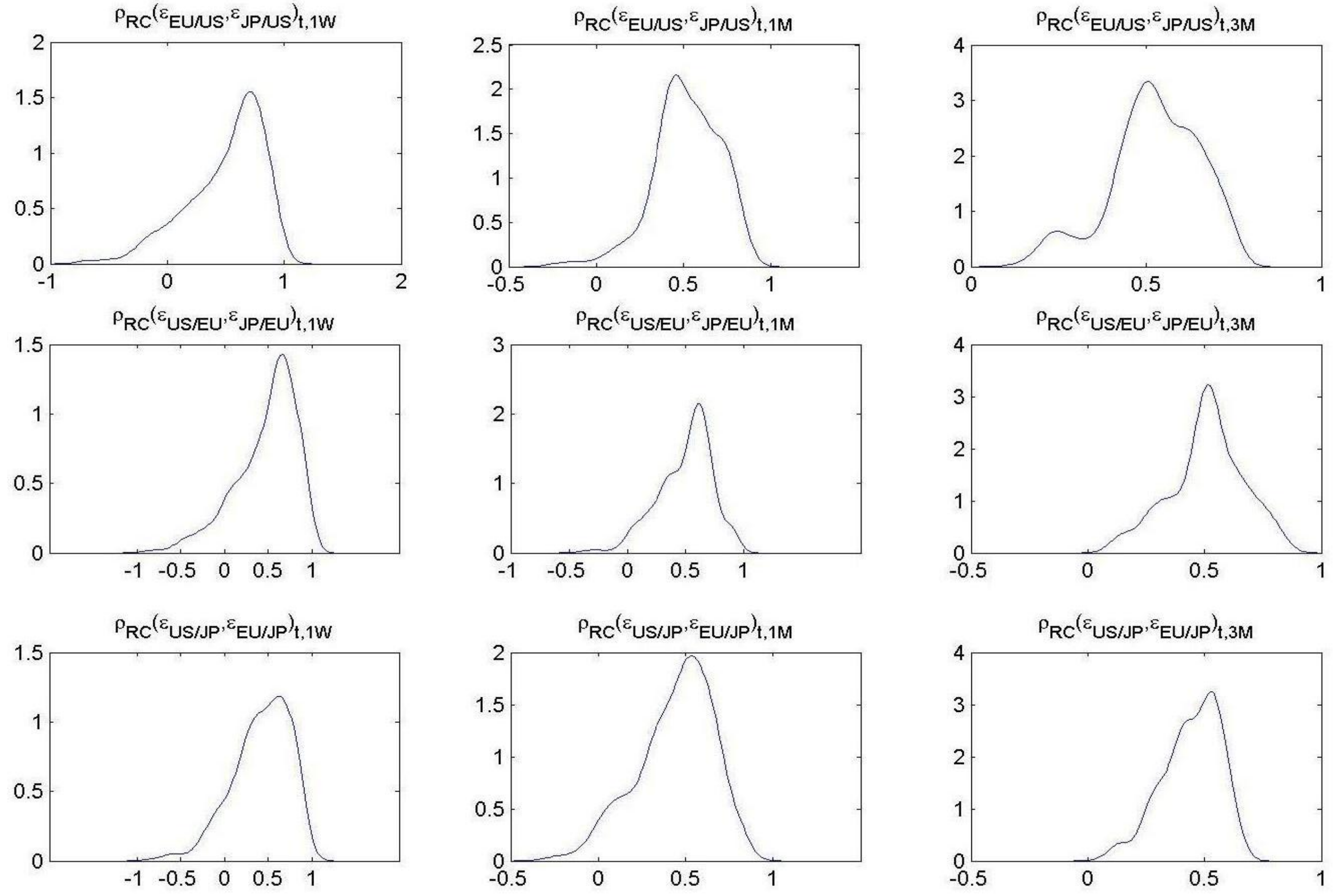


Figure 3.Panal A. Time series plot of realized correaltion versus various forecast with three forecast horizons in the pair of EU/US and UK/US

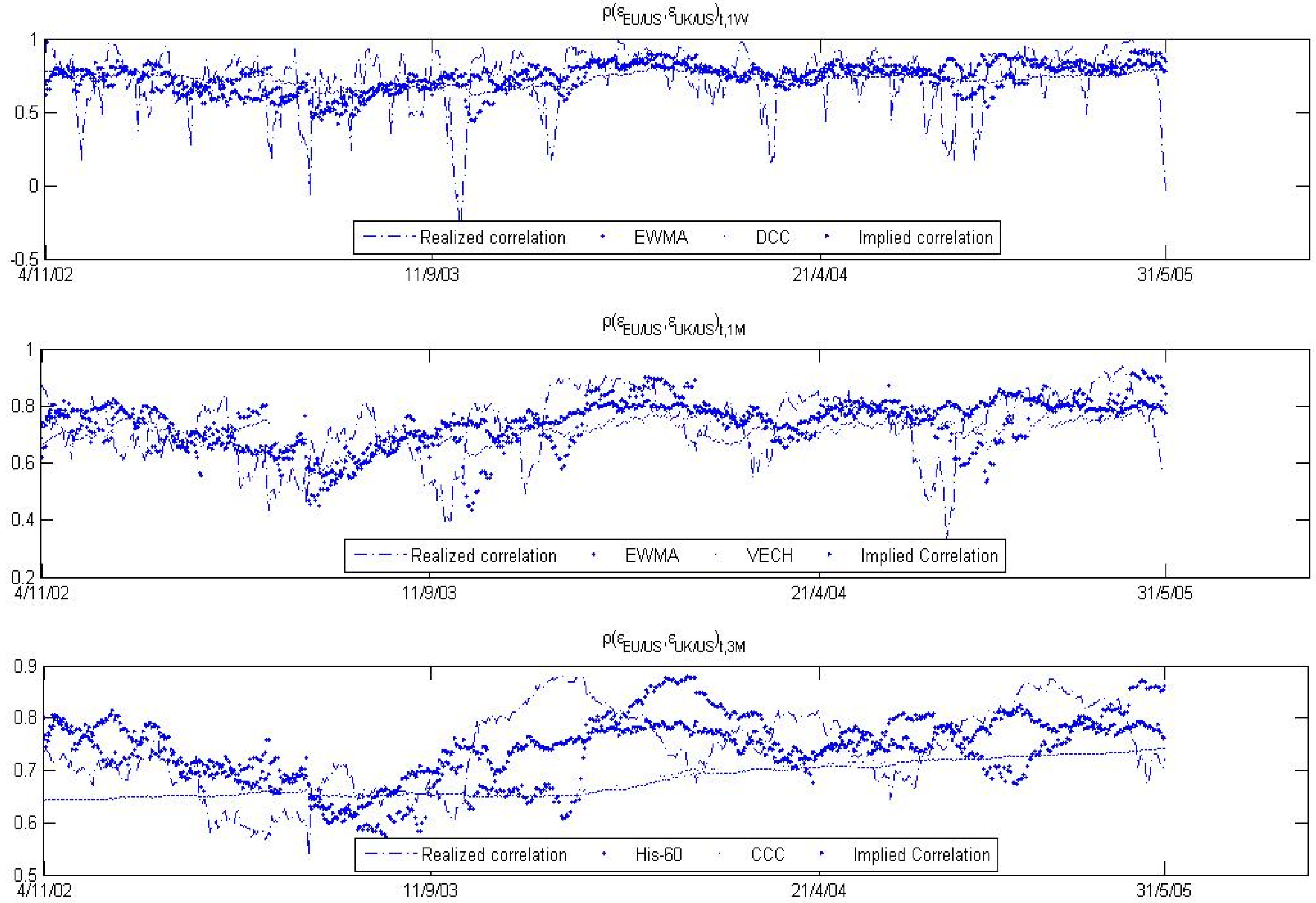
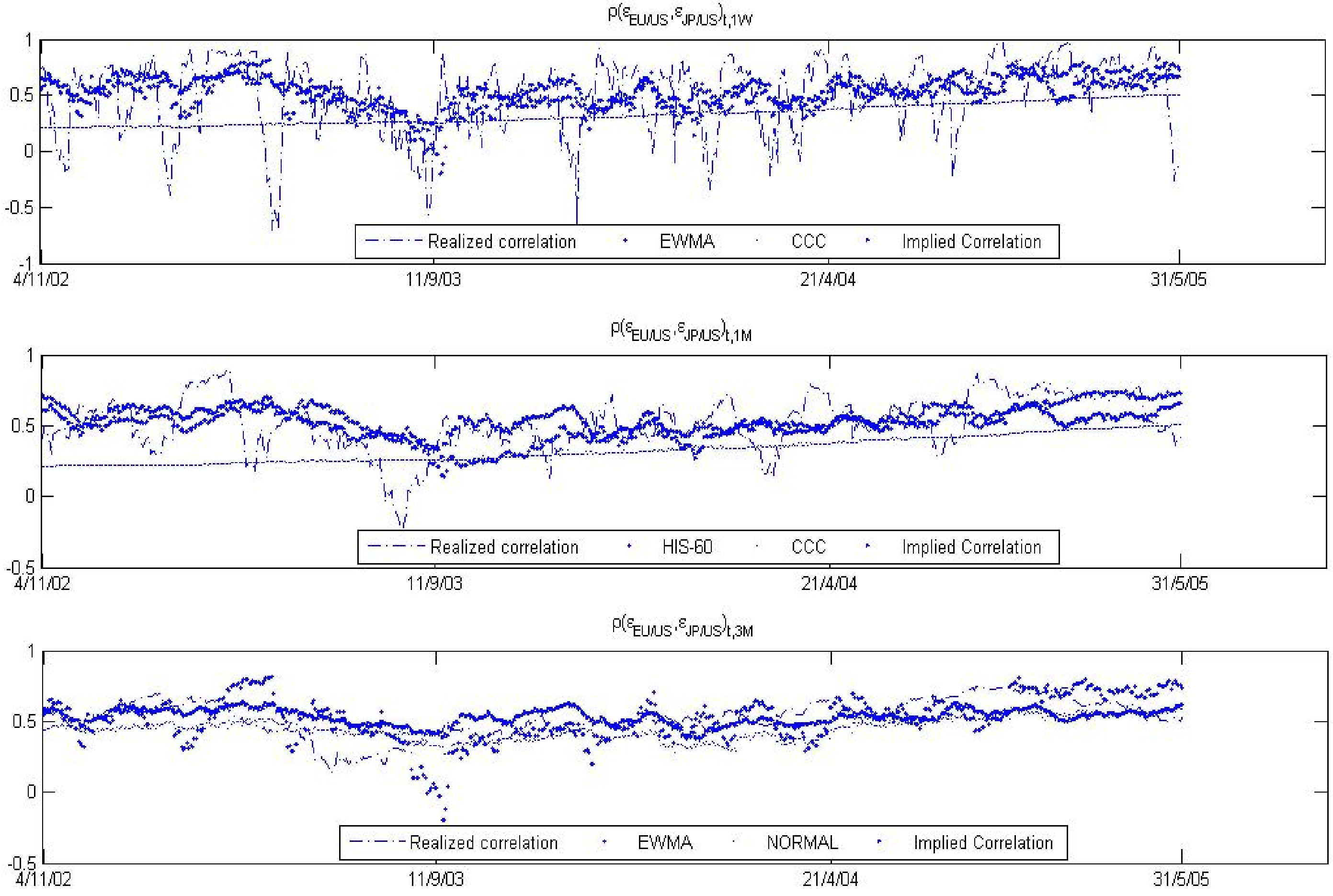


Figure 3. Panel B. Time series Plot of Realized correlation versus various forecasts with three horizons in the pair of EU/US and JP/US



Chapter 6

ADCC-MGM model and ADCC-MTM model

Introduction

In this chapter, we propose two new correlation mixture models, namely the ADCC-MGM model and ADCC-MTM model. The whole chapter is composed of five sections. In the first section, we present the motivation for proposing these two models for fitting correlation dynamics in financial time series and give the model specifications. In the second part, since statistical inferences are to be calculated using Griddy Gibbs sampler, a brief illustration of some preliminary settings for this simulator is provided. Then, the posterior sampling sequence of each model is respectively given in section three and four along with the simulating kernels for each parameter. Finally, in the last section, we also illustrate four methods of evaluating the performance of our models by carrying out in-sample analysis and calculating out-of-sample forecasts.

6.1 ADCC-MGM model and ADCC-MTM model

Research into time series models of dynamic correlation have exploded in recent years. While many models have been proposed for capturing the time-varying characteristics of this association measure, less attention has been paid to explaining the heavy tail and high kurtosis of return distributions concurrently. Since it is expected that a better prediction of future correlation can be generated when asymmetries and non-linearity in the response of covariance to past returns are taken into account, based on this motivation we propose two new correlation models in this chapter. Specifically, to ease the exposition, we use the most parsimonious form of Hafner and Franses's (2003) ADCC structure to model the correlation evolving process and assume the filtered returns of financial assets to follow a standard mixture of two symmetric distributions. Here, for each component, we allow them to have unique time-varying covariance matrixes so that correlation dynamics can be jointly determined by innovations showing different statistical characteristics, and the correlation process modelled by ADCC can allow for asymmetric feedback on good news and bad news.

Since using traditionally distributional assumptions, such as Gaussian and T, for modelling asset returns can ease the calibration, in this thesis we use these densities as components to propose mixtures.[62] Thus, dynamic correlation models given such assumptions are respectively called ADCC-MGM and ADCC-MTM. Besides, since it is known that estimation of mixture models is usually associated with a very complicated log-likelihood function, we adopt the MCMC Bayesian approach to calculate their inferences. To see the details of how to implement this approach and generate posterior draws for each parameter, we dedicate the next section to this topic. However, for now it is warranted to present these two models' specifications first.

Consider a d-dimensional return process y_t now and an unknown multivariate distribution Φ from which y_t is generated, we let the mean of Φ be time-varying and denoted by μ_t

[62] Two main advantages of using Gaussian and T for modelling asset returns are their numerical flexibility and analytical tractability in inference calculation.

and its covariance matrix by Σ_t . The dynamic process of y_t is defined by $y_t \mid F_{t-1} \sim \Phi(\mu_t, \Sigma_t)$ and $y_t = \mu + \Sigma_t^{1/2} \varepsilon_{t-1}$ where $\varepsilon_t \in R^d$ is an *i.i.d* random vector (process) independent of F_{t-1} and $E(y_t \mid \mu, F_{t-1}) = \mu_t$; $Var(y_t \mid \mu, F_{t-1}) = \Sigma_t^{1/2}(\Sigma_t^{1/2})' = \Sigma_t$. Here, since Σ_t can be written as a multiplication product of individual standard deviations and time-varying correlation, the covariance evolving process of ADCC mixture model then can be defined as

$$\begin{aligned}
&\Sigma_t = D_t R_t D_t; \\
&D_t = w + \alpha \sum\nolimits_{i=1}^{p} (y_{t-i} - \mu)(y_{t-i} - \mu)' + \beta \sum\nolimits_{j=1}^{q} D_{t-j} \\
&\varepsilon_{t-1} = y_t D_t^{-1} \\
&Q_t = \left(\bar{Q} - \eta^2 \bar{Q} - \varsigma^2 \bar{Q} - \iota^2 \bar{N}\right) + \eta^2 \sum\nolimits_{i=1}^{g} \varepsilon_{t-i} \varepsilon_{t-i}' + ... \\
&\quad ... + \varsigma^2 \sum\nolimits_{j=1}^{h} Q_{t-j} + \iota^2 \sum\nolimits_{i=1}^{g} \vartheta_{t-i} \vartheta_{t-i}' \\
&R_t = diag(Q_t)^{-1/2} Q_t diag(Q_t)^{-1/2}
\end{aligned} \tag{6.1}$$

where D_t represents the individual volatility calculated by a univariate heteroskedastic model with *p* ARCH lag and *q* GARCH lag; it is a $d \times d$ diagonal matrix with $\sqrt{\Sigma_{it}}$ on its i^{th} diagonal. R_t, which is a function of an authentic variable Q_t, denotes the time-varying correlation matrix. It is modelled by another heteroskedastic GARCH(*g,h*) process with asymmetric response of correlation to the negative shocks now taken into account. ϑ_t denotes the coefficient of this asymmetric effect. It is a variable that takes the value one when $\varepsilon_t < 0$, and zero otherwise. In the matrix form, it can also be written as $\vartheta_t = I[\varepsilon_t < 0] \Theta \varepsilon_t$.[63] Besides, to allow for covariance targeting we also let $\bar{Q} = E[\varepsilon_t \varepsilon_t']$ and $\bar{N} = E[\vartheta_t \vartheta_t']$.

Here, two things need to be noted before proceeding. First, to ensure the parsimony of the proposed mixture model, we set the lagged terms (*p,q,g,h*) in above GARCH processes all equal to one. Second, coefficients of Q_{t-1} and standardized past innovations ε_{t-1} are all set to be squared products so as to ensure the positive definitiveness of resultant covariance

[63] Θ here denotes the Hadamard matrix operator, It means elementwise multiplication.

matrix. As for the second setting, in other versions of ADCCs these parameters are often assumed to be either scalar or diagonal matrices. For example, in Sheppard (2002) the asymmetry correlation is captured by using parameters modelled as diagonal matrices. Indeed, through the inclusion of additional elements, a very general evolving process for dynamic correlation is proposed. However, the estimation cost of his model, compared to ours, is much higher. Since the MCMC algorithm to be used in the later part of this thesis is already known as a computationally demanding technique, it is then preferred to use a relatively simple model when this Bayesian approach is adopted.[64] For example, concerning the mixture models proposed above, the parameter set of interest now contains a lot of elements. Therefore, there is motivation to perform some trimming in this set before the estimation starts. Although it is certain that, after applying this strategy, some flexibility of the mixture models would be inevitably scarified, these losses seldom alter the correlation evolving processes fundamentally. Besides, the high computational cost of performing MCMC also explains why, with the availability of even more generalized choices in literature for modeling correlation process, such as AGDCC of Cajigas and Urga (2005), we still prefer to use the most parsimonious form of Hafner and Franses's (2003) ADCC here. For instance, in ADCC (1,1,1,1) model, there are only three parameters determining the correlation evolving processes. Thus, only three new draws need to be simulated in each iteration of posterior sampling. However, if a bivariate AGDCC (1,1,1,1) is trained, this number then increases to six, which implies a doubling of our estimation costs. Therefore, without losing much generality, we only consider using the simplest version of ADCC here to ensure the efficiency of MCMC simulator.

Besides, assuming a proper specification for Φ is also essential when proposing mixture models. In this thesis, since Φ is now assumed to be M-component mixture-distributed and no hybrid mixing is allowed, equation (6.1) is then considered as specifying a ADCC-MGM if all components are multivariate Gaussian distributed, and ADCC-MTM if all are

[64] High computational cost of calculating Bayesian inference could be due to various reasons. For example, it maybe due to the inclusion of a large parameter set when a complex model is assumed. Meanwhile, another possibility is sampling kernels of hyper-parameters not having analytical forms so that sophisticated simulation techniques need to be applied to generate their posterior draws.

multivariate *T* distributed. Given that parameter set of m^{th} component in Φ is now denoted by φ_m , these two models' specifications can then be respectively defined as,

$$\begin{array}{llll} \Phi(y_t \mid F_{t-1}) = \sum_{m=1}^{M} \pi_m p_m(y_t, \varphi_m) & & & m = 1, 2, \cdots M \\ if \quad p_m(y_t, \varphi_m) \sim N(\mu, \Sigma_t), & then & ADCC - MGM & \\ if \quad p_m(y_t, \varphi_m) \sim t(\mu, \Sigma_t, \nu), & then & ADCC - MTM & \end{array} \tag{6.2}$$

As for the training data y_t, we assume it to be *i.i.d* now in accordance with the convention. Although this assumption is stronger than the local independence that is frequently used in theoretical analysis of mixture models, it will not affect the validity of our inferential results. Besides, since the sampling technique (Griddy Gibbs sampler) to be used for later posterior simulation is a computation-intensive algorithm, to circumvent the 'curse of dimensionality' we only consider bivariate experimental data here and include only two component distributions for each mixture.[65] Thus, the proposed models, given all these settings, are respectively called bivariate two-component ADCC-MGM model and bivariate two-component ADCC-MTM model.

6.2 Sampling Procedure and Preliminary settings

In this section, we show how to calculate the Bayesian inference for above two correlation models. Since MCMC algorithm is to be used for estimation, some preliminary settings concerning the implementation of this sampling technique need to be stated first. Then, for each model we start the illustration of their posterior simulation procedure by firstly specifying a proper prior density for each parameter, and then deriving their joint and marginal sampling kernels respectively. For those whose kernel has an analytical form, we show how to generate its conjugate posterior density and perform direct sampling to simulate *i.i.d* draws. However, for others, where non-conjugacy is presented, a numerical integration-based Griddy Gibbs sampler is then used and a brief discussion concerning the choice of grid for this sampler is provided.

6.2.1 Component Label

[65] As confirmed in McLachlan and Peel (2000), a two-component mixture distribution is generally flexible enough to capture the stylized factors exhibited in financial time series.

In equation (6.1), since training data y_t is now assumed to be generated from a mixture distribution, it is essential to introduce a latent variable z_t for conducting Bayesian inference so that the current information can be augmented and complete information set can be formed. This technique of introducing component latent variable is very common when the task is to solve missing data problems and estimate mixture distribution. A brief discussion of its uses has already been given in Section 3.4.2 and Section 4.3.5 (see EM algorithm of Dempster *et al.,* 1977, and Data augmentation of Tanner and Wong, 1991, for details). Now, for the purposes of this thesis, we use a dichotomous quantity to form $z_t=(z_1$, $z_{2,}$, ... $z_t)$ so that $z_t=m$ is equal to saying it is the m^{th} component that generates y_t. Given this information, joint posterior density of mixture models then can be defined.

6.2.2 Joint Posterior density

Here, consider an example. If, for a specific observation, say $y^{'}$, its component label $z^{'}$ is now known to be equal to 1, density value of this observation then can be calculated by

$$\begin{aligned} p(y^{'} \mid z^{'}) &\sim \sum \phi\left(\mu_m, \Sigma_{mt} \mid z^{'}=1\right) \\ &\sim \phi(\mu_1, \Sigma_{1t}) \\ &= (2\pi)^{-d/2} \left|\Sigma_{1t}\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(y_t-\mu_1\right)^{'} \Sigma_{1t}^{-1}\left(y_t-\mu_1\right)\right\} \end{aligned} \tag{6.3}$$

provided that Φ is assumed to be MGM distributed. The likelihood function of the whole mixture model can be defined after all observations have been labelled

$$\begin{aligned} l\left(y \mid \varphi, z\right) &\propto \prod_{t\in\{z_t=m\}} \pi_m p\left(y_t \mid \varphi_m, z_t=m\right) \\ &= \prod_{z_t=1}\left[\pi_1 p\left(\mu_1, \Sigma_{1t}\right)\right] \cdot \prod_{z_t=2}\left[(1-\pi_1) p\left(\mu_2, \Sigma_{2t}\right)\right] \end{aligned} \tag{6.4}$$

$$\pi_1 = p(z_t=1); \qquad \pi_2 = p(z_t=2); \qquad \sum \pi_m = 1; \qquad m=1,2$$

Here, $p(\varphi_m \mid y, z)$ denotes the density function of m^{th} mixture component conditioned on the complete information set *(*y, z*).* Since it is known that, according to Bayesian inference,

$$\textit{Posterior distribution} \propto \textit{Prior distribution} \times \textit{likelihood function} \tag{6.5}$$

joint posterior density of φ can be defined

$$\kappa(\varphi \mid y, z) \propto p(\varphi) \times l(y \mid \varphi, z) \tag{6.6}$$

if prior distributions of all parameters, $p(\varphi)$, have also been properly assumed.

Concerning φ, in ADCC-MGM model this parameter set in its simplest form now can be given by

$$\varphi = \{z, \pi, \mu, \varpi, \alpha, \beta, \eta, \varsigma, \iota\} \tag{6.7}$$

where π denotes the weight parameter, μ denotes the mean parameter, ϖ, α, β represent the univariate GARCH parameters used to model individual volatility, and η, ς and ι are ARCH, GARCH and asymmetric parameters controlling the correlation evolving process. Note that z here is not actually a parameter. But this variable is also included in equation (6.7) because of its unoberservablity and the property of also requiring simulation to obtain new updates when Bayesian inference is calculated. In addition, if modelling of individual volatility and modelling of time-varying correlation is allowed to be demarcated, we can also obtain an even more simplified version of (6.7). That is

$$\varphi = \{z, \pi, \mu, \theta, \psi\} \tag{6.8}$$

where $\theta = \{\omega, \alpha, \beta\}$ and $\psi = \{\eta, \varsigma, \iota\}$.

6.2.3 Parameter set of interest φ

Given equation (6.7) and (6.8), it is now clear that φ is actually a large set containing multiple elements. Take the bivariate two-component ADCC-MGM for example: there are a total of 21 elements included in φ which means that, in each iteration of posterior simulation, a total of 21 new draws of $\varphi = \{\varphi_1, \varphi_2, \cdots, \varphi_{21}\}$ where $\varphi \in \Theta$ need to be simulated.[66] Although it is true that these elements can be categorized into just eight different groups and, for those of the same type, their sampling kernels are actually the same, the economic cost of sampling draws for so many (analytical and non-analytical) densities could still be easily accumulated and exceed a staggeringly high level very quickly. Thus, a proper trimming of this parameter set is usually desirable.

To perform this task, we impose some parameter restrictions here. For example, in this thesis we respectively let $\mu_2 = -(\pi_1 / \pi_2)\mu_1$ and $\pi_1 > \pi_2$.[67] The first constraint is imposed

[66] In Chapter 7, we will illustrate a specific method to index different element in φ.

[67] Since we are now considering a two-component mixture model here, mean parameter of the whole

to ensure that the weighted average of means in mixture distribution equals zero so that the means of second component distribution can be calculated analytically once the mean of first component is obtained. Using the second restriction is to avoid the label-switching problem. Since each mixture is now allowed to have only two components, we do not need to sample π_2 in each loop. Its value can be readily computed by $\pi_2 = 1 - \pi_1$ once an updated value for π_1 is obtained.

6.2.4 Settings for Griddy Gibbs sampler

In equation (6.5), we have defined the posterior simulation kernel for φ. Now, it is important to choose a 'right' sampling technique. Specifically, in this research, if a resultant kernel belongs to a known distributional type, that is, its density function is analytical, we simulate random draws for this kernel using direct sampling technique. However, for most others not having such forms, Griddy-Gibbs sampler is then used. As for this MCMC simulator, in Section 4.3.4 a detailed illustration has already been given. However, here some necessary settings concerning its implementation are still worth mentioning.

a. Determination of the grid points

First, to use Griddy Gibbs sampler, we need to determine the number of points to be input and values of points to be accessed. Concerning the first issue, although it is certain that the more points included the less bias will be introduced to the numerical evaluation of *c.d.f*, the computational cost of implementing a massive-point grid is usually very high for multi-dimensional problems. Empirically, how many points are really enough to run Griddy Gibbs sampler both efficiently and accurately is still an open question. Bauwens and Lumbrano (1998) chose 33 points after making a performance comparison with the results generated from using 17-point grid and those using 65-point grid. Galeano and Ausin (2005) argued that a 40-point grid was enough for their research purposes. Here,

mixture, μ, can be decomposed into two parts μ_1 and μ_2. Since the training data is now assumed to be bivariate, both μ_1 and μ_2 are (2×1) vectors and their combination is a (2×2) matrix. Here, μ_1 denotes the mean vector of the first component distribution which has two elements corresponding to each dimension in the bivariate data respectively.

although various criteria are adopted, it is important to note that most of the previous research conducted on this issue investigated only univariate models, and the number of parameters in these models did not exceed ten. However, in our case, not only does the parameter set now contain more than twenty elements, the updating scheme is also complicated by ADCC specification. Thus, an immediate drawback of this sophistication is the difficulty of evaluating any problems related to the grid points in simulation. For example, calculating the Bayesian inference for ADCC-MGM using a 4000-observation sample on a modern *Intel P4* processor needs at least 7 minutes per iteration if we choose the 40-point grid for evaluating integral. To achieve the convergence (usually, at least 2000 iterations after imposing some ideal conditions), it will take more than 9 days, or even weeks. Such a long calibration process is obviously too expensive for industrial uses of correlation models for daily valuation and risk management purpose. To circumvent this difficulty, we thus abandon the traditional strategy of including a large quantity of points in each grid and turn to find experimental data that could be trained properly so that simulated Markov chains, once generated, can have a quick convergence.[68]

As for the determination of values for these points, we choose the fixed grid of equidistant points for each parameter so that a smooth estimation of the marginal posterior density can be achieved.[69] Concretely, we set an upper and lower bound for each parameter and choose 30 equally-spaced points within the interval constructed by these bounds. For a more detailed illustration of these settings, see Section 6.3 and Section 6.4 for their applications in ADCC-MGM and ADCC-MTM models respectively.

b. Integration rule and Interpolation technique

With respect to the integration technique, we use trapezoidal method in this thesis. Say that, for a parameter ρ , if its grid points ρ_i and corresponding density values $\kappa(\rho \mid \varphi_{-\rho}^{(n)}, y, z^{(n+1)})$ now have all been generated, we divide the area under

[68] In this research, we include 30 points in each grid.

[69] Here, we can also choose a variable grid which can be modified to have more points on masses where posterior distribution is concentrated.

$\kappa(\rho \mid \varphi_{-\rho}^{(n)}, y, z^{(n+1)})$ into S strips, each with width h=(*upperbound* - *lowerbound*)/S, and approximate the shape of each strip using a trapezium rule and sum these results up. By so doing, the sampling kernel is then evaluated on a finite number of integrands and *c.d.f* of this kernel can be calculated by,

$$\Phi_i \approx \sum_{s=1}^{S} (h/2) * \left(\kappa\left(\rho_1 + (s-1)h \mid y, z\right) + \kappa\left(\rho_1 + sh \mid y, z\right)\right) \tag{6.9}$$

Here, although we can also use other methods such as adaptive Simpson and Lobotto quadratures to perform the same task, these methods work in a similar way to the trapezoidal rule except that the integrand is approximated using a quadratic function rather than a straight line within each subinterval. Since the computational cost is now a major issue, we thus adopt only the simplest method for discrete integration. Besides, for the same reason, when interpolation between two adjacent points is required, we fit only linear function.

6.3 Posterior simulation of ADCC-MGM model

Now, we describe the posterior sampling procedure for bivariate two-component ADCC-MGM model. First, after replacing the distribution function $p(\cdot \mid y, z)$ in equation (6.2) with a Gaussian density, we obtain the joint posterior density of ADCC-MGM. That is,

$$\begin{aligned} \kappa(\varphi \mid y, z) &= p(\varphi) \cdot l(y \mid \varphi, z) \\ &= p(\varphi) \prod_{t \in \{z_t = m\}} \pi_m \phi(y \mid \varphi_m, z) \\ &= p(\varphi) \prod_{t: z_t = 1} \left[\pi_1 \phi(\mu_1, \Sigma_{1t})\right] \cdot \prod_{t: z_t = 2} \left[(1-\pi_1) \phi(\mu_2, \Sigma_{2t})\right] \\ \phi(\mu_m, \Sigma_{mt}) &= (2\pi)^{-d/2} \left|\Sigma_{mt}\right|^{-1/2} \exp\left\{-\frac{1}{2}(y_t - \mu_m)' \Sigma_{mt}^{-1} (y_t - \mu_m)\right\} \end{aligned} \tag{6.10}$$

6.3.1 Prior density assumption

Next, we assume a proper prior density for each element in φ so as to obtain their marginal posteriors. Here, although a clever choice might be made, one usually finds it very difficult to derive an analytical solution for sampling a specific parameter. For example, in this research sampling kernels of most parameters do not have an analytical form. This is because joint posterior is now a very complicated function due to the incorporation of both mixture models and ADCC specification. Given such sophistication, it does not seem very practical to expect much prior information before sampling is really performed. In such cases, a natural solution is then to assume uninformative priors (see Geweke, 1992, Van Dijk, 1993, and many others for examples). The advantage of making this choice is that the density value of these priors is constant (or approximately constant) which can be omitted when calculating the marginal posteriors from equation (6.10). Therefore, except for the mean parameter μ and weight parameter π, whose sampling kernels have analytical forms, prior densities of all other elements in φ are all assumed to be uniformly distributed. In addition, we assume these densities to be independent of one another so that their joint density can be written as

$$p(\varphi) = p(\pi) \cdot p(\mu) \cdot p(\theta) \cdot p(\psi) \tag{6.11}$$

6.3.2 Joint posterior density of θ and ψ

Here, as for the elements in volatility parameter set θ and correlation parameter set ψ, since their joint prior is now approximately constant,

$$p(\theta,\psi) = p(\theta)p(\psi) \propto C \tag{6.12}$$

To obtain their joint posterior, one only needs to rewrite the function (6.10) by absorbing $p(\theta,\psi)$ and eliminating all elements not related to θ and ψ.[70] That is

$$\begin{aligned}
\kappa(\theta,\psi \mid y,z) &\propto p(\varphi)l(\varphi \mid y,z) \\
&= p(\theta,\psi)l(\theta,\psi \mid y,z) \\
&= l(\theta,\psi \mid y,z) \\
&= \prod_{t\in\{z_t=m\}} \pi_m \phi(\theta_m,\psi_m \mid y_i,z_i) \qquad m=(1,\ 2) \\
&= \prod_{t:z_t=1}^{T} \pi_1 (2\pi)^{-d/2} \left|\Sigma_{1t}\right|^{-1/2} \exp\left\{-\frac{1}{2}(y_t-\mu_1)^{'}\Sigma_{1t}^{-1}(y_t-\mu_1)\right\} \\
&\quad \times \prod_{t:z_t=2}^{T} (1-\pi_1)(2\pi)^{-d/2} \left|\Sigma_{2t}\right|^{-1/2} \exp\left\{-\frac{1}{2}(y_t-\mu_2)^{'}\Sigma_{2t}^{-1}(y_t-\mu_2)\right\}
\end{aligned} \tag{6.13}$$

Note that $\kappa(\theta,\psi \mid y,z)$ is now a function of both mean parameters μ and covariance matrix Σ. However, in Engle's standard DCC, Σ_t can also be decomposed into two parts if a so-called 'parameter separation method' is adopted. Briefly, the idea is to separate volatility parameters and correlation parameters into different likelihood functions and

[70] Normalized constant can be absorbed in posterior kernel because it will not affect the updated information.

estimate them separately.[71] Since the conditional heteroskedasticity of innovations can be modelled by $\Sigma_{mt} = D_{mt} R_{mt} D_{mt}$, another version of (6.13) can be obtained after normalization constant $(2\pi)^{-d/2}$ is eliminated. That is,

$$\begin{aligned}\kappa(\theta,\psi \mid y,z) = \prod_{t:z_t=1}^{T} \pi_1 \left|D_{1t}R_{1t}D_{1t}\right|^{-1/2} \exp\left\{-\frac{1}{2}(y_t-\mu_1)^{'}\left(D_{1t}R_{1t}D_{1t}\right)^{-1}(y_t-\mu_1)\right\} \times \\ \prod_{t:z_t=2}^{T}(1-\pi_1)\left|D_{2t}R_{2t}D_{2t}\right|^{-1/2} \exp\left\{-\frac{1}{2}(y_t-\mu_2)^{'}\left(D_{2t}R_{2t}D_{2t}\right)^{-1}(y_t-\mu_2)\right\}\end{aligned} \qquad (6.14)$$

Besides, since the demeaned return in (6.14) can be written as $D_{mt}^{-1}\left(y_t-\mu_m\right) = \varepsilon_{mt}$, an even more simplified posterior for θ and ψ can be derived. That is,

$$\begin{aligned}\kappa\left(\theta,\psi \mid y,z\right) = \prod_{t:z_t=1}^{T} \pi_1 \left|D_{1t}\right|^{-1} \left|R_{1t}\right|^{-1/2} \exp\left\{-\frac{1}{2}\varepsilon_{1t}^{'} R_{1t}^{-1} \varepsilon_{1t}\right\} \times \\ \prod_{t:z_t=2}^{T}(1-\pi_1)\left|D_{2t}\right|^{-1} \left|R_{2t}\right|^{-1/2} \exp\left\{-\frac{1}{2}\varepsilon_{2t}^{'} R_{2t}^{-1} \varepsilon_{2t}\right\}\end{aligned} \qquad (6.15)$$

Above, we have presented two alternative kernels for sampling θ and ψ . Now, it is necessary to make a choice between them. Note that, although the decomposition of covariance matrix can provide a computationally cheaper way for estimating DCC models using classical inferential method (ML), it does not help to ease the burden when Bayesian inference is calculated. This is because, when Σ is decomposed, sampling volatility parameters and sampling correlation parameters both require computing two different functions. Since the generation of Markov chain is based on a homogeneous loop, this computation needs to be performed every time the integral is evaluated. However, when equation (6.13) is adopted, the function needing to be updated is only Σ on each loop because all elements in θ and ψ are now encompassed, thus, as far as the computational cost is concerned, using this method is relatively cheaper. However, one disadvantage is the effectiveness of the searching for random draws in these two sets' high-probability region then might be somewhat affected because intercorrelation between their generated

[71] Newey and MacFadden (1994) provided the theoretical proof for the robustness of two-stage GMM estimator. By exploiting their result, Engle (1999) derived the consistency and asymptotic normality for his two-stage DCC estimators.

chains is neglected. Usually, since this influence will not affect the unbiasedness of the resultant chain, in the research we use equation (6.13) to simulate θ and ψ .

6.3.3 Posterior sampling sequence

Given the joint posterior density defined in equation (6.10), we now devise a fixed sampling sequence for simulating new updates. Here, we start the posterior sampling of φ from the latent variable z because only after this information has been obtained can likelihood function of mixture models be defined and sampling kernels of all other elements in φ be generated. Once z is updated, new draws of weight parameter π and mean parameter μ are then simulated from two analytical functions, and elements in volatility parameter set θ and correlation parameter set ψ are sampled from a non-conjugate kernel. Here, to see the details of how to perform Griddy Gibbs sampler for this correlation mixture model, we present its sampling sequence for generating *N*-state Markov chains in the following.

1. Let *n=0* be the first state of chain and set the initial value of φ to be $\varphi^{(0)}$

2. Draw state variable $z^{(n+1)}$ from kernel $\kappa\left(z \mid \mu^{(n)},\pi^{(n)},\theta^{(n)},\psi^{(n)},y\right)$

3. Draw weight parameter $\pi^{(n+1)}$ from kernel $\kappa\left(\pi \mid z^{(n+1)},\mu^{(n)},\theta^{(n)},\psi^{(n)},y\right)$

4. Draw mean parameters $\mu^{(n+1)}$ from kernel $\kappa\left(\mu \mid z^{(n+1)},\pi^{(n+1)},\theta^{(n)},\psi^{(n)},y\right)$

5. Draw volatility parameters $\theta^{(n+1)}$ from kernel $\kappa\left(\theta \mid z^{(n+1)},\pi^{(n+1)},\mu^{(n+1)},\psi^{(n)},y\right)$

6. Draw correlation parameters $\psi^{(n+1)}$ from kernel $\kappa\left(\psi \mid z^{(n+1)},\pi^{(n+1)},\mu^{(n+1)},\theta^{(n+1)},y\right)$

7. Let *n=n*+1 and go to 2 until *n=N*

Above, we present a fixed sequence for sampling φ. However, it is worth noting that this method is not the only way to generate Markov chains which can satisfy the convergence theorem. Since the homogeneity condition only requires the sampling sequence, once simulation starts, does not change, it is then possible for us to devise another way which is different from above to generate the same posterior result. However, such alternatives, though available, still need to be initialised by the sampling of component label variable.

6.3.4 Sampling kernel of each parameter in φ

We now illustrate how to derive the sampling kernel for each parameter in φ. Before proceeding, two things need to be noted. First, according to Griddy Gibbs sampler, since, for any parameter, its sampling kernel is required to be a sole function of this parameter, all elements (or parameters) not related to this parameter is then eliminated in resultant posterior density. Second, in some cases, for a single parameter one may find more than one suitable kernel (i.e. for μ, θ and ψ parameters in our mixture models). This is because different priors might now be used for generating posteriors, or their sampling kernels themselves can be further decomposed. In the following, we illustrate the derivation of each kernel according to the sampling sequence provided in the last subsection.

a. Sampling z from $\kappa\left(z \mid \mu,\pi,\theta,\psi,y\right)$

First, given an *M*-component standard mixture distribution, we sample new updates for *z* by calculating the conditional posterior probability of each component in the mixture, followed by the simulation of a time series whose proportion of observations that belongs to each component corresponds to previous probability. As stated early, since component label variable *z* is not a parameter, we do not need to make any prior distributional assumption for it. However, as a necessary condition, mutual independence of the random draws still needs to be ensured when new updates are generated.

To perform its simulation, an example is given below. Assume that the current state of Markov chain is $\varphi^{(n)}$ and our purpose is to generate a new update for component label of the next state $z^{(n+1)}$. First we calculate the conditional probability of y_t being generated from m^{th} Gaussian component by

$$p(z_t^{(n+1)} = m \mid \varphi^{(n)}, y) = \frac{\pi_m \phi(y_t \mid \mu_m, \theta_m, \psi_m)}{\sum_{i=1}^{M} \pi_m \phi(y_t \mid \mu_m, \theta_m, \psi_m)} \tag{6.16}$$

so that this probability corresponding to the first component is

$$p(z_t^{(n+1)}=1)=\frac{\pi_1|\Sigma_{1t}|^{-1/2}\exp\left\{-\frac{1}{2}(y_t-\mu_1)^{'}\Sigma_{1t}^{-1}(y_t-\mu_1)\right\}}{\pi_1|\Sigma_{1t}|^{-1/2}\exp\left\{-\frac{1}{2}(y_t-\mu_1)^{'}\Sigma_{1t}^{-1}(y_t-\mu_1)\right\}+(1-\pi_1)|\Sigma_{2t}|^{-1/2}\exp\left\{-\frac{1}{2}(y_t-\mu_2)^{'}\Sigma_{2t}^{-1}(y_t-\mu_2)\right\}}$$

(6.17)

and that of the second can be computed by $1-p(z_t^{(n+1)}=1)$. Then, to obtain $z^{(n+1)}$, we simply simulate a series of random draws from a binomial distribution with its parameter set to be $p(z_t^{(n+1)}=1)$ to get the updated information on component label.

b. Sampling π from $\kappa(\pi\,|\,\mu,\theta,\psi,y,z)$

Once the updated information on *z* has been obtained, for all remaining elements in φ, its sampling kernel can be defined after all elements not related to this parameter are eliminated in joint posterior density. Take the weight parameter π for example: its kernel now can be easily written as

$$\begin{aligned}\kappa(\pi\,|\,\varphi_{-\pi}^{(n)},z^{(n+1)},y) &= p(\varphi)l(\varphi\,|\,y,z^{(n+1)})\\ &= p(\pi)\prod_{m=1}^{M}\pi_m^{T_m} \qquad (6.18)\\ &= p(\pi)\pi_1^{T_1}\left(1-\pi_1\right)^{T_2}\end{aligned}$$

where *T₁*, *T₂* denotes the number of observations generated by the first and the second Gaussian component in the mixture.

At first glance, one might have found that density function of (6.18) actually looks very similar to a binomial distribution. Thus, an important result concerning the use of this density in Bayesian inference can be exploited. Concretely, according to a famous Bayesian theorem that Dirichlet is conjugate to the multinomial observations, if the prior density $p(\pi)$ is now assumed to be Dirichlet distributed, its posterior $\kappa(\pi\,|\,\varphi_{-\pi}^{(n)},z^{(n+1)},y)$ then will also be distributing like a Dirichlet.

To illustrate this result through an example, we now let $p(\pi)\sim Dir(a_1,a_2)$, that is

$$p(\pi)=\frac{\Gamma(\sum_{m=1}^{2}\alpha_m)}{\prod_{m=1}^{2}\Gamma(\alpha_m)}\prod_{m=1}^{2}\pi_m^{\alpha_i-1} \qquad \alpha_m>0; m=1,2 \tag{6.19}$$

Thus, its corresponding posterior is just

$$\kappa\left(\pi^{(n+1)}\mid\varphi_{-\pi}^{(n)},z^{(n+1)},y\right)=\frac{\Gamma(\sum_{m=1}^{2}\alpha_m)}{\prod_{m=1}^{2}\Gamma(\alpha_m)}\prod_{m=1}^{2}\pi_m^{\alpha_m-1}\prod_{m=1}^{2}\pi_m^{T_m} \tag{6.20}$$

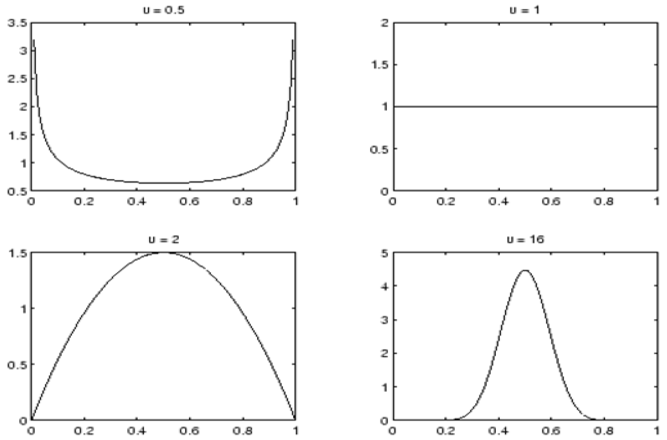
after (6.19) is inserted into (6.18). And this density can be further simplified to

$$\begin{aligned}\kappa\left(\pi^{(n+1)}\mid\varphi_{-\pi}^{(n)},z^{(n+1)},y\right)&\propto\prod_{m=1}^{2}\pi_m^{\alpha_m-1}\prod_{m=1}^{2}\pi_m^{T_m}\\&=\prod_{m=1}^{2}\pi_m^{\alpha_m+T_m-1}\end{aligned} \tag{6.21}$$

after all normalization constants are eliminated. It is clear that the posterior density now corresponds to another Dirichlet distribution, that is *Dir*(a_1+T_1, a_2+T_2).

Here, since T_1 and T_2 can be obtained after all component labels have been updated, we can easily generate a new update for $\pi_m^{(n+1)}$ once its prior distribution $p(\pi)$ has been properly specified. In the following, we selectively pick four Dirichlet densities as candidates for $p(\pi)$ and present their density shapes

Density estimates of *Dir*(*u*, *u*)



According to the above graphs, it is obvious *Dir*(1,1) is the only Dirichlet that can give equal weights to all values on the parameter space. Since, in empirical Bayesian learning, it is difficult to obtain early knowledge on posterior density shape of a parameter, using this uninformative prior is thus proper for our simulation purpose.

As for the sampling of posterior, which is now set to be *Dir*(*1*+T_1, *1*+T_2), we follow the traditional procedure suggested in Wilks (1962). Simply put, after simulating two independent Gamma variables c_m=(c_1, c_2), one from *Gamma*(*1*+T_1, *1*) and one from *Gamma*(*1*+T_2, *1*), we obtain a new update for $\pi_m^{(n+1)}$ using

$$\pi_m^{(n+1)} = \frac{c_m}{\sum_{m=1}^{2} c_m} \tag{6.22}$$

c. Sampling μ from $\kappa\left(\mu \mid \pi, \theta, \psi, y, z\right)$

Now, we illustrate the sampling of mean parameter. Coincidently, this task can also be performed in an analytical way and we can exploit a famous Bayesian conjugate result here. That is, the posterior density of the mean of a Gaussian distribution after assuming a Gaussian prior will also be Gaussian disturbed. Recall that, in preliminary settings, we have imposed a restriction, that is $\mu_2 = -(\pi_1 / \pi_2)\mu_1$. Thus, simulation of the whole mean parameters for mixture models can be resolved by only sampling the means of the first component μ_1 and then obtaining those of the second μ_2 through the updated π.

Here, to simulate μ_1, first we assume an arbitrary Gaussian prior $p(\mu_1) \sim \phi\left(\mu_{1*}, \Sigma_{1*}\right)$. Then, its corresponding posterior density is obtained

$$\begin{aligned} \kappa(\mu_1^{(n+1)} \mid \varphi_{-\mu_1}^{(n)}, \pi^{(n+1)}, z^{(n+1)}, y) &= p(\mu_1) l(y \mid \varphi_1, z^{(n+1)}) \\ &= \phi(\mu_{1*}, \Sigma_{1*}) \prod_{t \in \{z_t = 1\}} \pi_1 \phi(\mu_1, \Sigma_{1t}) \\ &\propto \phi\left\{ f(\Sigma_{1t}, \mu_{1*}, \Sigma_{1*}), \upsilon(\Sigma_{1t}, \Sigma_{1*}) \right\} \end{aligned} \tag{6.23}$$

after all elements not related to μ_1 are eliminated. Note that Σ_{1t} here denotes the time-varying conditional covariance generated by observations belonging to the first mixture component; $f(\Sigma_{1t}, \mu_{1*}, \Sigma_{1*})$ and $\upsilon(\Sigma_{1t}, \Sigma_{1*})$ represent the mean and covariance matrix of a new multivariate Gaussian.

To see the proof of how to obtain this density, first we expand the function (6.23) to

$$\begin{aligned}\kappa(\mu_1^{(n+1)} \mid \cdot) &= \phi(\mu_{1*}, \Sigma_{1*}) \prod_{i \in \{z_t = 1\}} \pi_1 \phi(\mu_1, \Sigma_{1t}) \\ &= (2\pi)^{-d/2} \left|\Sigma_{1*}\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(\mu_1 - \mu_{1*}\right)' \Sigma_{1*}^{-1}\left(\mu_1 - \mu_{1*}\right)\right\} \times \qquad (6.24) \\ &\prod_{t:z_t=1}^{T} \pi_1 (2\pi)^{-d/2} \left|\Sigma_{1t}\right|^{-1/2} \exp\left\{-\frac{1}{2}\left(y_t - \mu_1\right)' \Sigma_{1t}^{-1}\left(y_t - \mu_1\right)\right\}\end{aligned}$$

and then absorb the terms that do not depend on μ_1 into the constant of proportionality so that the resulting posterior can be integrated to one. Thus, the sampling kernel becomes

$$\kappa(\mu_1^{(n+1)} \mid \cdot) \propto \exp\left\{-\frac{1}{2}\left[\sum_{t \in \{z_t = 1\}}^{T} \Sigma_{1t}^{-1}(y_t - \mu_1)^2 + \Sigma_{1*}^{-1}(\mu_1 - \mu_{1*})^2\right]\right\} \qquad (6.25)$$

Now, by letting $\sum_{t:z_t=1}^{T}(y_t - \mu_1)^2 = \sum_{t:z_t=1}^{T}(y_t - \bar{y})^2 + \sum_{t:z_t=1}^{T}(\mu_1 - \bar{y})^2$ where $\bar{y} = \frac{1}{T_1}\sum_{t:z_t=1}^{T} y_t$ we rewrite (6.25) to

$$\kappa(\mu_1^{(n+1)} \mid \cdot) \propto \exp\left\{-\frac{1}{2}\left[\sum_{t:z_t=1}^{T} \Sigma_{1t}^{-1}(y_t - \bar{y})^2 + \sum_{t:z_t=1}^{T} \Sigma_{1t}^{-1}(\mu_1 - \bar{y})^2 + \Sigma_{1*}^{-1}(\mu_1 - \mu_{1*})^2\right]\right\} \qquad (6.26)$$

After eliminating $\sum_{t:z_t=1}^{T} \Sigma_{1t}^{-1}(y_t - \bar{y})^2$ in (6.26) due to its independence from the parameter of interest, the resulting kernel is just

$$\kappa(\mu_1^{(n+1)} \mid \cdot) \propto \exp\left\{-\frac{1}{2}\left[\sum_{t:z_t=1}^{T} \Sigma_{1t}^{-1}(\mu_1 - \bar{y})^2 + \Sigma_{1*}^{-1}\left(\mu_1 - \mu_{1*}\right)^2\right]\right\} \qquad (6.27)$$

Here, note that, after transformation, (6.27) can be written as the addition of a bivariate Gaussian density $\phi\{f(\Sigma_{1t}, \mu_{1*}, \Sigma_{1*}), \upsilon(\Sigma_{1t}, \Sigma_{1*})\}$ and terms that do not depend on μ_1 where

$$f\left(\Sigma_{1t}, \mu_{1*}, \Sigma_{1*}\right) = \frac{\sum_{t:z_t=1}^{T} \Sigma_{1t}^{-1}\bar{y} + \mu_{1*}\Sigma_{1*}^{-1}}{\sum_{t:z_t=1}^{T} \Sigma_{1t}^{-1} + \Sigma_{1*}^{-1}} \qquad \upsilon\left(\Sigma_1, \Sigma_{1*}\right) = \Big(\sum_{t:z_t=1}^{T} \Sigma_{1t}^{-1} + \Sigma_{1*}^{-1}\Big)^{-1} \qquad (6.28)$$

For those that will not affect the posterior information of μ_1, again, we simply eliminate them from the posterior density. Thus, the remaining terms, the simulating kernel of interest, is just a bivariate Gaussian density.

According to the conjugacy result just described, simulating a new random draw for mean parameters of the first component now needs at least three elements: the updated time-varying covariance matrix Σ_{1t}, mean parameter μ_{1*} of $p(\mu_1)$ and covariance matrix Σ_{1*} of $p(\mu_1)$. Here, to choose a proper prior, researchers usually like to pick a very large value for Σ_{1*} so that the resulting density can stretch widely over the parameter space and it is equal to assuming an uninformative prior. Since, in this paper, we are short of enough prior information on μ_1, using this assumption is thus proper for our simulation purpose. However, note that, in Bauwens, Hafner and Rombouts (2006), the authors have suggested another analytical way of sampling means in mixture distribution. Besides, some numerical methods are also sometimes used to perform the same task (see Galeano and Ausin, 2005 for details).

Sampling of volatility parameters and correlation parameters

Now, we return to the discussion of posterior simulation of volatility parameter set θ and correlation parameter set ψ. As illustrated at the beginning of this chapter, since covariance matrix-based sampling kernel has been chosen to simulate draws for elements in these two parameter sets, their joint posterior, which has been shown in (6.13), is then just equal to the likelihood function of standard Gaussian mixture because all their priors are now assumed to be independently and uniformly distributed.

In the following, we describe how to derive and simulate marginals for this joint posterior density. Since the resulting marginals are now to be non-analytical and need to be evaluated using grid-based simulation method, we define a proper upper and lower bound for each grid and make a fine tuning to their bounds so that the search for new draws can be directed to the most relevant areas.

d. Sampling θ from $\kappa\left(\theta \mid \mu,\pi,\psi,y,z\right)$

First, we show how to sample draws for elements in volatility parameter set $\theta = \left\{\varpi,\alpha,\beta\right\}$. We assume the prior of ARCH parameter α and GARCH parameter β to be independently and uniformly distributed on [0, 1] and grid points of intercept parameter ϖ equally spanning on a positive domain from zero to $\hat{\delta}_y^2$.[72] Then, to ensure the stationarity of covariance process, the summation of the ARCH and GARCH parameters of the same state is constrained to be less than one. Thus, adding up these conditions, the four restrictions to be imposed are

$$\alpha \in [0,1], \quad \beta \in [0,1], \quad \varpi \in [0, \hat{\delta}_y^2], \quad \alpha + \beta < 1 \tag{6.29}$$

Here, since the joint posterior density (6.13) has been given, we can easily derive the sampling kernels for $\kappa(\varpi^{(n+1)} \mid \cdot)$, $\kappa(\alpha^{(n+1)} \mid \cdot)$ and $\kappa(\beta^{(n+!)} \mid \cdot)$ respectively. For example, after absorbing the normalization constant $(2\pi)^{-d/2}$ and discarding the unrelated parameter π , $\kappa(\theta^{(n+1)} \mid \cdot)$ can be specified as a multiplication product of likelihood function of two unrelated component densities. That is,

$$\begin{aligned}\kappa(\theta^{(n+1)} \mid \cdot) &= \prod_{t:z_t=m}^{T} |\Sigma_{mt}|^{-1/2} \exp\left\{-\frac{1}{2}(y_t-\mu_m)'\Sigma_{mt}^{-1}(y_t-\mu_m)\right\} \\ &= \prod_{t:z_t=1}^{T} |\Sigma_{1t}|^{-1/2} \exp\left\{-\frac{1}{2}(y_t-\mu_1)'\Sigma_{1t}^{-1}(y_t-\mu_1)\right\} \prod_{t:z_t=2}^{T} |\Sigma_{2t}|^{-1/2} \exp\left\{-\frac{1}{2}(y_t-\mu_2)'\Sigma_{2t}^{-1}(y_t-\mu_2)\right\}\end{aligned} \tag{6.30}$$

Besides, if the decomposition of covariance matrix is allowed, an alternative kernel for sampling θ can also be derived. That is

$$\begin{aligned}\kappa(\theta^{(n+1)} \mid \cdot) &= \prod_{t:z_t=m}^{T} |D_{mt}|^{-1} \exp\left\{-\frac{1}{2}(y_t-\mu_m)' D_{mt}^{-1} R_{mt}^{-1} D_{mt}^{-1}(y_t-\mu_m)\right\} \\ &= \prod_{t:z_t=1}^{T} |D_{1t}|^{-1} \exp\left\{-\frac{1}{2}(y_t-\mu_1)' D_{1t}^{-1} R_{1t}^{-1} D_{1t}^{-1}(y_t-\mu_1)\right\} \times \\ &\quad \prod_{t:z_t=2}^{T} |D_{2t}|^{-1} \exp\left\{-\frac{1}{2}(y_t-\mu_2)' D_{2t}^{-1} R_{2t}^{-1} D_{2t}^{-1}(y_t-\mu_2)\right\}\end{aligned} \tag{6.31}$$

Here, note that, although (6.31) provides an alternative way to generate new updates for θ, we do not use it in our simulation due to the computational cost concerns. This kernel is presented here only for the completeness of analysis. For a more detailed illustration of the reason for abandoning its use, see Section 6.2.2.

[72] $\hat{\delta}_y^2$ here denotes the unconditional variance of training data y_t, it is set here as the upper bound of ϖ .

e. Sampling ψ from $\kappa\left(\psi \mid \mu,\pi,\theta,y,z\right)$

As for the sampling of correlation parameters $\psi = \{\eta,\varsigma,\iota\}$, the same strategy as that used for sampling volatility parameters is adopted here. Prior densities of all elements in ψ are assumed to be uniformly and independently distributed and we impose the restrictions

$$\eta \in [0,1], \quad \varsigma \in [0,1], \quad \iota \in [0,1], \quad \eta^2 + \varsigma^2 + \iota^2 < 1 \tag{6.32}$$

so that random draws of these correlation parameters can be drawn from the most relevant space and resulting covariance process is stationary.

To derive each marginal posterior, if the covariance matrix is now the only function that needs to be updated, $\kappa(\eta^{(n+1)} \mid \cdot)$, $\kappa(\varsigma^{(n+1)} \mid \cdot)$ and $\kappa(\iota^{(n+!)} \mid \cdot)$ can be written as

$$\begin{aligned} \kappa(\psi^{(n+1)} \mid \cdot) &= \prod_{t:z_t=m}^{T} |\Sigma_{mt}|^{-1/2} \exp\left\{-\frac{1}{2}(y_t-\mu_m)'\Sigma_{mt}^{-1}(y_t-\mu_m)\right\} \\ &= \prod_{t:z_t=1}^{T} |\Sigma_{1t}|^{-1/2} \exp\left\{-\frac{1}{2}(y_t-\mu_1)'\Sigma_{1t}^{-1}(y_t-\mu_1)\right\} \prod_{t:z_t=2}^{T} |\Sigma_{2t}|^{-1/2} \exp\left\{-\frac{1}{2}(y_t-\mu_2)'\Sigma_{2t}^{-1}(y_t-\mu_2)\right\} \end{aligned} \tag{6.33}$$

However, if the decomposition technique is allowed, $\psi^{(n+1)}$ then becomes

$$\begin{aligned} \kappa(\psi^{(n+1)} \mid \cdot) &= \prod_{t:z_t=m}^{T} |R_{mt}|^{-1/2} \exp\left\{-\frac{1}{2}\varepsilon_{mt}' R_{mt}^{-1} \varepsilon_{mt}\right\} \\ &= \prod_{t:z_t=1}^{T} |R_{1t}|^{-1/2} \exp\left\{-\frac{1}{2}\varepsilon_{1t}' R_{1t}^{-1} \varepsilon_{1t}\right\} \times \prod_{t:z_t=2}^{T} |R_{2t}|^{-1/2} \exp\left\{-\frac{1}{2}\varepsilon_{2t}' R_{2t}^{-1} \varepsilon_{2t}\right\} \end{aligned} \tag{6.34}$$

after all terms not depending on R_{mt} are eliminated.

6.4 Posterior simulation of ADCC-MTM model

Apart from the Gaussian mixture, a more flexible way to account for the skewness and leptokurtosis that are frequently presented in the financial time series using a finite mixture model, given a limited number of components, is to utilize a multivariate T mixture. In (6.1) and (6.2), we have already presented the specification of ADCC-MTM. Now, to calculate its inference, we illustrate its posterior sampling procedure bleow.

6.4.1 Joint posterior density

Consider a series of *d*-dimensional *M*-component multivariate T mixture distributed observations y_t whose density is

$$
\begin{aligned}
& y_t \mid F_{t-1} \sim \sum_{m=1}^{M} \pi_m t\left(\mu_m, \Sigma_m, v_m\right) \\
& t\left(\mu_m, \Sigma_m, v_m\right) = \frac{\Gamma\left(\dfrac{v_m + d}{2}\right)}{\left(\pi \cdot v_m\right)^{d/2} \Gamma\left(\dfrac{v_m}{2}\right)} \left|\Sigma_{mt}\right|^{-1/2} \left(1 + \frac{(y_t - \mu_m)' \Sigma_{mt}^{-1} (y_t - \mu_m)}{v_m}\right)^{-\frac{v_m + d}{2}}
\end{aligned}
\qquad (6.35)
$$

where μ_m, Σ_m, v_m are mean, covariance and degree of freedom parameters of m^{th} T component.[73] Since y_t is now assumed to be generated from a bivariate two-component mixture distribution, *d* and *M* are both set equal to two and we obtain

$$
y_t \mid F_{t-1} \sim \sum_{m=1}^{2} \frac{\Gamma\left(\dfrac{v_m + 2}{2}\right) \pi_m}{\pi v_m \Gamma\left(\dfrac{v_m}{2}\right)} \left|\Sigma_{mt}\right|^{-1/2} \left(1 + \frac{(y_{\mathrm{t}} - \mu_m)' \Sigma_{mt}^{-1} (y_t - \mu_m)}{v_{\mathrm{m}}}\right)^{-\frac{v_m + 2}{2}} \qquad (6.36)
$$

Here, concerning the above function, one thing needs to be noted before we proceed further to derive the likelihood function of the whole mixture model and joint posterior density for parameter set of interest. Since it is known that, for any gamma function $\Gamma(.)$, it satisfies $\Gamma(x) = x\Gamma(x)$, (6.36) then can be rewritten to

$$
\begin{aligned}
& y_t \mid F_{t-1} \sim \sum_{m=1}^{2} \frac{\pi_m}{2\pi} \left|\Sigma_{mt}\right|^{-1/2} \left(1 + \frac{(y_{\mathrm{t}} - \mu_m)' \Sigma_{mt}^{-1} (y_t - \mu_m)}{v_{\mathrm{m}}}\right)^{-\frac{v_m + 2}{2}} \\
& \textit{due to} \quad \Gamma\left(\frac{v_m + 2}{2}\right) = \Gamma\left(\frac{v_m}{2} + 1\right) = \frac{v_m}{2} \Gamma\left(\frac{v_m}{2}\right)
\end{aligned}
\qquad (6.37)
$$

The likelihood function of *t* mixture can be derived after all current information on component label variable is obtained. That is,

[73] In some textbooks, $\left(\pi \cdot v_m\right)^{d/2}$ presented in the denominator of equation (6.35) is written as $\left[\pi(v_m - 2)\right]^{d/2}$ and $\left(1 + \frac{f(y_{\mathrm{t}} \mid \mu_m, \Sigma_m)}{v_{\mathrm{m}}}\right)^{-\frac{v_m + d}{2}}$ is replaced by $\left(1 + \frac{f(y_{\mathrm{t}} \mid \mu_m, \Sigma_m)}{v_{\mathrm{m}} - 2}\right)^{-\frac{v_m + d}{2}}$. This setting is imposed to ensure the value of degree of freedom parameter larger than two.

$$
\begin{aligned}
l\left(y \mid \varphi, z\right) &\propto \prod_{t\in\{z_t=m\}} \pi_m t\left(\mu_m, \Sigma_m, \nu_m\right) \\
&= \prod_{t:z_t=1}^{T} \pi_1 \left|\Sigma_{1t}\right|^{-1/2} \left(2\pi\right)^{-1} \left(1+\frac{(y_t-\mu_1)^{'}\Sigma_{1t}^{-1}(y_t-\mu_1)}{\nu_1}\right)^{-\frac{\nu_1+2}{2}} \times \\
&\quad \prod_{t:z_t=2}^{T} \pi_2 \left|\Sigma_{2t}\right|^{-1/2} \left(2\pi\right)^{-1} \left(1+\frac{(y_t-\mu_2)^{'}\Sigma_{2t}^{-1}(y_t-\mu_2)}{\nu_2}\right)^{-\frac{\nu_2+2}{2}}
\end{aligned}
\qquad (6.38)
$$

$$
m=1,2; \quad \pi_m = p(z_m=1); \quad \sum \pi_m = 1; \quad \pi_m \in [0,1]
$$

Besides, if the decomposition of covariance matrix is allowed, we can also derive another form of (6.38) after Σ_{mt} is replaced by $D_{mt}R_{mt}D_{mt}$

$$
\begin{aligned}
l\left(y \mid \varphi, z\right) &= \prod_{t:z_t=1}^{T} \pi_1 \left|D_{1t}R_{1t}D_{1t}\right|^{-1/2} \left(2\pi\right)^{-1} \left(1+\frac{(y_t-\mu_1)^{'}(D_{1t}R_{1t}D_{1t})^{-1}(y_t-\mu_1)}{\nu_1}\right)^{-\frac{\nu_1+2}{2}} \times \\
&\quad \prod_{t:z_t=2}^{T} (1-\pi_1) \left|D_{2t}R_{2t}D_{2t}\right|^{-1/2} \left(2\pi\right)^{-1} \left(1+\frac{(y_t-\mu_2)^{'}(D_{2t}R_{2t}D_{2t})^{-1}(y_t-\mu_2)}{\nu_2}\right)^{-\frac{\nu_2+2}{2}}
\end{aligned}
\qquad (6.39)
$$

and

$$
\begin{aligned}
l\left(y \mid \varphi, z\right) &= \prod_{t:z_t=1}^{T} \pi_1 \left|D_{1t}\right|^{-1} \left|R_{1t}\right|^{-1/2} \left(2\pi\right)^{-1} \left(1+\frac{\varepsilon_{1t}^{'}R_{1t}^{-1}\varepsilon_{1t}}{\nu_1}\right)^{-\frac{\nu_1+2}{2}} \times \\
&\quad \prod_{t:z_t=2}^{T} (1-\pi_1) \left|D_{2t}\right|^{-1} \left|R_{2t}\right|^{-1/2} \left(2\pi\right)^{-1} \left(1+\frac{\varepsilon_{2t}^{'}R_{2t}^{-1}\varepsilon_{2t}}{\nu_2}\right)^{-\frac{\nu_2+2}{2}}
\end{aligned}
\qquad (6.40)
$$

if $D_{mt}^{-1}\left(y_t-\mu_m\right)$ is replaced by ε_{mt}.

The joint posterior density of φ can be defined after all priors have been properly specified.

$$
\begin{aligned}
\kappa(\varphi \mid y, z) &= p(\varphi)\cdot l(y \mid \varphi, z) \\
&= p(\varphi) \prod_{i\in\{z_i=m\}} \pi_m t\left(y_t \mid \varphi_m, z_t\right)
\end{aligned}
\qquad (6.41)
$$

Note that, in (6.41), parameter set of interest φ now contains a total of 23 elements. That is $\varphi = \left\{z, \pi_m, \mu_m, \theta_m, \psi_m, \nu_m; m=1,2\right\}$ where θ and ψ still represent two subsets corresponding to the volatility parameters and correlation parameters respectively.

6.4.2 Prior distributions assumption

In this section, we give the prior distributional assumption for each parameter in ADCC-MTM. As for this issue, most of the settings used in the previous sampling of ADCC-MGM are retained here. For example, the weight parameter π is still associated with a Dirichlet prior so that its posterior can be simulated analytically. The priors of θ and ψ are set to be uniformly distributed since no information is obtainable at the initial stage on the density shape and likely values of their posterior draws. Meanwhile, several changes also need to be mentioned. For instance, the prior of μ is no longer assumed to be Gaussian but uniformly distributed due to the appearance of a non-analytical kernel. As for the degree of freedom parameter, extra care then needs to be taken when its prior is assumed. When we consider the behaviour of likelihood function in equation (6.38) with respect to π , θ , ψ and μ , their posteriors are reasonable (integrable) if every covariance Σ_{mt} is kept strictly positive. However, sufficient prior information is needed on v to force its posterior to approach zero quickly enough at the tails in order to be integrable. Concerning this issue, although various priors might now be used, many empirical results show that the density shape of this parameter usually does not have a consistent style. Thus, in this research we use an uninformative density on a finite domain as its prior.

6.4.3 Posterior sampling sequence

Next, we describe the posterior sampling sequence for ADCC-MTM. To perform the simulation for φ, we start by augmenting the existing observations y_t with a latent variable z_t to form a complete information set and then simulating all parameters according to a fixed sequence to ensure the homogeneity of resulting chains. Here, the only difference between this sequence and that of ADCC-MGM is the addition of a sampling kernel for degree of freedom parameter at the end of each loop. That is,

1. Let $n=0$ be the first state of Markov chain and set initial value to be $\varphi^{(0)}$

2. Draw component label variable $z^{(n+1)}$ from kernel $\kappa(z \mid \mu^{(n)}, \pi^{(n)}, \theta^{(n)}, \psi^{(n)}, v^{(n)}, y)$

3. Draw probability measure $\pi^{(n+1)}$ from kernel $\kappa(\pi \mid z^{(n+1)}, \mu^{(n)}, \theta^{(n)}, \psi^{(n)}, v^{(n)}, y)$

4. Draw mean parameters $\mu^{(n+1)}$ from kernel $\kappa(\mu \mid z^{(n+1)}, \pi^{(n+1)}, \theta^{(n)}, \psi^{(n)}, \nu^{(n)}, y)$

5. Draw volatility parameters $\theta^{(n+1)}$ from kernel $\kappa(\theta \mid z^{(n+1)}, \pi^{(n+1)}, \mu^{(n+1)}, \psi^{(n)}, \nu^{(n)}, y)$

6. Draw correlation parameters $\psi^{(n+1)}$ from kernel $\kappa(\psi \mid z^{(n+1)}, \pi^{(n+1)}, \mu^{(n+1)}, \theta^{(n+1)}, \nu^{(n)}, y)$

7. Draw degree of freedom parameter $\nu^{(n+1)}$ from $\kappa(\nu \mid z^{(n+1)}, \pi^{(n+1)}, \mu^{(n+1)}, \theta^{(n+1)}, \psi^{(n+1)}, y)$

8. Le t *n*=*n*+1 and go to 2 until *n*=*N*

6.4.4 Sampling kernel of each parameter in φ

In the following, we illustrate the derivation of simulating kernels for each element in φ according to the sampling sequence given above.

a. Sampling z from $\kappa(z \mid \mu, \pi, \theta, \psi, \nu, y)$

First, as for the sampling of component label variable, we follow the same procedure as that illustrated in the last section. Provided that the current state is $\varphi^{(n)}$, we start by calculating the conditional posterior probability of y_t generated by m^{th} T mixture component at $(n+1)^{th}$ iteration using

$$p(z_t^{(n+1)} = m \mid \varphi^{(n)}, y) = \frac{\pi_m t(y_t \mid \mu_m, \Sigma_m, \nu_m)}{\sum_{i=1}^{M} \pi_m t(y_t \mid \mu_m, \Sigma_m, \nu_m)} \qquad m = 1,2 \tag{6.42}$$

Thus, for the first mixture component, its conditional probability is

$$p(z_t^{(n+1)} = 1) = \frac{\pi_1 |D_{1t}|^{-1} |R_{1t}|^{-1/2} \left(1 + \frac{\varepsilon_{1t}' R_{1t}^{-1} \varepsilon_{1t}}{\nu_1}\right)^{-\frac{\nu_1+2}{2}}}{\pi_1 |D_{1t}|^{-1} |R_{1t}|^{-1/2} \left(1 + \frac{\varepsilon_{1t}' R_{1t}^{-1} \varepsilon_{1t}}{\nu_1}\right)^{-\frac{\nu_1+2}{2}} + \pi_2 |D_{2t}|^{-1} |R_{2t}|^{-1/2} \left(1 + \frac{\varepsilon_{2t}' R_{2t}^{-1} \varepsilon_{2t}}{\nu_2}\right)^{-\frac{\nu_2+2}{2}}} \tag{6.43}$$

and that of second is $1 - p(z_t^{(n+1)} = 1)$. Here, once these two proportional measures have both been updated, we simulate $z^{(n+1)}$ by sampling a new series from a binomial distribution.[74]

b. Sampling π from $\kappa(\pi \mid \mu, \theta, \psi, \nu, y, z)$

[74] Here, the length of this new series is equal to that of y_t and parameter of binomial distribution is set to be $p(z_t^{(n+1)} = 1)$.

With respect to the weight parameter π , its simulation process is now similar to the previous one. After assuming a *Dir*(1, 1) prior for π , we obtain an analytical density for $\kappa(\pi\,|\,\cdot)$. That is, *Dir*(*1*+ *T*$_1$,*1*+ *T*$_2$), or

$$\begin{aligned}\kappa\left(\pi\,|\,\varphi_{-\pi}^{(n)},z^{(n+1)},y\right) &\propto \prod_{m=1}^{2}\pi_m^{\alpha_m-1}\prod_{m=1}^{2}\pi_m^{T_m}\\ &=\prod_{m=1}^{2}\pi_m^{\alpha_m+T_m-1}\end{aligned} \tag{6.44}$$

where *a*$_1$= *a*$_2$=1; *T*$_1$, *T*$_2$ denotes the number of observations generated by the first and second T component in the mixture.

c. Sampling μ from $\kappa(\mu\,|\,\pi,\theta,\psi,\nu,y,z)$

However, when mean parameters are simulated, a different strategy is then adopted. Since the mixture component (multivariate T distribution) now no longer belongs to any exponential distribution family, we cannot use the same conjugate solutions as those documented for sampling ADCC-MGM.[75] But the parameter restrictions imposed before are kept unchanged. For example, the means of two components are still constrained to have a weighted average value equalling zero. That is, in the matrix form we let $\pi^T\mu=0$.[76] Thus, simulation of the whole matrix of mixture model's mean parameters can be resolved by only sampling the mean of the first component and then deriving those of the second analytically.

Since no prior information is available for sampling μ , we use a uniform distribution as its prior. Thus, given the joint posterior density (6.38), its sampling kernel can be derived after all terms not related to μ are eliminated.

$$\begin{aligned}\kappa(\mu_m^{(n+1)}\,|\,\varphi_{-\mu}^{(n)},\pi^{(n+1)},z^{(n+1)},y) &= \prod_{t\in\{z_t=m\}}^{T}\left(1+\frac{(y_t-\mu_m)'\Sigma_{mt}^{-1}(y_t-\mu_m)}{\nu_m}\right)^{-\frac{\nu_m+2}{2}}\\ &=\prod_{t:z_t=1}^{T}\left(1+\frac{(y_t-\mu_1)'\Sigma_{1t}^{-1}(y_t-\mu_1)}{\nu_1}\right)^{-\frac{\nu_1+2}{2}}\times\prod_{t:z_t=2}^{T}\left(1+\frac{(y_t-\mu_2)'\Sigma_{2t}^{-1}(y_t-\mu_2)}{\nu_2}\right)^{-\frac{\nu_2+2}{2}}\end{aligned} \tag{6.45}$$

[75] The conjugate Gaussian prior for the mean parameter is only feasible for the variants in the exponential distribution family such as normal, Laplace, multinomial etc. Since *Student t* distribution is more often categorized as a scaled mixture of normal, it does not belong to the exponential distribution class and thus does not possess the conjugacy for mean.

[76] In a two-component mixture distribution, this restriction is the same as $\mu_2=-(\pi_1/\pi_2)\mu_1$

Besides, to ensure that the resulting density values are large enough to contribute the integral, we set the high-probability region of this parameter to be $(\bar{y}-4\hat{\sigma}_y/\sqrt{T},\bar{y}+4\sigma_y/\sqrt{T})$.

d. Sampling θ and ψ from $\kappa(\theta;\psi\,|\,\mu,\pi,v,y,z)$

To generate the sampling kernels for volatility parameters and correlation parameters, just as before, we do not decompose covariance matrix Σ_{mt} into D_{mt} and R_{mt}. Thus, in each iteration of posterior simulation, only one function needs to be updated. Given that all priors of θ and ψ are now assumed to be independently and uniformly distributed, $\kappa(\theta\,|\,\mu,\pi,\psi,v,y,z)$ and $\kappa(\psi\,|\,\mu,\pi,\theta,v,y,z)$ then have the same density form. That is,

$$\begin{aligned}\kappa(\theta^{(n+1)}\,|\,\cdot)=\kappa(\psi^{(n+1)}\,|\,\cdot)&=\prod_{t\in\{z_t=m\}}^{T}\left|\Sigma_{mt}\right|^{-1/2}\left(1+\frac{(y_{\mathrm{t}}-\mu_m)'\Sigma_{mt}^{-1}(y_{\mathrm{t}}-\mu_m)}{v_{\mathrm{m}}}\right)^{-\frac{v_m+2}{2}}\\&=\prod_{t:z_t=1}^{T}\left|\Sigma_{1t}\right|^{-1/2}\left(1+\frac{(y_{\mathrm{t}}-\mu_1)'\Sigma_{1t}^{-1}(y_{\mathrm{t}}-\mu_1)}{v_1}\right)^{-\frac{v_1+2}{2}}\times\\&\quad\prod_{t:z_t=2}^{T}\left|\Sigma_{2t}\right|^{-1/2}\left(1+\frac{(y_{\mathrm{t}}-\mu_2)'\Sigma_{2t}^{-1}(y_{\mathrm{t}}-\mu_2)}{v_2}\right)^{-\frac{v_2+2}{2}}\end{aligned}\qquad(6.46)$$

Besides, to ensure that the new updates are located in the most relevant high mass, the same restrictions provided in (6.29) and (6.32) are also used here.

e. Sampling v from $\kappa(v\,|\,\mu,\pi,\theta,\psi,y,z)$

Finally, we discuss the derivation of sampling kernel for degree of freedom parameter. To generate its posterior, some reviews concerning the selection of a proper prior for this parameter need to be presented first. Kleibergen and Van Dijk (1993) calculated the Bayesian inference of a *Student-t* GARCH model by using an unrestricted uniform distribution as v 's prior. A similar decision is made in Lin, Lee and Ni (2004), where a fixed interval is imposed onto its space. However, in Geweke (1993), the author challenged the appropriateness of using uninformative density as an appropriate prior for degree of freedom parameter and argued that the posterior density, given such ambiguous information, might be not integrable if Gibbs sampler is the target simulator. Here, it is necessary to note that his arguments do not contradict the previous results. In Kleibergen and Van Dijk (1993), integration of posterior is actually performed by using importance sampling technique. Although a truncated density shape is presented, it will not bias the integration results since high mass is still located in the untruncated space. To find a more proper substitute, Geweke (1993) himself assumed an exponential prior. That is $p(\nu) = \lambda \exp(-\lambda \nu)$. He found that, as long as the posterior draws for ν are not drawn from [0, 2], the variance of T-distributed innovations will not approach infinity and their empirical moments exists. Thus, parameter space for ν in his paper is set to be [2, +∞]. Besides, Mendoza-Blanco and Xin (1997) have also proved the appropriateness of using exponential priors. To see other ways of proposing priors for this parameter, Bauwens and Lubrano (1998) provided a detailed illustration.[77]

As for the purpose of this research, following Lin, Lee and Ni (2004), we now assume $p(\nu)$ to be uniformly distributed in a finite space from two to one hundred. That is $p(\nu) \sim U[2,100]$. Thus, given the joint posterior density (6.41), we can easily derive the sampling kernel for ν

$$\begin{aligned}\kappa(\nu \mid \mu,\pi,\theta,\psi,y,z) &= \prod_{t\in\{z_t=m\}}^{T}\left(1+\frac{(y_t-\mu_m)'\Sigma_{mt}^{-1}(y_t-\mu_m)}{\nu_m}\right)^{-\frac{\nu_m+d}{2}} \\ &= \prod_{t:z_t=1}^{T}\left(1+\frac{(y_t-\mu_1)'\Sigma_{1t}^{-1}(y_t-\mu_1)}{\nu_1}\right)^{-\frac{\nu_1+d}{2}} \times \prod_{t:z_t=2}^{T}\left(1+\frac{(y_t-\mu_2)'\Sigma_{2t}^{-1}(y_t-\mu_2)}{\nu_2}\right)^{-\frac{\nu_2+d}{2}}\end{aligned} \tag{6.47}$$

after all elements not related to this parameter are eliminated.

Here, it is important to note two things before proceeding. First, compared to the ADCC-MGM model, ADCC-MTM now contains a larger parameter set with more elements needing to be simulated but less analytical solutions available. Thus, the computational burden for calculating its inference using Bayesian method is much heavier. Second, apart from the sampling procedures we have just described, there are also other ways available for sampling parameters in ADCC-MTM. For example, by utilizing the hierarchical form of multivariate T, one can devise a hybrid method to estimate this mixture model. Since

[77] In Bauwens and Lubrano (1998), they argued that, in order to make posterior of ν integrable, the prior distribution assumed should force the posterior tending to zero quickly enough at both tails. Thus, a proper prior should be at least $o(\nu^{1+d})$ where d is the dimensionality of the training data.

degree of freedom parameter can be absorbed if such a hierarchical structure is assumed, a new parameter denoting the missing weight vector of the training data then needs to be introduced to the likelihood function (or joint posterior density). McLachlan and Peel (2000) described a so-called ECM estimation procedure for T mixture models when they are specified in a hierarchical way; a similar investigation through Bayesian inference is illustrated in Lin, Lee and Ni (2004). (See Appendix III for a more detailed description of hierarchical form of student T distribution)

6.5 In-Sample and Out-of-Sample analysis

Once the inferences have been calculated, assessing ADCC mixture models' performance in approximating in-sample correlation and forecasting out-of-sample correlation are also two topics of interest in this research. To perform these analyses, we calculate four quantities as follows. First, given the training data y_t and simulated N-state Markov chains $\varphi^{\{n\}}$, we generate the in-sample correlation at each time point as the posterior mean of conditional correlation calculated by inputting n^{th} simulated parameter values to the target models. Then, in a similar way, out-of-sample correlation forecasts, return forecast and next day's VaR of two mixture models are also generated.

6.5.1 In-Sample correlation estimation

First, for the in-sample analysis, since the true parameter value $\hat{\varphi}$ is now approximated by the empirical summary of a series of random draws, we can easily obtain a sample from the posterior distribution of conditional correlation by calculating $R_t^{(n)}$ for each $\varphi^{(n)}$ simulated. For example, provided that we have run posterior simulation of mixture models for N times and obtained a total of N draws for each parameter, then, if we assume all Markov chains have converged after S iterations, the average values of $R_t^{(n)}$ for $n \in [S+1, N]$ can be used to compute the posterior mean of in-sample correlation.[78] That

[78] Note that the calculation of posterior mean of in-sample correlation above is not based on the simulation from a sampling kernel, but by putting all updated parameter values of the same state to a

is,

$$E\left[R_t \mid y, \varphi\right] \sim \frac{1}{N-S} \sum_{n=S+1}^{N} E\left[R_t^{(n)} \mid y, \varphi^{(n)}\right] \tag{6.48}$$

6.5.2 Out-of-Sample correlation forecasting

Besides, to calculate the future correlation, we can adopt a similar approach. As Engle (2001) puts it, "volatility models are created to forecast volatility"; accordingly, correlation models are also invented for the same purpose. Since the intent of proposing ADCC-MGM and ADCC-MTM is actually to increase the model flexibility of capturing the stylized characteristics exhibited in financial data, it is then important to see whether this increased sophistication can improve the accuracy of dynamic correlation forecasts.

Here, since the conditional correlation is now modelled by ADCC of Hafner and Franses (2003), only one-step-ahead correlation forecast can be generated. This is because, if we want to obtain multi-step-ahead forecast, say R_{T+2} , based on the information currently available at *T*, the exact value of y_{T+1} then needs to be known so as to determine the value of ϑ_{T+1} to be input to the forecasting function. For example, if the current task is to forecast R_{T+2}, an essential variable that needs to be calculated is Q_{T+2} which is a function of $\varepsilon_{T+1}\varepsilon_{T+1}^{'}$, Q_{T+1} and $\vartheta_{T+1}\vartheta_{T+1}^{'}$. Here, although we can make the approximations through either $E(\varepsilon_{T+1}\varepsilon_{T+1}^{'}) \approx Q_{T+1}$ or $E[Q_{T+1}] \approx E[R_{T+1}]$ in equation (6.1) to obtain a recursive function for forecasting (see Sheppard and Engle, 2001, for details), the expected value of $E(\vartheta_{T+1}\vartheta_{T+1}^{'})$, which depends on y_{T+1}, is not obtainable at time *T*. Thus, we can only generate one-step-ahead correlation forecast (R_{T+1}) here.

Now, consider a series of converged Markov chain $\varphi^{\{n\}}$; after replacing $\bar{Q}$ in (6.1) with $\bar{R}$ and $\varepsilon_{t-1}\varepsilon_{t-1}^{'}$, Q_{t-1} with R_T, we can easily generate the predictive distribution of one-step-

covariance model. Although this result is called 'posterior', one needs to make clear its difference from the simulated parameter values that are generated by applying a simulation technique.

ahead correlation forecasts by calculating $R_{T+1}^{(n)}$ for all *n* in $[S+1, N]$. Its expected value is just equal to the conditional mean of these samples. That is,

$$\begin{aligned}
&E\left[R_{T+1} \mid y, \varphi\right] \sim \frac{1}{N-S} \sum_{n=S+1}^{N} E\left[R_{T+1}^{(n)} \mid y, \varphi^{(n)}\right] \qquad where \\
&R_{T+1}^{(n)} = [1-(\eta^{(n)})^2 - (\varsigma^{(n)})^2]\bar{R} - (\iota^{(n)})^2 \bar{N} + [(\eta^{(n)})^2 + (\varsigma^{(n)})^2]R_T + (\iota^{(n)})^2 \vartheta_T \vartheta_T^{'} \qquad (6.49) \\
&\bar{R} = E[\varepsilon_t \varepsilon_t^{'}]; \qquad \bar{N} = E[\vartheta_t \vartheta_t^{'}]
\end{aligned}$$

6.5.3 Next day's return forecast and VaR estimation

Besides, in a similar way we can also estimate the predictive density of next day's return *y_{T+1}*. According to equation (6.1) and (6.2), since distributional assumptions for training data have already been given (That is, the density of next day's return y_{T+1} would be a bivariate two-component Gaussian mixture with mean $E[\mu]$ covariance $E[\Sigma_{T+1} \mid y, \varphi]$ if y_t is now modelled by ADCC-MGM and a bivariate two-component T mixture if ADCC-MTM model is fitted),

$$p(y_{T+1} \mid y) = \int_{\varphi} p(y_{T+1} \mid \varphi, y) p(\varphi \mid y) d\varphi \qquad (6.50)$$

to obtain the predictive density, we then only needs to calculate the one-step-ahead covariance forecast Σ_{T+1}. Now, given that $E[\Sigma_{T+1} \mid \varphi, y] = E[D_{T+1} R_{T+1} D_{T+1} \mid \varphi, y]$ and R_{T+1} can be generated from (6.49), the remaining task is just to forecast D_{T+1}.

Here, if the posterior sampling sequence of a ADCC-mixture model has been iterated *N* times and associated burn-in period is set to be *S*, the posterior mean of one-step-ahead volatility forecast $D_{T+1}^{(n)}$ can be calculated by

$$\begin{aligned}
&E\left[D_{T+1} \mid y, \varphi\right] = \frac{1}{N-S} \sum_{n=S+1}^{N} E\left[D_{T+1}^{(n)} \mid y, \varphi^{(n)}\right] \qquad where \\
&D_{T+1}^{(n)} = w^{(n)} + \alpha^{(n)}(y_T - \mu)(y_T - \mu)^{'} + \beta^{(n)} D_T^{(n)}
\end{aligned} \qquad (6.51)$$

Its corresponding predictive return density y_{T+1} can be derived as the mean of a series of either MGM or MTM distributed random variables whose mean and covariance are evaluated by inputting all equilibrium draws to corresponding correlation mixture models.

$$p(y_{T+1} \mid y) \sim \frac{1}{N-S} \sum_{n=S+1}^{N} p(y_{T+1} \mid y, \varphi^{(n)}) \qquad (6.52)$$

In Ausin and Galeano (2005), the authors presented a way to compute the predictive mean and variance for (6.50). However, to generate its *VaR* estimates, quantile information of this density is also required.[79] Over the last decade, *VaR* has become the major risk management tool in financial industry. As proposed in 1995 by the Basle Committee, banks are now required to calculate the capital requirements for their trading books based on this measure and a large amount of literature are then devoted to producing better point estimates for this quantity to cover the potential maximum loss of next day, next month or an even longer period. For example, one can use either a parametric method (quasi-maximum likelihood and bootstrap resampling) or a non-parametric historical simulation approach to generate a *VaR* estimate. However, it is certain that it would be better if the distribution of this quantity is also known. Thus, its variability can be quantified and its precision is obtainable. Here, since, in Bayesian inference, parameters are now characterized by Markov chains, uncertainty in *VaR* (uncertainty of future returns) then can be described in a distributional form.

Applied into our cases, since it is known that the probability of a future volatility larger than a given threshold can be estimated by the proportion of observations in the sample larger than this threshold, given a specific $\varphi^{(n)}$ (or say $\Sigma_{T+1}^{(n)}$), if we now replicate *P* times the simulation of a sample from a mixture distribution with mean $E[\mu]$ and covariance $E[\Sigma_{T+1}^{(n)}]$, then we can obtain samples $y_{T+1}^{(n,p)}$ for *p=1,…,P*, which allow us to construct a predictive interval for y_{T+1} and finally generate next day's *VaR*. To illustrate this simulation procedure more clearly, consider now a task of calculating the next day's κ% *VaR* for an initial outlay *A* given that all MCMC posterior draws have been generated. First, for each replication *p=1,…P,* we obtain an estimate for *VaR*$^{(p)}$ using

$$VaR^{(p)} = A \times I_{\kappa}^{(p)} \qquad (6.53)$$

[79] VaR is the maximum potential loss associated with an unfavourable movement in market prices during a given time period with certain probability.

where $I_{\kappa}^{(p)}$ is the empirical κ -quantile of the samples, $y_{T+1}^{(1,p)}, y_{T+1}^{(2,p)} \cdots y_{T+1}^{(N,p)}$. Then, by iterating this process P times, we construct a predictive distribution for *VaR* and use conditional mean, median or mode of generated samples $VaR^{(p)}$ to approximate the true *VaR*.

6.5.4 Comparison with other correlation models

Apart from the in-sample and out-of-sample analysis, another important aspect of this chapter is model comparison and our aim is to see whether the ADCC mixture models proposed in this thesis can provide a better fit to observed data than other correlation models. Here, we consider four competitors for ADCC-MGM and ADCC-MTM. Specifically, they are CCC of Bollerslev (1990), DCC of Engle (1999), scalar ADCC of Hafner and Franses (2003) and diagonal AGDCC of Capiello *et al.*, (2004). For these models' specifications and characteristics, in Appendix II a detailed illustration has already been given.

As for their comparison criterion, here we consider using an economic loss function. Following Hafner and Franses (2003), we employ a so-called minimum variance criterion as a specification test. The main purpose is to construct an arbitrary portfolio using each time series (each asset) included in the bivariate training data $y_{t.}$ and then compare the variance of this portfolio after each asset is proportionally weighted. Here, we purposely constrain the average return of this portfolio close to a pre-specified value so that outperformance of a correlation model can be confirmed if it can generate the lowest variance among all alternatives. [80]

[80] In asset allocation problems, to discriminate the performance of different correlation (or volatility) models, usually there are three approaches. One is to constrain the portfolio return to a target level so that the outperformance of a model over its competitors can be confirmed if it can generate the lowest variance. Besides, we can through constraining the portfolio variance and locating the one which can generate the highest return to find the best. Meanwhile, if sufficient flexibility is allowed, in an unconstrained environment we can also freely compute the portfolio variance and return as models suggest and locate the optimal choice after comparing their Sharpe ratios. Note that the first two methods are suggesting a constrained optimal whilst a more balanced view is provided in the third approach. However, in terms of the stability of comparison results, it is then usually considered that the first two methods, especially the one constraining the portfolio return, can perform better than the third.

Concretely, if we now use $\Sigma_{t(i)}$ to represent the time-varying covariance matrix generated by i^{th} correlation model at time *t*. The weight vector for each constituent included in this portfolio is then calculated by

$$w_{t(i)} = \frac{\Sigma_{t(i)}^{-1} l}{l' \Sigma_{t(i)}^{-1} l} \tag{6.54}$$

where *l* is a (2×1) vector of ones. And the variance of target portfolio is computed by $V_{t(i)} = w_{t(i)}^{'} \Sigma_{t(i)} w_{t(i)}$. Here, one thing needs to be noted concerning the above portfolio is short selling is actually allowed. Therefore the weight vector is not constrained to be strictly positive definitive. Given these settings, consider a model, say *i*, if its portfolio variance $V_{t(i)}$ is now the smallest among all V_t obtainable from competing models, then we can say this model is the best-specified.

6.6 Summary

In this chapter, we introduce two new dynamic correlation models based on the mixture modelling techniques. By incorporating a variety of statistical characteristics, these two models are capable to account for multiple features frequently presented in financial time series such as fat tails, leptokurtosis, leverage effect and correlation targeting. And we describe how to calculate their inferences through a Bayesian approach. Specifically, for the inferential procedure, we use Griddy Gibbs sampler as the target simulator to sample draws for each parameter and use empirical summary of simulated Markov chains to approximate the exact inference. Besides, we also illustrate several ways of evaluating our models. For example, for the in-sample analysis, we calculate the dynamic correlation at each time point as the posterior mean of conditional correlation generated by inputting simulated parameter values to target models. For out-of-sample analysis, we derive next-day's correlation forecast, return forecast and VaR forecast.

Chapter 7

Simulation results and Empirical results

Introduction

In this chapter, we present the posterior results of ADCC-MGM and ADCC-MTM fitted to two simulated data and three empirical data. The whole chapter comprises four sections. In the first section, we describe the inferential results of two simulation studies. That is, we respectively simulate a series of multivariate Gaussian mixture distributed samples and a series of multivariate T mixture distributed samples both with ADCC-covariance evolving process incorporated, and then estimate them using posterior sampling procedure illustrated in the last chapter. After simulation, not only unconditional moments of posterior draws for each parameter are calculated, their kernel densities and convergence are also plotted and assessed. In the second section, to monitor the consistency and flexibility of our models, we consider three empirical applications where assets of different classes and assets in different markets are utilized. Their posterior results are illustrated in a similar way as previously. In the third section, we estimate in-sample correlation and forecast future correlation using two mixture models and apply the results into asset allocation and VaR calculation. Besides this, their performance is also compared to a variety of alternative conditional correlation models including ADCC, AGDCC and their variants. Finally, in the last section we summarize all major findings documented in this chapter.

7.1 Simulation studies

7.1.1 Simulated data

First, we describe two data-generating processes DGP1 and DGP2. The first corresponds to bivariate two-component ADCC-MGM model; the second corresponds to bivariate two-component ADCC-MTM model. For each process, we simulate 2000 observations and let the unconditional correlation of simulated data (bivariate) equal 0.8.

According to equation (6.1), since massive parameters are now incorporated to φ, a proper method for indexing them needs to be illustrated before we proceed. For example, for certain parameters like $\mu, \varpi, \alpha, \beta$, it is preferred a (2×2) matrix now can be used to express their values since these parameters contain elements corresponding to different components in the mixture and different series in the sample data simultaneously. Using μ for instance, this parameter is defined by

$$\mu = \begin{bmatrix} \mu_{a1} & \mu_{a2} \\ \mu_{b1} & \mu_{b2} \end{bmatrix} \tag{7.1}$$

where *1*, *2* denotes the first and second component distribution included in the mixture and *a*, *b* respectively correspond to first and second series of bivariate data. Here, μ_{b1} represents the mean parameter used to model the second series of sample data which is generated by the first component distribution in the mixture. However, for others such as $\pi, \eta, \varsigma, \iota$ and *v*, configuration of $\varphi = [\varphi_1, \varphi_2]$ is used because these parameters no longer contain elements corresponding to each series in the resultant data.

Now, in order to simulate a series of random sample from DGP1, we start by sampling two series of bivariate MGM-distributed random variables with ADCC-generated covariance separately incorporated into them using parameter values given below,

Mean parameters

$$\mu = \begin{bmatrix} \mu_{a1} & \mu_{a2} \\ \mu_{b1} & \mu_{b2} \end{bmatrix} = \begin{bmatrix} 0.001 & -0.002 \\ 0.010 & -0.023 \end{bmatrix}$$

Volatility Parameters θ

$$\varpi=\begin{bmatrix}\varpi_{a1} & \varpi_{a2}\\ \varpi_{b1} & \varpi_{b2}\end{bmatrix}=\begin{bmatrix}0.005 & 0.005\\ 0.050 & 0.004\end{bmatrix} \quad \alpha=\begin{bmatrix}\alpha_{a1} & \alpha_{a2}\\ \alpha_{b1} & \alpha_{b2}\end{bmatrix}=\begin{bmatrix}0.03 & 0.09\\ 0.04 & 0.05\end{bmatrix} \quad \beta=\begin{bmatrix}\beta_{a1} & \beta_{a2}\\ \beta_{b1} & \beta_{b2}\end{bmatrix}=\begin{bmatrix}0.90 & 0.60\\ 0.95 & 0.75\end{bmatrix}$$

Correlation Parameters ψ

$$\eta=[\eta_1,\eta_2]=[0.10 \quad 0.15]; \qquad \varsigma=[\varsigma_1,\varsigma_2]=[0.80 \quad 0.60]; \qquad \iota=[\iota_1,\iota_2]=[0.30 \quad 0.25]$$

and then proportionally mixing them using

Weight parameters

$$\pi=[\pi_1,\pi_2]=[0.7 \quad 0.3]$$

so that a single series of bivariate two-component mixture-distributed innovations with ADCC covariance can be obtained.

Here, it should be noted that we have purposely let the first mixture component follow a stable process with high probability and let the second be weak-stationary and low probability. This is because the financial market is often characterized by tranquil periods, suggesting a strong covariance stationary process. Since this feature is very common, it is then reasonable to associate it with a high probability. As a comparison, any structural changes that could lead to a substantial increase of the volatility are then much less frequently observed in markets and covariance in such periods is usually modelled by a weakly stationary process.

For simulation from DGP2, apart from retaining all settings just illustrated for DGP1, we also introduce the initial values for

Degree of freedom parameters

$$\nu=[\nu_1 \ \ \nu_2]=[10 \ \ 7]$$

and let $\varphi=(\pi,\mu,\varpi,\alpha,\beta,\eta,\varsigma,\iota,\nu)$, since multivariate T mixture distribution is to be assumed for modelling conditional returns.

7.1.2 Summary Statistics

Once all simulated data have been obtained, we illustrate their statistical characteristics below.

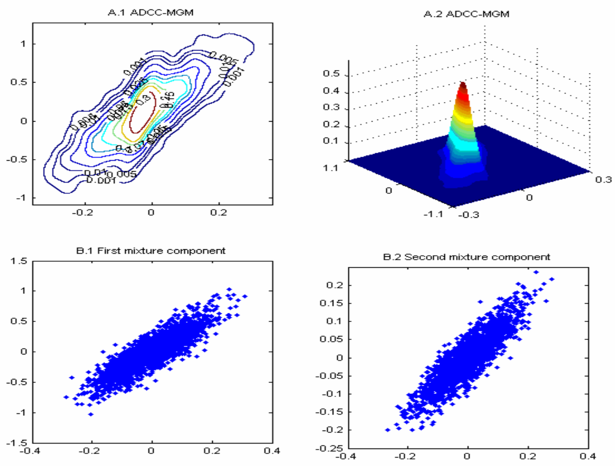
DGP1

First, for data simulated using DGP1, we report their first four unconditional moments and perform two hypothesis tests to examine their normality in both univariate and bivariate context.

<Insert Table 7.1 Panel A >

From Table 7.1, it can be easily seen that the first time series, individually, is a less volatile process than the second, with thinner tails observed on both sides. Both series now present a symmetric density shape (skewness≈0) while their normality results are different. After performing the Shapiro-Francia test, we found evidence of univariate normality for the first time series, with a *p-value* 0.4121. However, concerning the second, the null hypothesis is then rejected with calculated kurtosis generating a value exceeding 3.92. To evaluate multivariate normality, in this research we exploit a result from Doornik and Hensen (1994). Since the reported *p*-value of their test for DGP1 is 0.1143, statistically speaking, the null of multivariate normality cannot be rejected for any significance level stricter than 90%. Overall, since unconditional correlation calculated for this simulated bivariate data is also around 0.8, it is then fair to say that the first simulation provides a good refection of DGP1 in the sense that designated distributional characteristics are mostly captured. Graphical evidences for this argument are provided in Figure 7.1 where kernel density, contour plot and scatter plot of resultant data is presented.

Figure 7.1 Kernel density estimate plots, contour plots and scatter plots
Panel A. Bivariate data simulated using DGP1



Concerning the bivariate kernel density plot (right on the top row), we now use initial values given in DGP1 to calculate the unconditional variance (a fixed value) of two mixture components and then combine them with means to calculate the density estimates of two Gaussians.[81] It is known that, if the modes of two components were sufficiently far apart, one would expect their resulting mixture to resemble two Gaussian densities side by side, that is, a bimodal density. However, in Figure 7.1 only one mode can be observed. It then reveals the fact that means of two components (Gaussian distribution) are now set to be very close to each other and both near zero. Thus, the overlap between these components would tend to obscure the distinction between them. However, clear evidence for mixture can still be observed if we look at their contour plot. From the left graph of the top row, it can be easily seen that two symmetric densities are now combined on different domains so that density shape of their mixture does not appear to be round. Besides, from the same graph it is implied that the proportion of this mixture is unequal. As for this case, the one having flat tails is now given more weights than the one showing heavy tails.

DGP2

As for the training data obtained from DGP2, we perform the same standard analysis on its distributional characteristics as previously and, in Table 7.1, Panel B reports its results. As expected, means of both series are still central around zero and their unconditional correlation is close to the corresponding theoretical value. However, standard deviation is now found to be slightly larger than those reported in the previous cases and values of simulated data become more dispersed. For example, in the bivariate data generated from DGP1, the most volatile series is the second, whose values range from -0.87 to 1.02. However, when data corresponding to DGP2 is analyzed, minimum and maximum values of the same series are respectively -1.56 and 1.20. Given this feature, it is then implied that more tail behaviour is now incorporated into the current density and this finding can also be confirmed using increased values of kurtosis estimates for both series. Concerning their normality test, as can be seen, null hypotheses defined in the univariate context and multivariate context are now both firmly rejected with a close-to-zero probability value. This result is as expected because the components of the mixture are now both assumed to

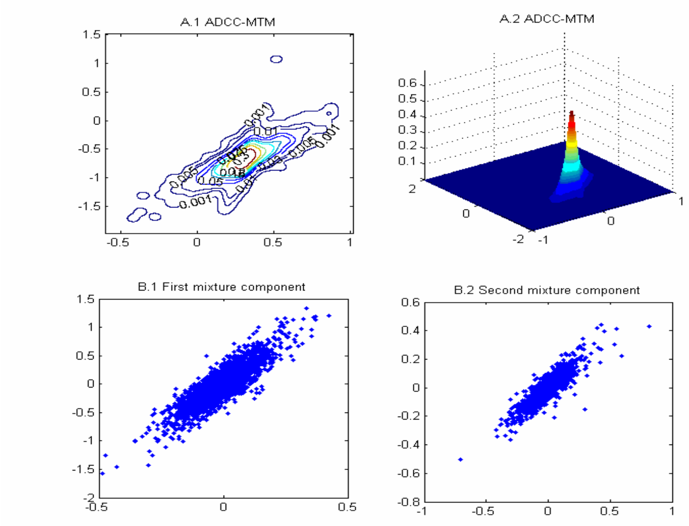
[81] We use Gaussian kernel here with bandwidth computed by rule of thumb.

be multivariate *t*-distributed which, only when its degree of freedom parameter is set to be infinity, will tend to show Gaussian features.

<Insert Table 7.1 Panel B>

To obtain a more comprehensive view of the aforementioned characteristics, we present, in Figure 7.1 Panel B, the bivariate kernel density plot and contour plot of data simulated using DGP2. As can be seen, the resulting mixture still appears to be unimodal with a seemingly symmetric shape in the centre but a far-from-symmetric shape around the edges. Features of high peakedness and fat tails are evident due to the inclusion of multivariate *t*. Tails in the positive domain are slightly heavier than those in the negative domain. And, occasionally, some extreme events can be observed.

Figure 7.1 Kernel density estimate plots, contour plots and scatter plots
Panel B. Bivariate data simulated using DGP2



7.1.3 Estimation results

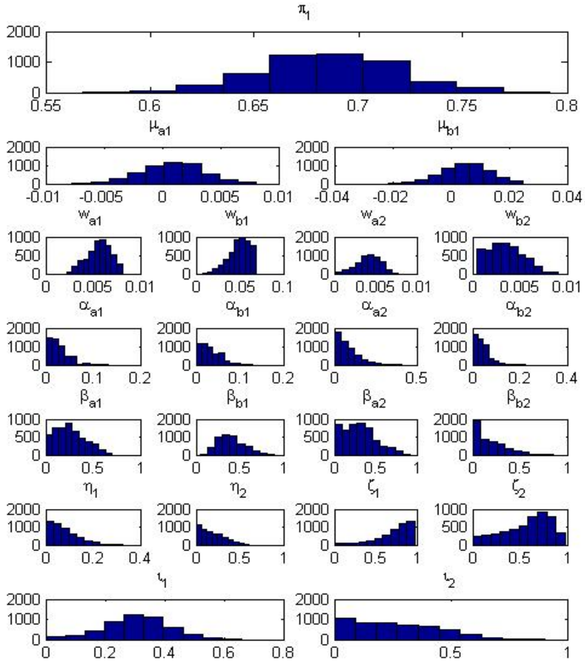
Next, we present the posterior estimation results for the two simulated data just obtained. First, for those generated using DGP1, we apply the sampling sequence illustrated in Section 6.3 to calculate its inference since the target model is assumed to be ADCC-MGM. However, while those obtained from DGP2 are estimated, procedure illustrated in Section 6.4 for fitting ADCC-MTM is then utilized. Here, for both cases, we use Griddy Gibbs sampler to run a total of 15000 simulations and discard the first 10000 for warming-up. Thus, posterior parameter value is approximated using the remaining 5000 draws in the simulated chains and, for these draws, we calculate their location measures (mean, median and mode) as well as dispersion measures (s.t.d, max and min) to obtain an idea of the central value and variability of each parameter of interest.

DGP1

In Table 7.2 Panel A, we first present the summary statistics of chains simulated for ADCC-MGM parameters and, in Figure 7.2 Panel A, give their histogram plots. In most cases, we find posterior location estimators now approximate their corresponding true parameter values reasonably well and, statistically, convergence of most parameters can be confirmed (See Table 7.3 for convergence result). For example, posterior mean, median and mode of weight parameter π_1 (0.6864, 0.6855 and 0.6910) are all very close to their corresponding theoretical values set in DGP1 (0.7) in the sense that theoretical value is located in a reasonable confidence interval.[82] *Z*-test results (0.9231) confirm that the first 1500 and the last 1500 samples included in the equilibrium state of π_1 have equal medians, suggesting that posterior draws have all statistically converged. Besides, since location measures (mean, median and mode) themselves are now very close to one another, it is then reasonable to expect a symmetric density shape for π_1 . Note that in Bayesian inference this feature is especially desirable because it proves the high mass of target parameter has been sufficiently explored around the most likely values.

Figure 7.2 Histogram plots of posterior draws of mixture models' parameters
Panel A. ADCC-MGM estimated on simulated data obtained using DGP1

[82] Here, we use confidence interval to depict the closeness between true parameter value and simulated values. For example, in a normal distribution, such confidence intervals, say, $\mu \pm 3\sigma$ are often used to depict the 99% of probability events around the mean. Applied in this case, if the theoretical value of a parameter is located within $\mu \pm 3\sigma$, then we say the posterior moments approximate this value reasonably well.



<Insert Table 7.2 Panel A and Table 7.3>

However, as can be seen, in all except some special cases (such as $\mu_{a1}, \mu_{a2}, \iota_1$), this bell-shape is seldom observed in Figure 7.2. Often, posterior density is characterized by either significant evidence of excess skewness or sometimes multi-modality. In such cases, to identify a good estimator, great care then needs to be exercised. Here, if asymmetry is the only factor to be accounted, we can use either posterior median or mode as a better estimator than mean for approximating true parameter value. This is because these empirical moments in the case of asymmetry can provide a more reliable representative of the whole density than the mean as a single statistic. For instance, in the posterior density of ς_1 , since negative skewness is observed, it is then reasonable to expect posterior mode (0.9310) to be a more proper estimator than mean (0.7299) for approximating theoretical value (0.9). However, when multi-modality is present, the necessity of rechecking the convergence result is then highlighted. For example, in the case of β_{a2} , posterior density presents a seemingly flat shape in the range of [0, 0.5] and then gradually decreases its density values from 0.5 to 1. If we now look at its results from *Z*-test, *PSRF* test and *IPSRF* test, the null hypothesis of statistical convergence cannot be rejected on any significance level, suggesting that its posterior moments are already sound enough to be used for approximation (See Table 7.3). However, the problem is that, given a numerically-confirmed converged Markov chain, its posterior moments (mean: 0.29; median: 0.28; mode: 0.32) are still far from the theoretical value (0.6) they are supposed to be close to. Since, from the reported maximum and minimum values, one can know that the search for new updates for β_{a2} has already been directed to a relevant area [0.0001 0.9230], this problem then cannot be simply explained by the improper choice of either grid points or integration (interpolation) techniques because, given a correctly specified sampling kernel, only mild difference would be generated if these points and techniques are chosen differently.[83] Thus, doubts are spontaneously cast onto the effectiveness of our simulated data and simulation procedure. For example, one may want to question 'whether the coding of our posterior sampling sequence is problematic or sufficient enough' or 'whether the simulated data used in our sample is a good reflection of DGP1.' Concerning the first hypothesis, it is firmly rejected if we look further into the estimation result of empirical investigation where sound posterior results are found throughout the cases. However, as for the second, its possibility cannot be simply ruled out, although desirable distributional characteristics have been documented in the last subsection. This poor posterior performance is especially the case for β in volatility parameters. As can be seen from Table 7.2, posterior moments of these parameters are all around 0.2-0.4, whilst their corresponding true values actually range from 0.60 to 0.95. Moreover, in one case (β_{a1}), convergence of the resultant chain is rejected if we reset the significance level to 95%. Although a large gap is now observed, it is necessary to note that β is the only exception here; for all others parameters, such as $\psi = (\eta, \varsigma, \iota)$, μ and, needless-to-say, the weight parameter ϖ, their posterior moments are then good estimators for approximating true values using confidence intervals, and one can find a clear peak in their posterior distributions to represent the whole posterior density.

[83] For example, when we use Taylor method to expand a function *f(x)* around zero, the only difference between first order approximation $f(0)+C_1\ f'(0)x+o(x^n)$and second order approximation $f(0)+C_1\ f'(0)x+C_2 f''(0)x^2+ o(x^n)$ is only a function of x^2, say $g(x^2)$. Since integration and interpolation techniques applied in this paper are both based on the first order approximation, its difference from other alternatives is then a high-order function.

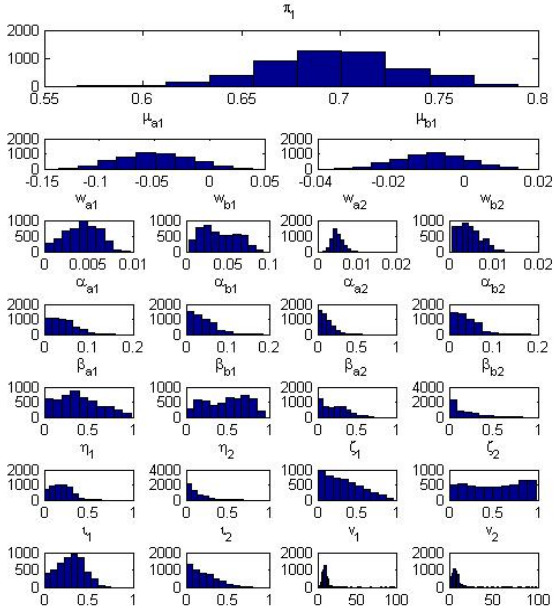
Besides, concerning DGP1, there is also another interesting finding worth mentioning here. That is, simulated draws of parameters belonging to the second component often present a more volatile process than those of the first. For example, if α and η are now of interest, *s.t.d* of posterior values simulated for α_{a2} and α_{b2} (0.063 and 0.031) are larger than those simulated for α_{a1} and α_{b1} (0.020 and 0.023), and the chain simulated for η_2 (*s.t.d*: 0.1366) is more volatile than that generated for η_1 (*s.t.d*: 0.059). This result is as expected because likelihood curvature is indeed much smaller for the second component. For example, in this simulation study only 2000×0.3=600 observations are expected to be generated from the second Gaussian-component. Compared to the first (with 1400 observations), since sample size is now much smaller, the appearance of the same probability event would then be less frequent and it is natural to expect a more dispersed posterior distribution for its associated parameters, suggesting a higher standard deviation. In this research, since the first mixture component is always assumed to have a higher proportion than the second so as to ease the label-switching problem, similar results are documented throughout this chapter.

DGP2

Now, we turn to analyze the MCMC outputs of the second simulation study where ADCC-MTM model is fitted. Respectively, in Table 7.2 Panel B and Table 7.3 we report summary statistics and convergence results for each parameter and in Figure 7.2 Panel B present their histogram plots. At first glance, it is easy to see that, for most parameters, their posterior values reported now are very close to those illustrated previously. This is because initial values in two simulations (DGP1 and DGP2) are mostly set as equal. However, if a detailed comparison is launched, several changes are still not difficult to be found.

<Insert Table 7.2 Panel B and Table 7.3>

Figure 7.2 Histogram plots of posterior draws of mixture models' parameters
Panel B. ADCC-MTM estimated on simulated data obtained using DGP2



For instance, a notable difference between ADCC-MGM and ADCC-MTM is the inclusion of a degree of freedom parameter. Concerning υ, posterior moments of its two elements in different component distributions now both approximate their corresponding theoretical values reasonably well. For υ_1, the best estimator for approximation is posterior mean (9.7546). However, for υ_2, posterior median (6.8628) then emerges as a better location measure. Indeed, in terms of the absolute value, *s.t.d* of this parameter is now the largest of all. However, their resultant Markov chains are not the most volatile ones. For example, if we compare the posterior samples of η_2 with υ_1, relative volatility of η_2 is 0.88 while that of υ_1 is only 0.67.[84]

As for volatility parameters, their posterior results now improve a lot. Using empirical summary of β_{a1} for example, mean, median and mode of this parameter are now respectively 0.38, 0.35 and 0.34, much larger than the same estimators calculated in the previous case. Besides, as a reflection of DPG2, larger parameter values are now generated for parameters of the first component (like β_{a1} and β_{b1}) than those of the second (like

[84] Relative volatility measure is the value of *s.t.d* estimate divided by mean, that is, $s.t.d / \mu$.

β_{a2} and β_{b2}), suggesting that our configuration designed for DGP2 has been realized. That is, the first component is now associated with a stronger covariance stationary process than the second to represent the 'common' market behaviour.

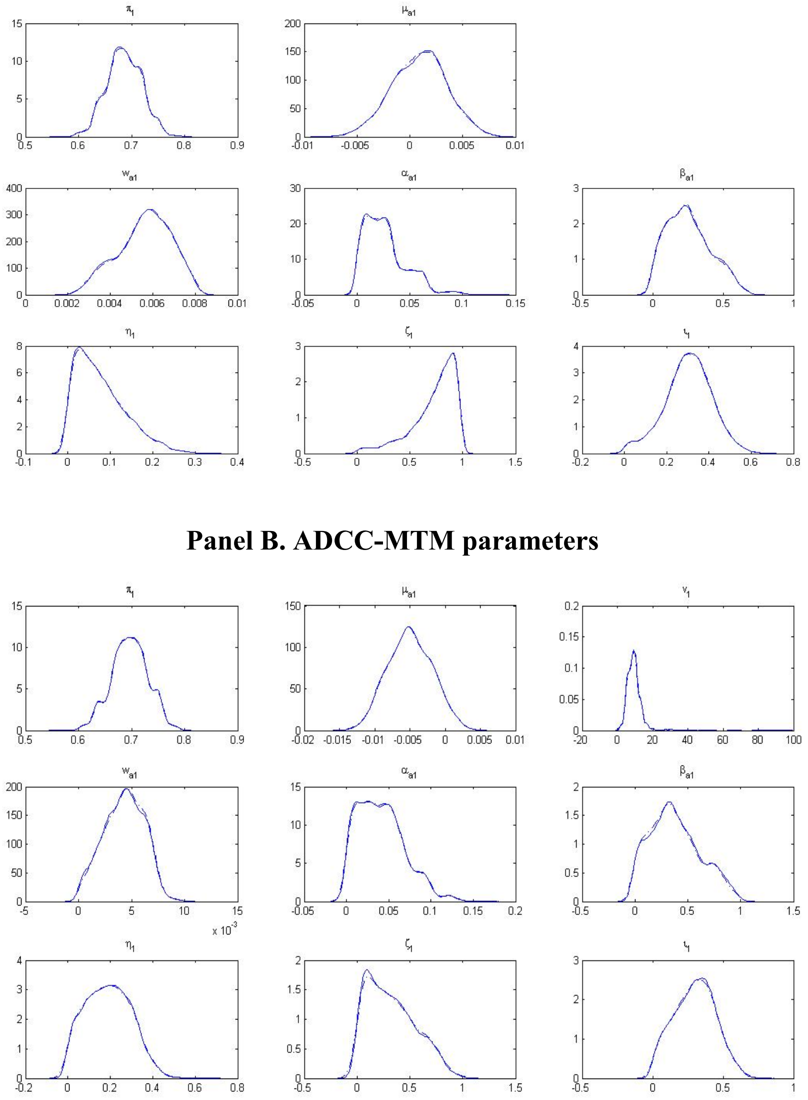
Besides, here it is also necessary to note something about the correlation parameters $\psi = (\eta, \varsigma, \iota)$. As can be seen from Table 7.2 Panel B, their posterior results now change a lot compared to those documented in the previous study. Although, for some elements, the distance between posterior mean and its corresponding true value is drastically shortened, i.e. ι_2, this improved performance cannot be applied to all. On the contrary, in some cases some evidence of deterioration is even found. For example, posterior mean of ς_1 obtained in ADCC-MGM (0.7299) can provide a close approximation to its theoretical counterpart (0.80) if one standard deviation (0.2088) confidence interval is imposed. However, the same estimator reported now (0.325), even if augmented with two standard deviations (0.227), still cannot approach the target value closely enough. Given such a large gap, it is then fair to say that the resultant Markov chain fails to provide sufficient information for posterior approximation and, unavoidably, doubts then need to be cast again on the validity of our simulated data. Concerning this issue, a more detailed explanation is to be provided in Section 7.2 and 7.4.

7.1.4 Individual Convergence speed

Above, convergence diagnostic results for two simulation studies have been shown through the implementation of three hypothesis tests. Although most of the chains are now confirmed as statistically converged, nothing has been said about their individual converging speed. Here, with respect to this issue, a brief illustration is provided below. First, for some particular parameters, we compare their posterior densities with different numbers of equilibrium draws to see how their convergence improves as simulation proceeds. Here, for the first simulation results, we plot in Figure 7.3 Panel A kernel density estimates of a variety of its chains with 2000 equilibrium draws (after 10000 burn-in draws are discarded) and compare them with those plotted with 5000 draws. Then, in the same figure, a similar graph for ADCC-MTM parameters is also drawn.

Figure 7.3 Posterior density comparisons of correlation mixture models' parameters after 2000 draws and after 5000 draws

Panel A. ADCC-MGM parameters



Panel B. ADCC-MTM parameters

Solid line denotes the posterior density after 2000 draws; dotted line denotes the density after 5000 draws. (Both chains are derived after initial 10000 warming-up draws are discarded)

As can be seen, in most cases, posterior densities plotted using dotted line and solid line are now overlapped, suggesting that the extra 3000 draws simulated do not contribute much information to purify the posterior distributions of interest, and an early distributional convergence has already been achieved. Given this feature, we then use

ADCC-MTM, for example, to calculate the correlation for all its resultant chains and present the result in Table 7.4.

<Insert Table 7.4 >

The purpose of calculating this statistic is to assess the individual converging speed based on the virtue that, the more one parameter appears to be correlated to the others, the lower the chance its realizations would be independent and the slower would its chain tend to converge. On the contrary, if a chain seems to be independent of all others, the probability of its obtaining a fast convergence will then be relatively high.[85]

Applied to our cases, from last chapter, since it is already known that sampling kernels of π, μ and υ are independent of one another, their resultant chains are then very unlikely to converge slowly. And this is especially the case for π because its simulating kernel has an analytical form. Since the simulated draws, for this particular parameter, are now to be *i.i.d*, its convergence is then expected to be very fast. However, concerning the others, such as volatility parameters and correlation parameters, the appearance of a much slower convergence speed, due to the strong intercorrelation found in their resultant chains, is then not surprising. For example, the correlation between ϖ_{a1} and β_{a1} is as high as -0.95 and that of ϖ_{a2} and β_{a2} also reaches -0.59. Given a large number of such highly-correlated chains, generating new updates for one will then inevitably be affected by the correlated simulated values of another.

7.2 Empirical investigation

Above, in simulation studies, we have documented an interesting finding, that is, for some parameters, their posterior performance of using empirical moments to approximate the target theoretical value is not uniform. To explain this phenomenon, we cast doubts on the validity of our simulated data while confirming the correctness of our estimation procedure.

[85] Here, as a natural result of Markovian property, autocorrelation within a chain is also a factor that could relate to the individual convergence rate. However, since dependence of adjacent points in equilibrium state gradually decreases after a sufficiently long run of burn-in iterations, we thus do not consider it here.

In the following, to provide the proof for this argument, three empirical studies were carried out.

First, in this research, since the motivation for proposing mixture models is due to the bi-modality observed in kernel density plot of realized correlation in the foreign exchange market, we model the bivariate daily return of USD/GBP (*US/UK*) and EUR/JPY (*EU/JP*) using samples from 01/01/1999 to 22/06/2005.[86] A total of 1689 daily observations are obtained. Second, following Engle and Colacito (2003), we investigate dynamic correlation between US bond index and US stock index. Daily prices of *S&P 500* index and *US 10-yr bond* are respectively collected from *DataStream* using code '*ISPCS00*' and '*CTYCS00*', and we select the sample from 03/01/1995 to 04/07/2006 with 3000 observations included. Finally, the third empirical analysis is performed to study the co-movement between UK stock index and US stock index. We choose 1000 daily observations for FTSE100 and S&P500 starting from 29/08/2003 to 30/06/2007.

For each sample, we now estimate it using both ADCC-MGM and ADCC-MTM and generate 15000 run Markov chains with initial 10000 draws deleted as burn-in points. Therefore, there are in total six posterior sampling procedures to be followed and model performance is now examined not only in a portfolio with assets of different classes (stock index, bond, currency) but also with assets in different markets (US, UK).

7.2.2 Summary statistics

As always, first we briefly summarize some statistical characteristics of these samples. Specifically, in Table 7.5 we present their descriptive statistics and the results of three normality tests (two univariate and one multivariate), and then, in Figure 7.4, we plot their sample paths and kernel densities.

<Insert Table 7.5 >

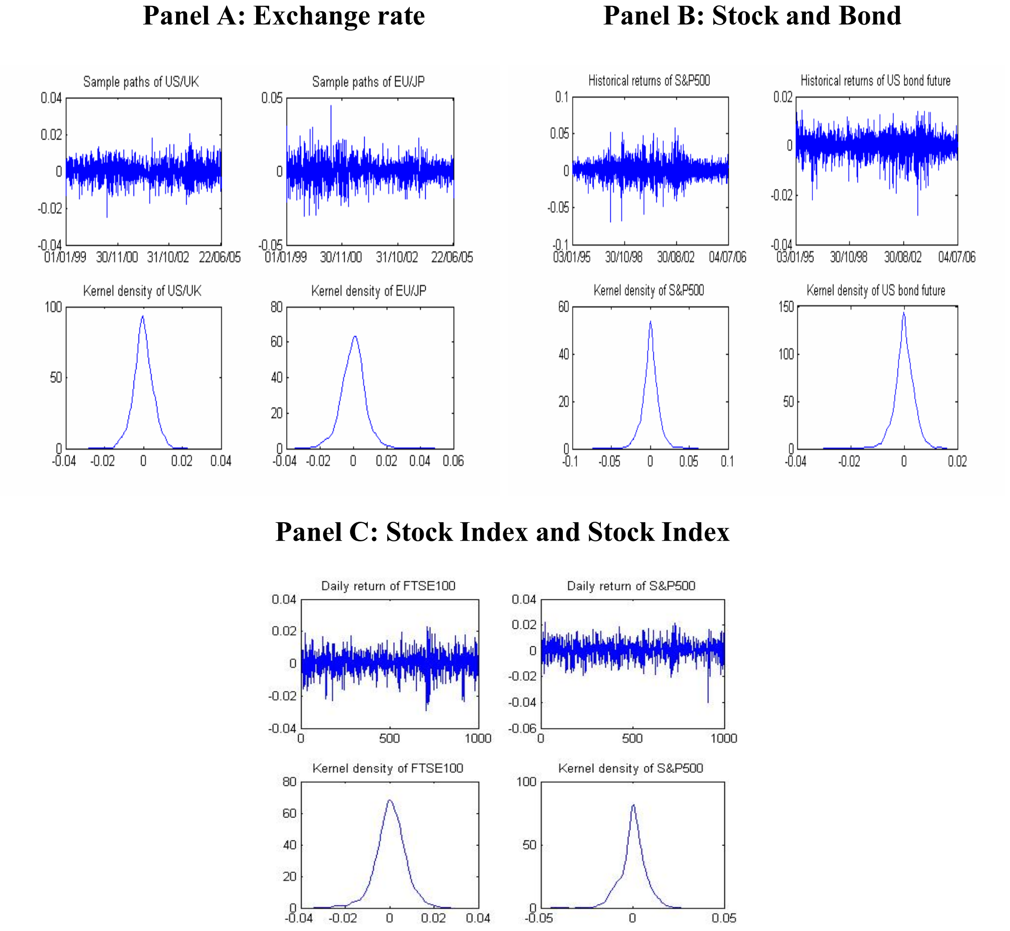
[86] Undoubtedly, using these pairs may be of little economic sense. But they are included mainly due to the significant evidences of bi-modality observed in unconditional distribution of their realized correlation. Besides, it is also because of their negative correlation. Since the other two samples used in this research now present either positive or near-zero correlation, these data are included mainly for the completeness concerns because our purpose is to examine proposed models in different correlation scenarios.

According to the results, it can be clearly seen that three different scenarios are now considered here, with US and UK stock indexes showing positive correlation, exchange rate data showing negative correlation and stock and bond data showing near-zero correlation. Concerning the first case, it is an expected result because equity markets in most developed countries are already known to be positively correlated to one another. Negative correlation between US/UK and EU/JP is also unsurprising because, during the sample period we choose, the spot rate of US/UK experienced a steady appreciation as EU/JP was heading for a sharp devaluation. Here, the only thing worth noting is the -0.083 correlation found between S&P500 and US bond future. Although, by their different market natures, it is reasonable to expect a low correlation, in interpretation great care needs to be exercised. Recall that, in Chapter two, it has already been stated that correlation is an association measure only defined on linear space. Therefore, even if given this near-zero correlation, it is still too early to say that there is no underlying dependence structure between these two financial assets.[87]

As for their statistical characteristics, historical returns of most samples now follow stable processes with standard deviations only in one case observed exceeding 0.01. Most of the kernel densities present a seemingly symmetric shape with negative skewness. The only exception here is for EU/JP return where a positive skewness of 0.0135 is reported. Concerning the normality test, overwhelming evidences have been found showing the rejection of Gaussian in either univariate context or multivariate context. Only in one case, (S&P500 used in the second sample data), does the *p*-value of Shapiro-Francia normality test generate a value exceeding 0.042, suggesting that, for this particular time series, univariate Gaussian cannot be rejected at 99% level but will still not be accepted once we relax the significance level to 95%. Finally, in all cases fat tails and high peakedness are confirmed with no sample showing calculated kurtosis being able to generate a value of less than three.

[87] On the contrary, in many different ways bond market could be said to be correlated to equity market. For example, when expectation of an interest rate rise goes up, bond price will dive immediately. At the same time, equity prices in this scenario are also expected to fall due to the concurrent fear of stricter monetary policy to be imposed.

Figure 7.4 Plot of historical returns and kernel densities of three empirical data


Panel A: Exchange rate **Panel B: Stock and Bond**

Panel C: Stock Index and Stock Index

7.2.3 Estimation results

a. starting value setting

Now, we start to illustrate the posterior results of ADCC-MGM and ADCC-MTM fitted to these empirical data. Before proceeding, some illustrations on the initial values to be set for each parameter need to be provided as MCMC outputs are often found sensitive to these prior information (either distribution or value) assumed. Specifically, in this research we use techniques such as mode-finding or fitting a relevant model to search for the proper starting values. As for weight parameter, since it is assumed that first component can always obtain a higher proportion, we retain the same setting for π in simulation study so that sampling of π_1 starts by using 0.7 as the origin for resultant chain. Concerning mean parameters, in order to minimize the potential bias, simulation of target chain is initiated by using unconditional mean of sample data as the first state. With respect to volatility and correlation parameters, a more numerically-efficient searching method is then adopted.

That is, first we fit a standard ADCC (1,1,1,1) model to each sample so that initial values of θ and ψ for the first mixture component can be obtained. Then, after a mild modification, those for the second are given. Here, note that this modification is now made by mildly decreasing the parameter values calculated for the first component. This is because we are now inclined to give the first component stronger volatility persistence and stronger correlation persistence so that, overall, its covariance process would appear to be more persistent and stationary than the second. In so doing, a tranquil period frequently observed in financial markets can then be modelled. Moreover, when ADCC-MTM is used, we also fit a bivariate T distribution to the sample data so that the initial value of degree of freedom parameter can also be obtained.

b. foreign exchange rate result

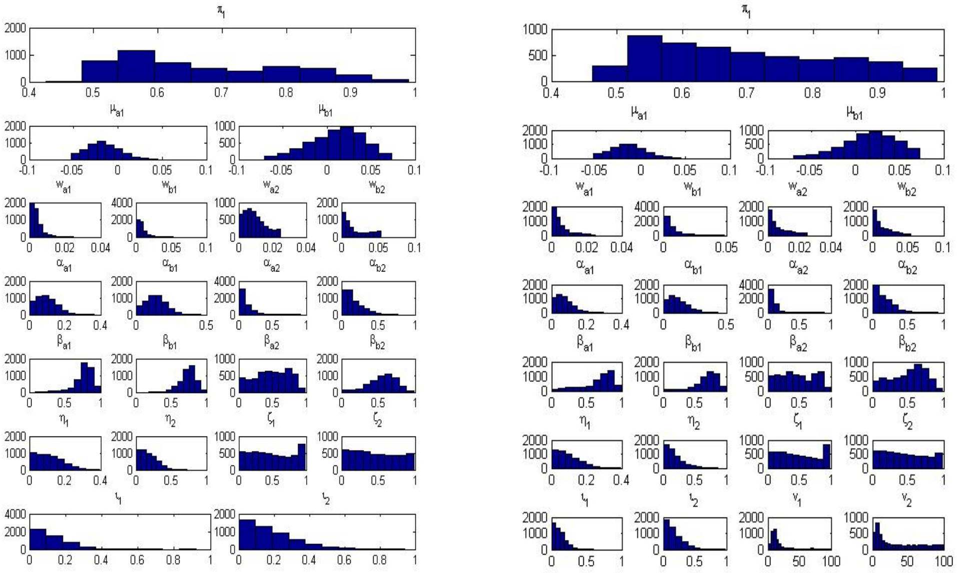
Next, we report posterior estimation results of fitting foreign exchange data. Posterior moments of their parameters are given in Table 7.6 Panel A and Panel B and their corresponding histograms plotted in Figure 7.5 Panel A and Panel B.

<Insert Table 7.6 >

After a brief comparison, we find that, for most parameters, their resultant posterior moments in two models are now very close to each other. This result is as expected because correlation mixture models proposed in this paper are already known to be closely related. Not only are their mechanisms of generating correlation dynamics based on the same specification, density assumption (multivariate Gaussian mixture) given in ADCC-MGM is also a limiting case of that (multivariate T mixture) assumed in ADCC-MTM.[88] Given that training data is also used in the same way, it is then reasonable to generate similar results for the same parameters. For example, the posterior mean, median and mode of weight parameter in estimated ADCC-MGM are respectively 0.667, 0.639 and 0.566. The same statistics in ADCC-MTM are reported to be 0.694, 0.671 and 0.542. Besides, if their kernel densities are analyzed, a similar degree of positive skewness can also be observed, suggesting that posterior mode is now a better location estimator than mean to approximate the true parameter value.

[88] MGM is a limiting case of MTM when the degree of freedom parameter v in MTM approaches infinity.

Figure 7.5 Histogram plots of parameters in empirical fitting of exchange rate data



Panel A. ADCC-MGM **Panel B. ADCC-MTM**

In addition, this asymmetry can also be found in a variety of other cases. Here, it is especially worth noting ς (the parameter governing the correlation persistence) and υ. For example, contrary to most findings that posterior draws smoothly disperse over a wide space, posterior density of ς_1 concentrates only on a tiny region with most probability events occurring around $\varsigma_1 \approx 1$. Thus, in approximation, posterior modes (0.9308 for ADCC-MGM and 0.9310 for ADCC-MTM) once again outperform means (0.4914 for ADCC-MGM and 0.4854 for ADCC-MTM). However, concerning ς_2, its density shapes is then much flatter. Since the convergence of all chains have be statistically confirmed, for this particular parameter, using posterior mean as a representative of the whole density is more proper. As for degree of freedom parameter, similar asymmetry is also found in its two elements. Skewness estimated for υ_1 is 2.1329, much larger than that calculated for υ_2 (0.5552). An interesting finding here is their posterior modes are now nearly the same. Given that weight parameters calculated in the above case are also roughly the same $(\tilde{\pi}_1 \approx 0.5)$, the appeal of using two-component mixture models to quantify the correlation between currency-pairs is then almost lost.[89] This is because, what is assumed for correlation models are now two equally-weighted Gaussian samples or two equally-weighted T samples with same tail behaviours. Provided that the means of two component distributions are also not far from each other, this is then equal to assuming just one-component ADCC-Gaussian or one component ADCC-T.

Besides this, another interesting finding here is the close-to-zero posterior values reported for μ and ϖ. As will be shown in later sections, since this is a common result (See Table 7.7 and 7.8 for proofs), illustrations of this particular issue are then necessary. First, given this feature, it is necessary to rule out the possibility of non-convergence of relevant chains and this can be easily proved by calculated *Z*-statistics. Then, concerning these small values, the question of 'whether we could eliminate them in the target correlation mixture models' is naturally raised. Definitively, it would be a potentially beneficial strategy if we could adopt it. Since Griddy Gibbs sampler is already known as a numerically demanding algorithm, massive computational work can be saved if six parameters (both ϖ and μ), which could previously only be updated using non-analytical kernels, are now available for elimination. More importantly, even without the inclusion of these elements, the mechanism assumed for covariance (volatility and correlation) process in target model will not be fundamentally altered. In this research, we have tried fitting ADCC-MGM and ADCC-MTM again for the same empirical data with these parameters omitted, and similar results are generated to those just reported while computational expense is substantially reduced. For example, for ADCC-MGM, now, 28.57% of the original computational work can be saved in each loop of posterior sampling. This quantity for ADCC-MTM is 26.09%. Although a much cheaper solution for calculating Bayesian inference is now available, to retain the origin virtue, here we still report the posterior result using the same style as previously, with ϖ and μ both included.

c. Stock and Bond data

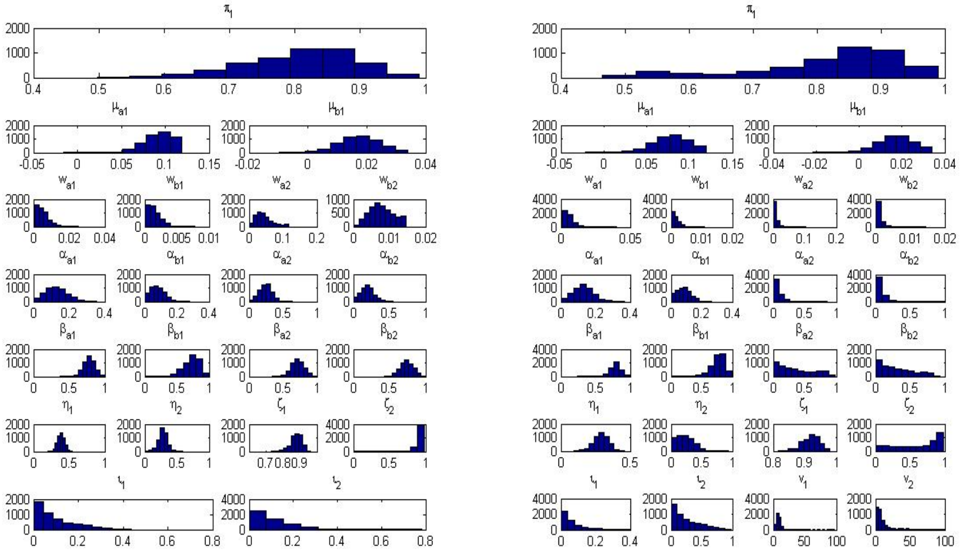
Now, we proceed to illustrate the posterior results of empirical fitting for stock and bond

[89] $\tilde{\pi}_1$ denotes the theoretical value for π_1

data. For two mixture models, their posterior parameter values are respectively documented in Table 7.7 Panel A and Panel B and corresponding histograms plotted in Figure 7.6 Panel A and Panel B.

<Insert Table 7.7>

Figure 7.6 Histogram plots of parameters in empirical fitting of stock and bond data



Panel A. ADCC-MGM **Panel B. ADCC-MTM**

Compared to the previous case, at first glance one can already obtain an idea that a better fitting result is now generated. Not only resultant parameter values become more sensible and posterior density presents a desirable shape, the inherent mechanisms we assume for mixture models are also realized. For example, in both ADCC-MGM and ADCC-MTM models, the first components now can easily obtain the dominating power in modelling entire covariance dynamics. Posterior means of weight parameter of this component are around 0.8. These components are given stronger volatility persistence and stronger correlation persistence than the second. Take the correlation process for instance. Coefficient determining the stationarity in the first component is now given by $\eta_1^2+\varsigma_1^2$ (0.917), which is a value larger than the one $\eta_2^2+\varsigma_2^2$ assumed in the second component (0.826). Thus, it is reasonable to expect that a large change in the target correlation dynamics is to be followed by another large change.[90] The majority of the observations

[90] Here, we use $\eta_1^2+\varsigma_1^2$ instead of $\eta_1+\varsigma_1$ to depict the correlation stationarity because of the model

(≈0.8*3000=2400) support this finding. This is then equal to saying that weak stationary process, governing the sudden changes in correlation dynamics, can only be observed using 600 samples. Concerning the degree of freedom parameter, again, asymmetric posterior density is documented for both its elements included in the mixture. However, slightly different modes are now observed with υ_1 equalling 5.79 while υ_2 equals to 5.94.

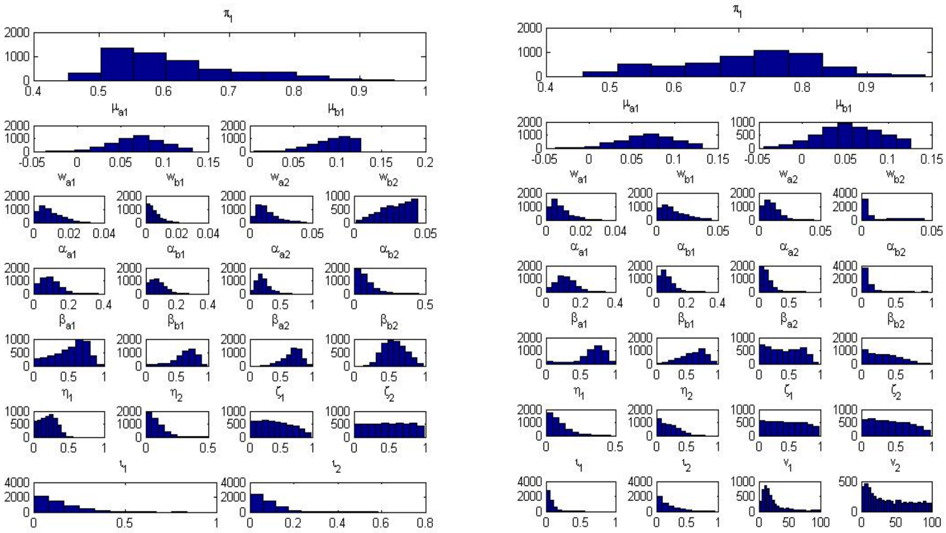
d. Stock index data

Finally, we report the posterior result of fitting ADCC-MGM and ADCC-MTM to stock index data. Their parameters are illustrated in Table 7.8 while densities are plotted in Figure 7.7.

<Insert Table 7.8>

Figure 7.7 Histogram plots of parameters in empirical fitting of stock index data

Panel A. ADCC-MGM **Panel B. ADCC-MTM**



Here, for most parameters, it can be easily seen that their posterior results are now found quite similar to those previously reported. However, three interesting findings are still documented. First, new concerns have been raised over weight parameter. As can be seen from its posterior moments, for ADCC-MGM the best location estimator to describe π_1 is posterior mode (0.528) whilst, for ADCC-MTM, the same statistic then increases the proportion of first component to 0.750. Second, as a response to the roughly equal-mixing

specification assumed in chapter 6.

assumed in ADCC-MGM, correlation process generated in two components now shows a similar degree of stationarity. For example, parameter value generated for η_1 is very close to η_2. Besides, for ADCC-MTM, different tail behaviours of multivariate T are finally observed. Posterior mode for υ_1 is 15.83, whilst that for υ_2 is only 5.95. Given these features, it is then natural to draw that conclusion that ADCC-MTM is actually a correlation model more capable than ADCC-MGM of distinguishing different components in the resultant mixture. This is because, unlike Gaussian mixture, where the task of identifying different components can only be performed by imposing a restriction on π or μ, in a multivariate T mixture model it can also be done by giving different values to the degree of freedom parameter of different components (See Chapter 4 for more details on label-switching).

7.2.3 Implementational issues

Above, we have illustrated the posterior results of two simulation studies and three empirical investigations. Now, before proceeding to examine the in-sample and out-of-sample performance for mixture model, we illustrate some implementational issues that have been raised in previous sections but not stated completely enough. These issues included approximation error in integration and computational cost of Griddy Gibbs sampler.

a. Approximation error

First, as have been mentioned several times in Chapter 6, approximation error is a critical issue when posterior result of a MCMC algorithm is analysed. Applied to this research, since sampling kernel of most parameters in correlation mixture models does not have an analytical form, simulation of new draws then has to rely on the principle of '*Inverse of C.D.F*'. According to Griddy Gibbs sampler, since the integration now needs to be calculated by evaluating a number of grid points on a relevant space, potential bias is then unavoidable and it is not surprising to see some difference between calculated posterior mean and true parameter value. Although we have chosen as many points as possible after balancing the efficiency of algorithm with accuracy of results, in some cases this error is still a major factor biasing the posterior result. For example, for some parameters, one has observed that posterior mean of a converged chain fails to approach its corresponding theoretical value closely enough. However, after being augmented with a reasonable interval, these approximation results then improve a lot and most theoretical values can be identified as just located in the high mass.

b. High computational cost

Besides this, as have been stated repeatedly, computational cost is also a major concern here. To obtain a brief idea of how expensive it actually was in this research, we now set out two examples. In the first simulation study where training data (DGP1) is estimated by ADCC-MGM, each loop of posterior sampling involves 600 instances of evaluating likelihood function and requires 5.6 minutes of computational time.[91] Thus, it would take a modern Intel *P4* computer more than 9.7 days for each chain to successfully generate a moderate number of states.[92] However, for estimating ADCC-MTM, the computational cost is even higher since degree of freedom parameters are now also included. The average computational time for this mixture model is about 6.5 minutes per iteration and it would take the same computer nearly 11.28 days to complete the overall calculation of Bayesian inference.

7.3 Correlation models comparison

Starting from this section, we examine the in-sample and out-of-sample performance of ADCC-MGM and ADCC-MTM and compare the results with a variety of alternative DCCs. Here, for competing models, firstly we consider the inclusion of correlation-targeting technique in some traditional DCC's dynamics, and then relax this assumption to see whether any improvements in model capability can be derived. Besides, two new multivariate asymmetric DCCs are also proposed in this chapter to contribute to the current literature.

[91] We obtain the number 600 because there are a total of 20 parameters in ADCC-MGM that require Griddy Gibbs sampler to simulate their new updates, and each simulation needs to evaluate 30 grid points.

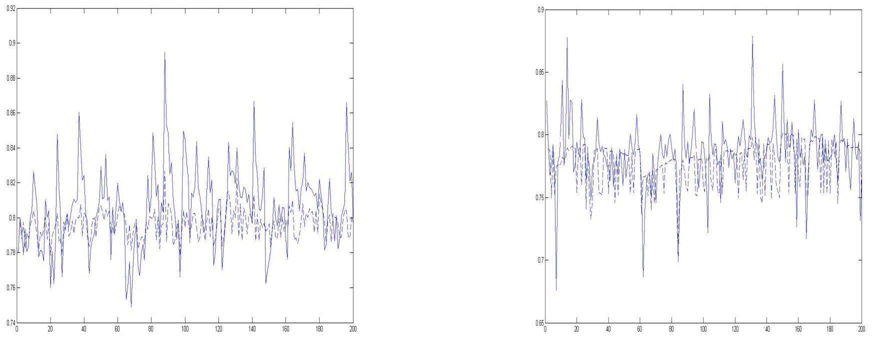
[92] Programming codes of this paper are written in Matlab and are available upon request.

7.3.1 In-sample and Out-of-sample analysis of Mixture models

First, concerning the in-sample analysis, one thing needs to be noted before proceeding. That is, this analysis is now performed only on simulated data because true parameter values of mixture model, only for these data, are obtainable prior to estimation so that realized correlations can be computed and compared to in-sample correlations. [93] Specifically, for each sample respectively simulated using DGP1 and DGP2 and fitted by ADCC-MGM and ADCC-MTM, we calculate predictive means of their last 200 conditional correlations and compare the results to realized correlations generated by applying true parameter values to corresponding mixture models. Since most of the chains, after burn-in period is eliminated, can now be confirmed as statistically converged, we obtain a total of 5000 equilibrium draws and for each $\varphi^{(i)}$ where $i \in [1 \quad 5000]$ we use equation (6.2) to generate a conditional correlation estimate. [94] Thus, the predictive mean is just equal to the sample mean of these estimates.[95]

Figure 7.8 Realized correlations and Predictive means of the last 200 conditional correlations estimated on simulated data

Panel A. ADCC-MGM (DGP1) **Panel B. ADCC-MTM (DGP2)**



Solid line denotes the sample paths of predictive means of the last 200 conditional correlations; dotted line represents the corresponding realized correlation.

From the above graph, it can be seen that (predictive means of) conditional correlations

[93] The same comparison for the empirical samples is not possible due to the lack of high-frequency data.

[94] In simulation study, posterior sampling has run a total of 15000 times. Given that burn-in period is now set at 10000, equilibrium draws then correspond to the last 5000. That is, we can obtain 5000 different φ after sampling, and each can be applied to a correlation mixture model, either ADCC-MGM or ADCC-MTM, to generate the conditional correlation estimate.

[95] Although we term this quantity as the predictive mean, it does not mean we are actually forecasting because conditional correlation estimates generated from different simulated φ are now given the in-sample advantage, i.e., the sampling of φ is based on the whole 2000 observations rather than 1800.

now follow a more volatile evolving process than realized correlation although their pattern of changes in the same time frame is quite similar. This result is as expected because, in section 7.2, it is already documented that the covariance stationary process estimated using simulated data is weaker than those assumed in DGP1 and DGP2.

To investigate out-of-sample performance, since parameter uncertainty is now allowed, by using a similar approach to the one just illustrated we generate the predictive densities of one-step-ahead correlation forecast, return forecast and VaR forecasts for both simulated and empirical data. Meanwhile, also provided is the predictive density of minimized variance of an authentic portfolio.[96] In Table 7.9, we report the summary statistics for these densities.

<Insert Table 7.9>

As can be seen, unconditional moments of correlation forecasts generated by two mixture models are now quite close to each other. For example, when stock index data is fitted to compute the next day's correlation, predictive results obtained using ADCC-MGM are 0.4395 for sample mean, 0.4293 for sample median and 0.4242 for sample mode. Corresponding statistics calculated by ADCC-MTM are respectively 0.4334, 0.4256 and 0.4111. This result is as expected because target models used for forecasting are already known to have many closely related characteristics and the only way to differentiate them is not by the mechanisms of how they generate correlation dynamics, but by their capability to account for extreme events.

Concerning the hedging performance, according to Table 7.9 ADCC-MTM in most cases is now found outperforming ADCC-MGM. The portfolio variance this mixture model can minimize is lower than the one calculated by assuming Gaussian mixture for unconditional returns and its performance is consistent whilst improvements are often found to be marginal in magnitude. Besides, from a risk manager's perspective, the superiority of ADCC-MTM over ADCC-MGM is also confirmed in the sense that extreme events now

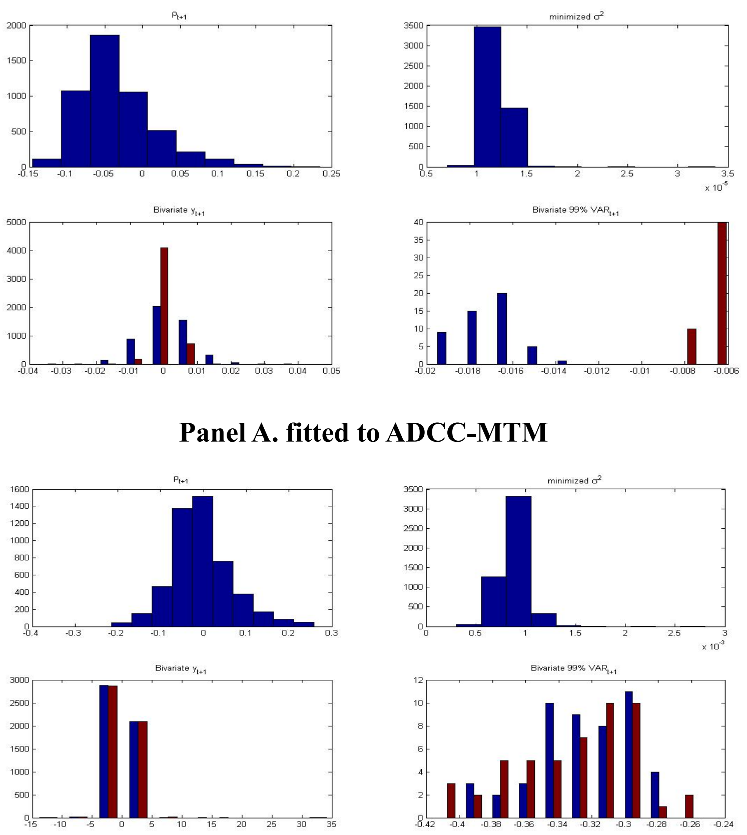
[96] This portfolio is constructed using two individual time series included in bivariate data.

can be more flexibility accounted using T mixture models. For example, when foreign exchange data is fitted, predictive mean of 99% next day's VaR calculated using ADCC-MGM are respectively -0.012 and -0.015, less than those implied by ADCC-MTM (-0.02807 and -0.02824) where one could even observe a loss of -0.0334 and -0.0342 in the worst scenario.

Next, to obtain a more concrete idea of how these predictive densities will distribute, we plot in Figure 7.9 their histograms using results generated from fitting stock and bond data for example.

Figure 7.9 Histogram plots of predictive densities of one-step-ahead correlation forecast, minimized variance, next day's return and VaR calculated for stock and bond data

Panel A. fitted to ADCC-MGM



Panel A. fitted to ADCC-MTM

Blue bar denotes the histogram statistics calculated for first time series included in bivariate data. Red bar corresponds to the second time series.

Here, it can be seen that, for most densities, it is now very easy to find a clear mode. However, only when next day's return forecast generated by ADCC-MGM is analyzed can a symmetric density be shown. As for their counterparts obtained using ADCC-MTM, outliers in large magnitude then become the dominating force in both tails of T distribution, suggesting that more capitals (VaR) than normal situations now need to be reserved to deal with the occurrence of these extreme events.

7.3.2 Fitting results of Traditional DCCs

Apart from the mixture models, a number of alternative DCCs are also fitted in this research to both simulated and empirical data to compare their performances in recovering time-varying correlations, generating forecasts and minimizing portfolio variances. The aim of making these comparisons is to see whether the increased sophistication introduced by mixture models is economically worthwhile. Specifically, to propose competitors to ADCC-MGM and ADCC-MTM, here we consider using some traditional MGARCH variants which assume Gaussian innovations such as Bollerslev's (1990) CCC, Engle's (2002) DCC, Hefner and Frances's (2003) ADCC and Capillio *et al.*'s (2004) AGDCC. In Appendix II, specifications of these correlation models have already been given. Now, we only use Gauss-Newton procedure to estimate them and illustrate the result.

a. Including Correlation Targeting

Before we proceed further, concerning the last two models, there is something that needs to be noted. For ADCC and AGDCC, in order to ensure the parsimony, we now incorporate correlation-targeting technique in their dynamics so that their covariance evolving processes can be modelled by

$$Q_t = (1-\eta^2-\varsigma^2)\bar{Q} - \iota^2\bar{N} + \eta^2\varepsilon_{t-1}\varepsilon_{t-1}^{'} + \varsigma^2 Q_{t-1} + \iota^2\vartheta_{t-1}\vartheta_{t-1}^{'} \qquad \text{(ADCC)} \qquad (7.2)$$

$$Q_t = \left(\bar{Q} - \eta^{'}\bar{Q}\eta - \varsigma^{'}\bar{Q}\varsigma - \iota^{'}\bar{N}\iota\right) + \eta^{'}\varepsilon_{t-1}\varepsilon_{t-1}^{'}\eta + \varsigma^{'}Q_{t-1}\varsigma + \iota^{'}\vartheta_{t-1}\vartheta_{t-1}^{'}\iota \qquad \text{(AGDCC)} \qquad (7.3)$$

where η , ς and ι in (7.2) respectively denote a scalar number whilst the same parameter in (7.3) represents a (2×2) diagonal matrix.[97] Besides, we constrain in ADCC $|(1-\eta^2-\varsigma^2)\bar{Q}-\iota^2\bar{N}|>0$ and for AGDCC $|\bar{Q}-\eta^{'}\bar{Q}\eta-\varsigma^{'}\bar{Q}\varsigma-\iota^{'}\bar{N}\iota|>0$ to ensure the positive definitiveness of the interception term. Since, in equation (7.2) and (7.3), other elements in Q_t are already know to be squared products, once these conditions are satisfied

[97] For example, for η in AGDCC model, it is now structured as $\begin{bmatrix}\eta_{11} & 0 \\ 0 & \eta_{22}\end{bmatrix}$.

the resultant covariance matrix will then become positively definitive. In addition, we also restrict the value of $\eta^2 + \varsigma^2$ in ADCC to less than one and $|\eta'\eta + \varsigma'\varsigma|$ in AGDCC to lie within the unit cycle so that the stationarity of resultant covariance can also be ensured.

In the following, we report the estimation results of fitting these correlation models. From Table 7.10 to Table 7.14, one can easily observe that DCC on most occasions now provides the best fit to simulated and empirical data, and estimated parameter values generated by using this model are frequently found to be significantly different from zero and imply strong correlation persistence.

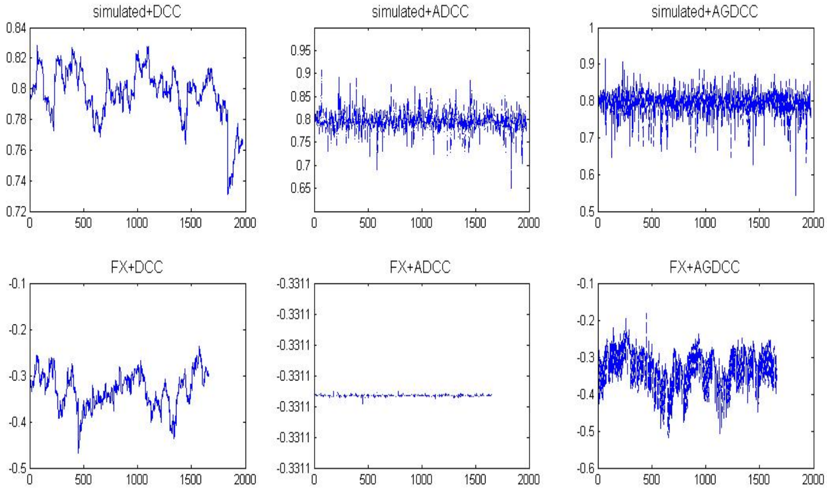
<Insert Table 7.10, 7.11, 7.12, 7.13 and 7.14>

However, here two things need to be noted. First, while fitting the empirical data, although we have introduced asymmetric news impact to DCC's specification, insufficient evidence is found to justify this assumption, suggesting that there is no leverage effect in correlation dynamics. Second, there is a subtle tendency in this research to favour the CCC model which assumes correlation to be constant rather than dynamic. Concerning these two issues, since similar results are to be documented again in a later part of this section, we leave our discussion until then.

b. Excluding Correlation-Targeting

Above, for estimating asymmetric DCCs, we included correlation-targeting. That is, to ensure parsimony, we let the conditional covariance converge to a pre-calculated long-term value. However, in a standard GARCH (see equation 2.5) this value is often computed by a need-to-estimate parameter to allow for more flexibility. Thus, an immediate drawback of imposing this targeting assumption is that the resultant correlation evolving process may be bounded within a small interval and centralized around a fixed value. Simply put, it will probably just not evolve as dynamically as we expected. To obtain the visual proof of this argument, see time series plots of dynamic correlation generated by DCC, ADCC and AGDCC for fitting simulated data (using DGP2) and foreign exchange rate data, for example.

Figure 7.10 Time series plot of dynamic correlation generated by fitting DCC, ADCC and AGDCC to second simulated data and FX data with correlation-targeting included



In the first row of Figure 7.10, we present the sample paths of various correlations calculated for simulated data, while the second row presents the results generated for foreign exchange data. Clearly, only DCC-generated correlations now can present an identifiable tendency of correlation changes whilst ADCC-generated correlations, at first glance, seem rather to be following standard mean-reverting processes or just remaining relatively constant in the second case (we will return to this point later). Concerning AGDCC correlation, its evolving process is very volatile, suggesting that further filtration might be needed for purification of information so that its correlation signals can be more clearly identified and extracted.

Given these features, it is then fair to say that both asymmetric DCCs now fail to produce the expected dynamics and there is motivation to slightly change their specifications so that more flexibility can be incorporated. For this purpose, we now consider relaxing the targeting assumption and introduce new interception parameters directly to the covariance dynamics. Specifically, by rewriting (7.2) and (7.3) to

$$Q_t = \varpi^2 + \eta^2 \varepsilon_{t-1}\varepsilon_{t-1}^{'} + \varsigma^2 Q_{t-1} + \iota^2 \vartheta_{t-1}\vartheta_{t-1}^{'} \quad \text{(ADCC)} \qquad (7.4)$$

$$Q_t = CC' + \eta^{'} \varepsilon_{t-1}\varepsilon_{t-1}^{'}\eta + \varsigma^{'} Q_{t-1}\varsigma + \iota^{'} \vartheta_{t-1}\vartheta_{t-1}^{'}\iota \quad \text{(AGDCC)} \qquad (7.5)$$

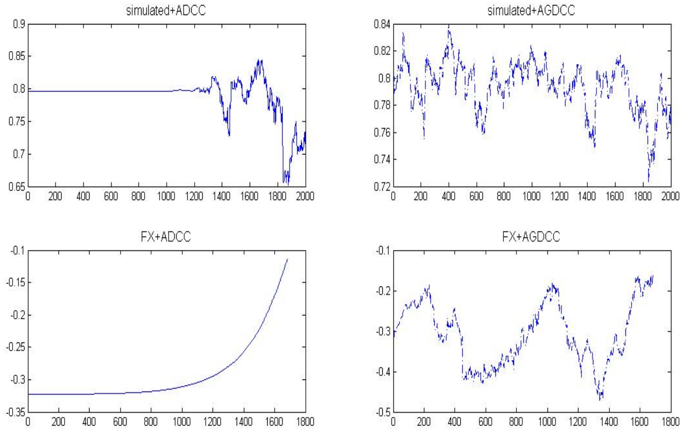
where ϖ denotes a scalar number and C represents a (2×2) matrix transformed by vech(.) function of a column vector with a total of 3 elements, it is now expected that a less restrictive model will lead calculated correlations to exhibit behaviour more like the true correlation, although these quantities themselves are unobservable in the real markets.

In equation (7.4), positive definitiveness of covariance matrix is guaranteed by its unique squared parameter settings. However, for AGDCC, to maintain the same property, it is then required $C^{'}C$ is kept positive definitive. Since decomposition of covariance matrixes, if we leave aside the asymmetric factor, can now be written as,

$$\begin{array}{ll} ADCC: & AGDCC: \\ & q_{ii,t} = c_{11}^2 + \eta_{11}^2 \varepsilon_{i,t-1}^2 + \varsigma_{11}^2 q_{ii,t-1} \\ q_{ii,t} = \varpi^2 + \eta^2 \varepsilon_{i,t-1}^2 + \varsigma^2 q_{ii,t-1} & q_{jj,t} = c_{22}^2 + \eta_{22}^2 \varepsilon_{i,t-1}^2 + \varsigma_{22}^2 q_{ii,t-1} \\ q_{jj,t} = \varpi^2 + \eta^2 \varepsilon_{i,t-1}^2 + \varsigma^2 q_{ii,t-1} & q_{ij,t} = q_{ji,t} = c_{11}c_{12} + \eta_{11}\eta_{22} \varepsilon_{i,t-1}^2 + \varsigma_{11}\varsigma_{22} q_{ij,t-1} \\ q_{ij,t} = q_{ji,t} = \varpi^2 + \eta^2 \varepsilon_{i,t-1}^2 + \varsigma^2 q_{ij,t-1} & C = \begin{pmatrix} c_{11} & 0 \\ c_{12} & c_{22} \end{pmatrix}, \eta = \begin{pmatrix} \eta_{11} & 0 \\ 0 & \eta_{22} \end{pmatrix}, \iota = \begin{pmatrix} \iota_{11} & 0 \\ 0 & \iota_{22} \end{pmatrix} \end{array} \quad (7.6)$$

this is then equal to saying that both c_{11} and c_{12} are required to have the same sign. To see whether the implementation of this strategy can improve the model performance of asymmetric DCCs, now we fit these two models to the same empirical data used above and plot the resultant correlation in the following.

Figure 7.11 Time series plot of dynamic correlation generated by fitting ADCC and AGDCC to second simulated data and FX data with interception term included



As before, the first row gives the fitting results of simulated data while the second presents the sample path of time-varying correlation calculated for exchange rate data.

From the graph, it can be seen that a more dynamic process for correlation can now be captured through the slight change in parameterization and its evolving process is no longer wandering around a fixed value (unconditional correlation), but showing upside-down changes. This decentralizing effect is especially evident for AGDCC model where a BEKK-type specification is imposed for modelling covariance dynamics. Here, given this dynamic feature, interpretation of the correlation results becomes much easier. See the second row and second column of the above chart for example (exchange rate data fitted by AGDCC model with interception term included): correlation now remains relatively stable for the initial period and then experiences a gradual change (either going up or going down) thereafter. This pattern accords with most empirical evidence documented in financial literature that support the findings of a strongly persistent correlation evolving process during the tranquil period. For risk managers, a correlation model, capable of capturing features like this, is especially useful. As has been stated previously, in finance, understanding the correlation risk correctly and hedging this risk properly is by no means an easy task. If realized correlation follows a process like that shown in Figure 7.10, definitively, it will be very difficult for a risk manager to react correctly to the given information since only mean reverting process is implied.

In addition to AGDCC, a clear sign of correlation changes can also be observed when correlation-targeting technique is removed from ADCC. For example, when this model is fitted to exchange rate data (second row, first column of Figure 7.11), a smoothly evolving process, with correlation steadily increasing its value from -0.3 to -0.1 during the whole sample period, is observed. Although, in the initial stage, a period of constant correlation still presents itself, it does not take long before this coefficient finally switches itself to a dynamic process. Given this evidence, we now fit again all five sample sets of data using asymmetric DCCs. Their parameter estimation results, together with one-step-ahead correlation forecasts are presented in Table 7.15 and Table 7.16 respectively.

<Insert Table 7.15 and 7.16>

First, for ADCC, as can be seen from Table 7.15, except for ς which governs the correlation persistence, other parameters included in the modelling of covariance dynamics are now found insignificantly different from zero.[98] As for AGDCC, similar results are also generated although one exception still needs to be noted. That is, when stock index data is fitted, most of the parameter values are now large enough to reject the null (parameter equalling zero). And this is the first time we successfully document evidence for leverage effect in correlation dynamics. That is, in equity markets, when overall market goes down, more shares tend to move in the same direction than in the case when the index goes up by the same magnitude. This result is as expected because asymmetric correlation is a stylized feature in equity market.

c. Comparison of ADCC with AGDCC

If we now summarize the information just provided for two asymmetric DCCs (comparison of ADCC and AGDCC with correlation-targeting included and excluded), one may be easily tempted to draw a conclusion that ADCC as a correlation model is, actually, much less flexible than AGDCC. Indeed, if we look back to Figure 7.10 and Figure 7.11, ADCC seldom leads its generated correlations to evolve dynamically, and constant correlation as a result of this model is not rare. Compared to AGDCC, its poorer performance can now be partially explained by the reduced number of parameters included in its specification for modelling.[99] Meanwhile, as have been implied in equation (7.6), it maybe also due to the common dynamics assumed for its individual variances $q_{ii,t}$, $q_{ij,t}$ and joint covariance $q_{ij,t}$.[100]

However, if we take a step further, more encouraging results can be generated. Take the occasional case where ADCC produces constant correlations for example (see second

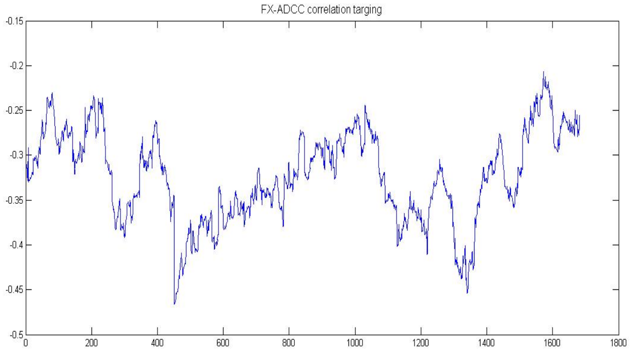
[98] Here, we cannot, based on this feature, say that estimation result of ADCC is then not good. It is because the standard error, from which *t*-statistics (the criterion of determining the significance of a parameter different from zero) are calculated, is now computed using inverse Hessian matrix. And this Hessian matrix is derived in optimization step of ML rather than through an explicitly analytical form. Thus, the results could be spurious.

[99] For example, in ADCC only three/four elements (targeting included/excluded) are incorporated to modelling correlation evolving process whilst the similar dynamics in AGDCC need to be estimated using six/nine parameters.

[100] In ADCC, $q_{ii,t,}$, $q_{jj,t}$ and $q_{ii,t,}$ are modeled using the same system of parameters.

row and second column of Figure 7.10). If in (7.2) the positive definitiveness constraint $|(1-\eta^2-\varsigma^2)\bar{Q}-\iota^2\bar{N}|>0$ is now replaced by $(1-\eta^2-\varsigma^2)\bar{Q}-\iota^2\bar{N}>0$, that is to constrain all elements, rather than the determinant, in the two-by-two interception matrix to be positive, dynamic property of the correlation process for ADCC then can be easily re-obtained (see below).

Figure 7.12 Exchange rate data estimated by ADCC model with correlation-targeting included and modified positive definitiveness constraint



Besides, compared to previously, its parameter estimation results also improve a lot.

Volatility parameters						
	ω_1	α_1	β_1	ω_2	α_2	β_2
value	1.12E-06	0.057319	0.89888	1.68E-06	0.083524	0.8899
s.t.d	(8.76E-14)**	(0.00013)**	(0.000315)**	(3.76E-13)**	(0.00019)**	(0.00044)**
	Correlation parameters					Logliklihood
	η	ς	ι			-12609
value	8.81E-02	0.989	2.59E-02			
s.t.d	(5.82E-02)**	(0.01904)**	0.15265			

As can be seen, except for asymmetry factor, all other parameters included in covariance equations are now found to be positive and significantly different from zero. In addition, under this new constraint the resultant covariance is kept positive definitive although it is not required to be so.[101] Given this evidence, to obtain a good fitting result, the necessity of properly tuning the constraints before estimating ADCC is thus highlighted. However,

[101] Although it is preferred we can manually control the parameterization in optimization, imposing a restriction like $(1-\eta^2-\varsigma^2)\bar{Q}-\iota^2\bar{N}>0$ for ADCC is not a sufficient condition to ensure positive definitiveness because, even with all positive elements, interception term $(1-\eta^2-\varsigma^2)\bar{Q}-\iota^2\bar{N}$ could still have a negative determinant. However, in this particular case the determinant of this matrix is now found to be positive.

in terms of the flexibility, undoubtedly, it is still the AGDCC that provides the better performance.

7.3.3 Asymmetric DCCs with *t* and skew *t* innovations

Above, we have examined the model performance of a range of traditional DCC variants. Now, to increase their flexibility, we consider introducing more sophistication, e.g. to combine ADCC/AGDCCC with a multivariate fat-tailed distribution. Note that, in this research such attempt has already been made through the implementation of two mixture models. However their inferences are studied in a Bayesian framework with massive computational cost associated. Here, to take another look at ADCCs by enhancing their tail behaviors. We consider incorporating a multivariate *t* distribution and one of its skewed version and use a classically inferential approach to estimate the proposed models. Similar experiments of combining a fat-tailed and skewed parametric distribution with dynamic correlation models have already been studied in the literature. For example, Cajigas and Urga (2005) combined GDCC with an asymmetric Laplace distribution. Examples of a multivariate elliptical distribution with symmetric DCC are illustrated in Pelagatti and Rondena (2004).

a. ADCC and AGDCC with multivariate *t*

For our purposes, first we assume innovations of ADCC/AGDCC to be bivariate *t* distributed. Since log-likelihood function of a model with *t* errors can be written as

$$\begin{aligned} L(.\mid\varphi) &= \sum_{t=1}^{T}\log f\left(y_t \mid \varphi\right) \\ &= \sum_{t=1}^{T}\left\{-\log(2\pi)-\frac{1}{2}\log\left|\Sigma_t\right|-\frac{v+2}{2}\log\left(1+\frac{(y_t-\mu)^{'}\Sigma_t^{-1}(y_t-\mu)}{v}\right)\right\} \end{aligned} \tag{7.7}$$

we can easily obtain

$$L(.\mid\varphi)=\sum_{t=1}^{T}\left\{-\log(2\pi)-\log\left|D_t\right|-\frac{1}{2}\log\left|R_t\right|-\frac{v+2}{2}\log\left(1+\frac{\varepsilon_t^{'}R_t^{-1}\varepsilon_t}{v}\right)\right\} \tag{7.8}$$

after DCC's unique covariance equation $\Sigma_t = D_t R_t D_t$ and error term $\varepsilon_t = (y_t-\mu)D_t^{-1}$ are inserted to (7.7). Then, volatility parameters θ and correlation parameters ψ can be respectively calculated after this log-likelihood function is maximized with respect to D_t and R_t, using Gauss-Newton approach. That is, for estimating ADCC/AGDCC, we decompose $L\left(.\,|\,\varphi\right)$ into two functions and maximize them separately. Here, one function is

$$L_\theta\left(.\,|\,\varphi\right) = -\sum_{t=1}^{T}\left\{\log(2\pi) + \log\left|D_t\right|\right\} \tag{7.9}$$

the other is

$$L_\psi\left(.\,|\,\varphi\right) = -\frac{1}{2}\sum_{t=1}^{T}\left\{\log\left|R_t\right| + (\nu+2)\log\left(1+\frac{\varepsilon_t^{'}R_t^{-1}\varepsilon_t}{\nu}\right)\right\} \tag{7.10}$$

Given this information, we now present the fitting result of ADCC-*t* and AGDCC-*t*, with correlation-targeting not included, to five sample data.

<Insert Table 7.17 and 7.18>

As can be seen, model performance, after fat tails are incorporated, now improves a lot compared to those where only Gaussian innovations are assumed. For example, calculated portfolio variances become much smaller than those reported in the previous cases. However, for some particular samples, such as foreign exchange data estimated using AGDCC-*t*, the fitting result is still not quite good, with none of its parameters found capable of generating a different-from-zero value.

a. ADCC and AGDCC with multivariate skew *t*

Here, apart from the symmetric distribution, to exploit the ADCC mechanism on a further basis, we also utilize a result documented in Fernandez and Steel (1998) and generalized by Bauwen and Laurent (2002) to see whether the introduction of a skewness factor to above *t* will again improve the fitting and forecasting results of dynamic correlation models. Concerning this topic, some recent works, contributing to the generalization of a symmetric distribution to a skewed one, need to be briefly reviewed first.

As we know, it has been a challenge for a while for econometricians to design a multivariate distribution that is both easy for inferential use and compatible with skewness and kurtosis of financial returns. Many efforts in this area are put onto searching for a new parametric function, different from standard ones, to fit the empirical data. Typically, one can choose either an asymmetric Laplace distribution, a hyperbolic distribution or a normal inverse Gaussian. Besides, in some research it is also suggested that we can use standard ones as base distribution to introduce non-linear dynamics using additional parameters. For example, in the univariate context Hansen (1994) proposed a skewed version of student *t*. By changing the scale of third moments (skewness) at each side of mode, Fernandez and Steel (1998) developed a similar method to introduce asymmetry to any continuous and unimodal distributions and the skewed normal discussed in their paper was soon generalized to other versions (See Lambert and Laurent, 2000, and Jones and Faddy, 2000). In the multivariate context, the first skewed Gaussian was proposed in Azzalini and Capitanio (1996) where the authors used a combination of a *p.d.f* and a *c.d.f* to form an asymmetric density. Branco and Dey (2000), based on their works, introduced a general class of multivariate skew-elliptical distribution, Arnold and Beaver (2000) proposed the multivariate skew Cauchy, Azzalini and Capitanio (2003) studied the multivariate skew *t*. Bayesian inference of the same model is calculated in Lochos, Cancho and Aoki (2008). Recently, skewed versions of these standard distributions have also been used to form mixture distribution. For example, to handle highly asymmetric data, Wang, Ng and McLachlan (2009) developed multivariate skew-*t* mixtures using EM as an inferential tool. Lin (2009) derived the maximum likelihood estimator for parameters in multivariate skew normal mixture distribution.

Concerning our research purpose, we now use Bauewen and Laurent's (2002) skew *t* to enhance DCC models. In terms of the generality, this distribution is so flexible that can nest a variety of alternatives such as normal, student *t*, Cauchy, skew-normal, skew-cauchy. Since only bivariate data is to be analyzed, its density function can be defined as

$$f(y^* \mid \zeta, \upsilon) = 4\frac{\zeta_1}{1+\zeta_1^2}\frac{\zeta_2}{1+\zeta_2^2} t_v(\kappa^* \mid \mu, \Sigma) \tag{7.10}$$

where $t_v(\cdot \mid \mu, \Sigma)$ denotes the *p.d.f* of a standard *t*, ζ represents the skewness parameter, $\kappa^* = (\kappa_1^*, \kappa_2^*)'$ where $\kappa_i^* = y_i^* \zeta_i^{-I_i}$ for *i*=1,2 and *I_i* equals one if y_i^* is positive and minus one otherwise. Usually, since empirical data before filtration is not centered on zero, standardization is required to transform the raw data first so that (7.10) can be fitted. Based on this virtue, we now consider using the method suggested in Fernandez and Steel (1998) to obtain standardized innovation. That is, we calculate $y = (y^* - m)./s$ where mean *m* and variance *s* are respectively given by

$$m_i = \frac{\Gamma(\frac{v-1}{2})}{\Gamma(\frac{v}{2})}\sqrt{\frac{v-2}{\pi}}\left(\zeta_i - \frac{1}{\zeta_i}\right) \quad \text{and} \quad s_i^2 = \left(\zeta_i^2 + \frac{1}{\zeta_i^2} - 1\right) - m_i^2 \qquad (7.11)$$

Thus, in an expanded form, (7.10) can also be defined using

$$f(y \mid \zeta, \upsilon) = \frac{4}{\pi}\frac{\zeta_1 s_1}{1+\zeta_1^2}\frac{\zeta_2 s_2}{1+\zeta_2^2}\frac{\Gamma((v+2)/2)}{\Gamma(v/2)(v-2)}\left(1+\frac{\kappa'\kappa}{v-2}\right)^{-(v+2)/2} \qquad (7.12)$$

where

$$\kappa = (\kappa_1, \kappa_2)'$$
$$\kappa_i = s_i y^* + m_i \zeta_i^{-I_i} = (s_i y_i + m_i)\zeta_i^{-I_i}$$
$$I_i = \begin{cases} 1 & if \quad y_i \geq -m_i / s_i \\ -1 & if \quad y_i \leq -m_i / s_i \end{cases}$$

Here, to understand this asymmetric distribution more clearly, it is necessary to bear in mind the concept of *M*-symmetry for a multivariate density. As Bauwen and Laurent (2002) illustrated, *"... a unimodal density g(x) defined on R^k (k dimension) is symmetrical if and only if for any x, g(x)=g(Qx), for all diagonal matrix Q whose diagonal elements are equal to 1 or to -1…*" . Thus, in a bivariate case it is required that

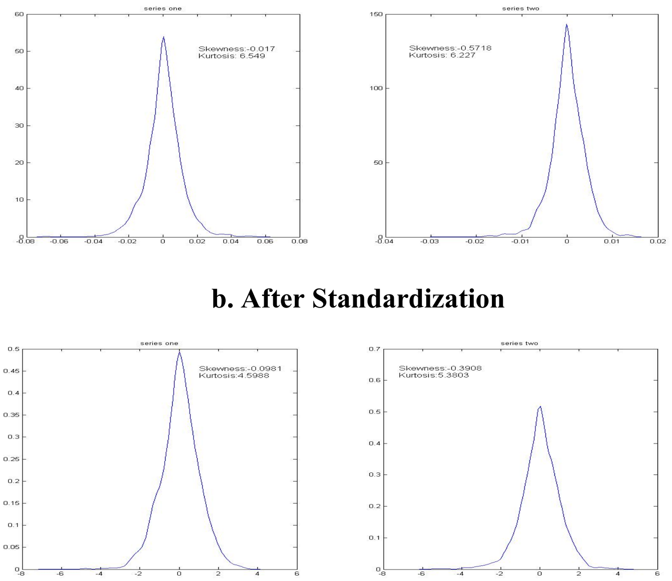
$$g(x_1, x_2) = g(-x_1, x_2) = g(x_1, -x_2) = g(-x_1, -x_2)$$

And, in maximum likelihood, four situations then need to be categorized and evaluated before a realization of x_i can be input to log-likelihood function. As for ζ^2, since it has been shown in (7.11) that this variable now determines the ratio of probability masses above and below the mode, skewness then can be defined based on it. For example, if ζ is significantly larger (less) than one, target distribution is then said to be skew to the right (left), and observations have a tendency to generate positive (negative) skewness. Analogously, ζ equaling one is an indication of symmetric distribution. (See Lambert and Laurent, 2000, and Bauwen and Laurent, 2002, for more interpretation of using $\log \zeta$ as an indication of skewness).

Moreover, since the sign of x_i is now an essential factor relating to the calculated skewness, a proper transformation of the original observations is needed before estimation. Since raw asset returns y now need to be transformed $y \rightarrow y^*$ and standardized $y^* \rightarrow \kappa$ before input to skew *t*, a potential problem for estimation then arises. Specifically, in calibrating DCC models, usually we adopt a two-step estimation procedure to fit the sample data. This method is valid because univariate volatility D_t and correlation R_t have their own parameters and these parameters will not contaminate the log-likelihood function specifically given for the other component of covariance matrix. For example, GARCH parameters used to fit D_t are not related to the inference concerning any parameters governing $R_{t.}$ Thus, their log-likelihood function is separable for estimation. However, in this case where a skewed *t* is assumed, asset returns then need to be properly transformed before they can be input to calculating D_t and $R_{t.}$ Since this transformation is now determined by a common factor ζ , decomposition of the log-likelihood function might be improper. However, it is not equal to saying that the two-step procedure suggested in Engle (2002) will then produce invalid parameter estimates. The major concern here is only 'when to impose this transformation'. As known from previous chapters, DCC correlation R_t is often modelled using a univariate GARCH where innovations are asset returns standardized by calculated $D_{t.}$ Since it is widely accepted that returns, after being standardized by GARCH volatility, will only show a lesser degree of fat tails (since GARCH can help capture the volatility clustering) whilst leaving skewness in most cases unchanged, there is then a motivation to apply the transformation not in the process of generating D_t but after return is standardized. That is, in step $\varepsilon_t = (y_t - \mu)D_t^{-1}$. To see the visional proof for this argument, we present below the comparison of kernel density plot of, say, stock and bond data before and after standardization by univariate GARCH volatility.

Figure 7.13 Kernel density plot of stock and bond data before and after standardization by univariate GARCH volatility

a. Before Standardization



b. After Standardization

From kurtosis estimates, one can easily confirm the downgrade of high peakedness for returns after being standardized by GARCH volatilities. However, as for skewness, no significant change is then observed, suggesting that two-step procedure is still a valid, but probably not very efficient, way of calculating DCCs' inference.[102]

Given this feature, we now illustrate the estimation procedure of ADCC-*skew-t* and AGDCC-*skew-t* in the following. As before, we first define the whole log-likelihood function for $\varphi = (\theta, \psi, \zeta, v)$. That is,

$$
\begin{aligned}
L_{\varphi}\left(.\,|\,\varphi\right) &= \sum_{t=1}^{T} \log f\left(y_t\,|\,\varphi\right) \\
&= \sum_{t=1}^{T}\left\{ \log(\frac{4}{\pi}) + \log[\frac{\zeta_1 s_1 \zeta_2 s_2}{(1+\zeta_1^2)(1+\zeta_2^2)}] + \log[\frac{\Gamma((v+2)/2)}{\Gamma(v/2)(v-2)}] - \frac{1}{2}\log\left|\Sigma_t\right| \cdots \right. \qquad (7.13) \\
&\quad \left. \cdots - \frac{v+2}{2}\log(1+\frac{(y_t-\mu)^{'}\Sigma_t^{-1}(y_t-\mu)}{v-2}) \right\}
\end{aligned}
$$

Then, we derive decomposed functions $L_{\theta}\left(.\,|\,\varphi\right)$ and $L_{\psi}\left(.\,|\,\varphi\right)$ respectively corresponding to volatility parameters and correlation parameters.[103] Here, since degree of freedom

[102] Here, it is necessary to bear in mind that parameters calculated from maximizing an un-decomposed log-likelihood function would be statistically more consistent, optimal and desirable in this particular case.

[103] In some textbooks, (v-2) is replaced by v so that $\log[\Gamma((v^{'}+2)/2)/\Gamma(v^{'}/2)v] = \log(0.5)$ and this quantity in estimating $L_{\psi,\zeta,v}\left(.\,|\,\varphi\right)$ can be eliminated.

parameter *v*, skewness factor ζ and $L_{\psi}\left(.\,|\,\varphi\right)$ can be defined in one single function, we use $L_{\psi,\zeta,v}\left(.\,|\,\varphi\right)$ to denote it. Thus,

$$L_{\theta}\left(.\,|\,\varphi\right)=\sum_{t=1}^{T}\left\{\log(\frac{4}{\pi})-\log\left|D_{t}\right|\right\} \tag{7.14}$$

and

$$\begin{aligned}L_{\psi,\zeta,v}\left(.\,|\,\varphi\right)=-\frac{1}{2}\sum_{t=1}^{T}\Bigg\{\log\left|R_{t}\right|+(v+2)\log(1+\frac{\varepsilon_{t}^{'}R_{t}^{-1}\varepsilon_{t}}{v-2})-2\log[\frac{\zeta_{1}s_{1}\zeta_{2}s_{2}}{(1+\zeta_{1}^{2})(1+\zeta_{2}^{2})}]\cdots\\ \cdots-2\log[\frac{\Gamma((v+2)/2)}{\Gamma(v/2)(v-2)}]\Bigg\}\end{aligned} \tag{7.15}$$

can be obtained after $\Sigma_{t}=D_{t}R_{t}D_{t}$ is input to (7.13) and return is standardized by $\varepsilon_{t}=(y_{t}-\mu)D_{t}^{-1}$ and transformed by a function of ζ . Here, if covariance matrix in (7.13) is modelled by (7.4), parameter values of ADCC-*skew-t* can be obtained after (7.14) and (7.15) are respectively maximized with respect to θ and $(\psi,\varsigma,\upsilon)$. Those of AGDCC-*skew-t* can be calibrated if (7.5) is now utilized.

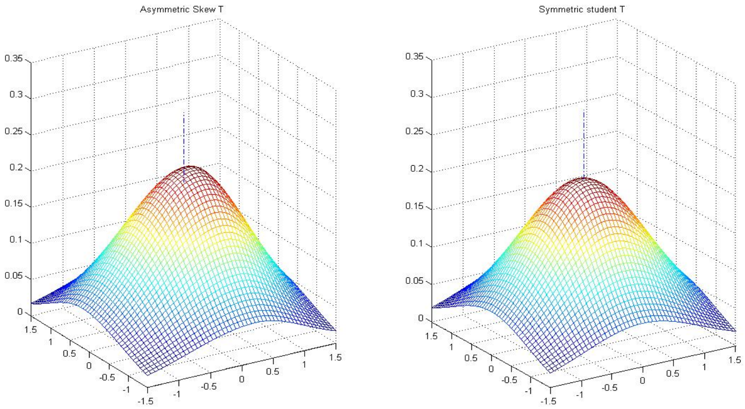
Next, in Table 7.19 and Table 7.20 we present the fitting results of these models to simulated and empirical data.

<Insert Table 7.19 and 7.20>

As can be seen, compared to previously, more parameters are now able to generate a sensible value which is significantly different from zero. Outperformance of skew-*t* versions of asymmetric DCC models over the ones assuming either Gaussian innovations or symmetric *t* innovations is documented throughout the cases and this outperformance is usually reflected through a reduced value for portfolio variance in optimization after skewness in asset return distribution is taken into account. For example, when the second simulated data is analyzed, minimized variance generated using ADCC-skew *t* is 0.5671, much less than those (0.6567 and 0.7892) generated by ADCC-*t* and ADCC-Gaussian (reported in Table 7.17 and Table 7.15). Concerning the skewness, although each marginal is now given a specific coefficient, only in fitting stock index data is significant evidence for asymmetry (negative skewness) observed.

Now, to obtain a brief idea of how symmetric *t* and skewed *t* would really affect the distributional characteristics of multiple returns being modelled, we use stock index data, for example, to present mesh plots using these two densities. Parameters (ζ, ν, Σ) used to generate these plots are obtained from Table 7.18 and Table 7.20 respectively. Besides, also presented is a dotted line, representing (0,0,Z) in 3-D surface, as a reference to the bivariate symmetry.

Figure 7.14 Mesh plot of multivariate symmetric *t*- and skew *t*- distributions generated from parameter estimation results of stock index data.



As can be seen from the above graph, on the right-hand side where a symmetric *t* is presented, the distribution shows clear evidence of symmetry around the reference line. However, when the left one is analyzed, bivariate density then clearly skews to the negative observations (peakedness is obtained on the positive domain), which confirms again most of the empirical findings that negative returns are more likely to be observed in equity market than positive ones. Concerning the peakedness of these two densities, they roughly stay at around 0.15, a value much lower than those observed in the simulated data (Figure 7.1) where the appearance of 0.3 or even 0.4 is not unusual. This is as expected because, after estimation, stock index data show a much lower degree of higher moments than those manually assumed in simulated ADCC-MGM data and simulated ADCC-MTM data. And the left one (multivariate skew *t*) presented a slightly higher degree of peakedness than the right one because, in a skewed version, asset returns need to be contaminated with a function of skewness factor and degree of freedom to reinforce the higher moments such as extra kurtosis so that these contaminated returns can be input to a standard symmetric *t* with skewness factor again added to tilt the distribution.

7.3.4 Comparison of portfolio variance

Above, while we propose competitors for ADCC mixture models, a range of comparisons of model performance among these alternatives have already been launched. For example, we have compared the correlation dynamics with and without targeting techniques incorporated, ADCC- structure with AGDCC- structure and correlations generated by mechanism assuming different distributions. Here, since our main aim is to see whether the increased sophistication introduced by mixture models is economically worthwhile and in this thesis we use minimized portfolio variance as the main tool for discriminating between different models, a summary of this result is then provided below.

<Insert Table 7.21>

From Table 7.21, it can be clearly seen that ADCC mixture models now perform the best among all DCC variants in terms of being able to generate the lowest portfolio variance. Averagely speaking, ADCC-MTM, which has the most sophisticated mechanism assumed for its correlation evolving process in this research, is also the most capable model to produce a portfolio which can generate the stable profit/loss. For example, if stock and bond data is used to construct a portfolio, the overall risk (portfolio variance) generated by using this mixture model is only 0.0009 whilst that generated for ADCC-MGM is 0.0012 and 0.00318 for ADCC-G. Here, note that compared to the asymmetric DCC model with only one Gaussian component incorporated, this mixture model now successfully reduce the portfolio variance a substantial amount (nearly 71.9%). Similar evidence can also been observed when exchange rate data is fitted. However, as for the stock index data and simulated data, outperformance of mixture model over its competitors is then not evident any more. Similar values for quantifying the portfolio risk are derived by all types of correlation models. However, it is fair to say that, generally speaking ADCC mixture model is still the best performer given that portfolio returns are now all set to be equal.

7.3.5 Comparison of correlations and returns

a. Comparison of In-Sample correlations

In previous sections, we have analyzed the correlation dynamics under different scenarios. Now, to obtain a more detailed idea of their relative performance, we plot their sample paths. First, for ADCC and AGDCC, whose innovations are assumed to be bivariate Gaussian-distributed, we calculate their dynamic correlations for two simulated and three empirical data. Then, the same models, combined with multivariate *t* and multivariate *skew-t,* are estimated with correlation dynamics respectively plotted in Figure 7.16 and 7.18. Here, note that, for these models, we use a constant term, instead of targeting, to model interception parameter. Finally, as a comparison, sample paths of ADCC-MGM-generated correlation and ADCC-MTM-generated correlation are also presented.

Figure 7.15 Time series plot of dynamic correlation generated from ADCC-*Gaussian* and AGDCC-*Gaussian* model

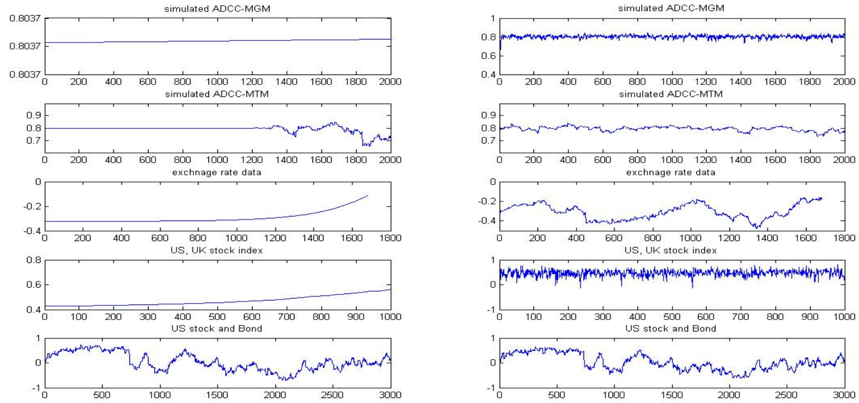


Figure 7.16 Time series plot of dynamic correlation generated by ADCC-*t* and AGDCC-*t*

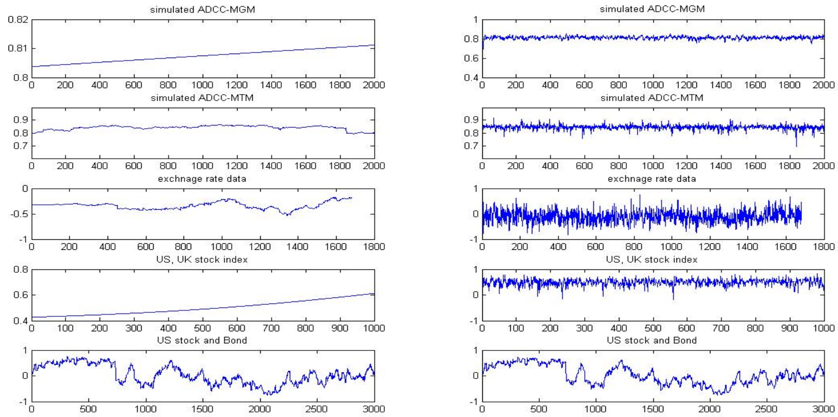


Figure 7.17 Time series plot of dynamic correlation generated from ADCC-*skew t* and AGDCC-*skew t*

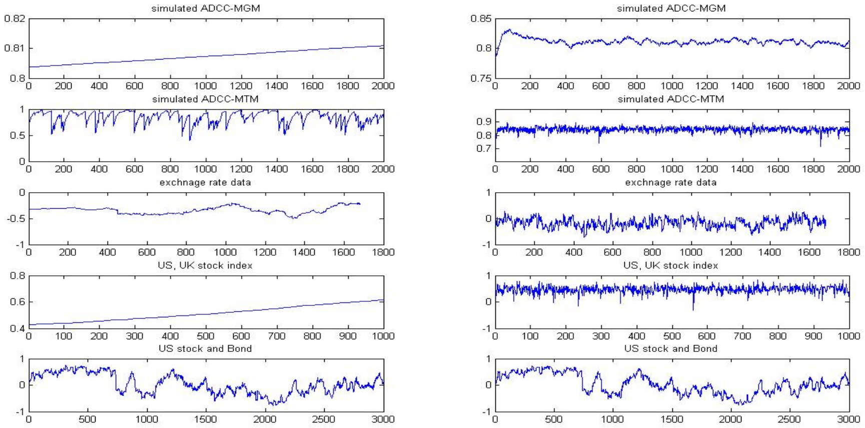
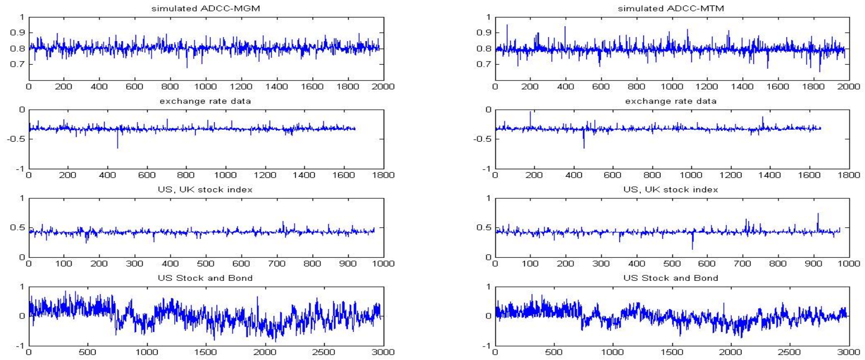


Figure 7.18 Time series plot of dynamic correlation generated from ADCC-MGM and AGDCC-MTM model



From Figure 7.15 to 7.18, the diagrams presented in the first column report the correlations generated by ADCC model. Analogously, those plotted in the second then correspond to AGDCC correlation. As can be seen, most of the bivariate relationships considered above do not fluctuate significantly and a smooth evolving process for correlation is observed throughout the time. The only exception here is for stock and bond data where the dynamic property of correlation can be confirmed on a consistent basis in all ADCC and AGDCC results.

b. Autocorrelation test for standardized return and volatility

Apart from the correlation, in this research we are also interested in seeing whether the asymmetric DCCs are adequate for capturing the dynamics in conditional returns and conditional volatilities. For examining these properties, two hypothesis tests are carried out. First, we perform the Jarque-Bera test to examine univariate normality of innovations after they are standardized by $\varepsilon_t = \Sigma_t^{-1/2}(y_t - \mu)$. Then, Box-Pierce statistics are calculated with 20 lags on resultant residuals and squared residuals to see whether autocorrelation is present. Here, for both tests, we set the significance level to 95% and report *p*-values in Table 7.22. Similar tests for raw (un-standardized) returns have already been performed and reported in summary statistics (See Table 7.1 and 7.5 for details).

<Insert Table 7.22>

Clearly, from the table, univariate normality is now firmly rejected in the majority of cases. This result is as expected because it is coherent with most empirical findings, that is, GARCH volatility can only account for, to certain degree, extra kurtosis exhibited in unconditional returns. And there are still some unaccounted factors that need to be taken care of in the modelling of conditional second or even higher moments. Concerning the autocorrelation result, it can be seen that randomness of standardized innovations and their second moments (square of standardized residuals) are now confirmed for both simulated data. However, for empirical data, a different situation then arises. Strong evidence is found in standardized exchange rate returns and their volatility to accept the null hypothesis of zero autocorrelation. However, for US and UK stock indexes it is then rejected with a close-to-zero *p*-value, suggesting that a higher order of autoregressive- AR or moving average- MA lag (larger than two) is now needed to enhance the mean equation to take extra serial dependence in conditional returns into account. However, such evidence for conditional volatility is not very prominent.

In the univariate context, since it is now known that, except for stock index data, asymmetric DCCs perform sufficiently well to strip the serial dependence in conditional mean (return) and conditional variance (volatility), it is then also interesting to see whether this performance will hold when multivariate cases are examined. For example, we can test whether the unconditional correlation calculated from standardized innovations will still remain at a level similar to those generated from unstandardized ones and whether the calculated correlation after return, already filtered by a dynamic correlation model, will still present a strong dynamic property or, in a similar vein, whether our asymmetric DCCs used here are good enough to capture all the dynamics implied in the unconditional correlation's evolving process.

c. Constant correlation test for standardized return and volatility

For this purpose, we now compare the unconditional correlation of sample data before- and after- standardization by GARCH volatilities and exploit a result from Engle and Sheppard (2001) to test the constant correlation hypothesis for the same sample period. Concerning this test, the null is now set to be *H₀*: $R_t = \overline{R}$ for $\forall \mathrm{t}$ and we test it against *H₁*: $\beta_1^* = \beta_2^* = \cdots = \beta_n^*$ in $X_t = \beta_0^* + \beta_1^* X_{t-1} + \beta_2^* X_{t-2} + \cdots + \beta_n^* X_{t-n}$ for all *n* lags where $X_t = vech^n(\varepsilon_t \varepsilon_t^{'} - I)$, ε_t is standardized residuals and *vech*^{n} is a *vech* operator only selecting elements under the main diagonal (for a similar test for constant correlation, see also Tse, 2000). Results of this test are documented in Table 7.23.

<Insert Table 7.23>

From the table, strong evidence now can be observed for a consistently good performance of asymmetric DCCs in modelling correlation dynamics. Before returns are standardized, unconditional correlation usually stays at a relatively stable level. For example, for simulated ADCC-MGM data and stock index data, it respectively equals 0.8 and -0.31. However, after ADCC and AGDCC are fitted with a range of distributional assumptions such as *Gaussian, t* and *skew t,* and returns are standardized by their calculated covariance, this quantity then immediately approaches zero in all of the cases, suggesting that correlation dynamics are now sufficiently well captured by given DCC models after filtration. To obtain a more objective opinion, here we perform Engle's constant correlation test. Concerning its results, now, on only several occasions are *p*-values of its statistic found below 0.05 (for example, when ADCC-Gaussian is used to fit stock and bond data, AGDCC-*t* is used to fit exchange rate and ADCC-*skew-t* is used to estimate simulated ADCC-MTM data), suggesting that the null of a constant correlation is rejected and the dynamics in correlation evolving process are not sufficiently captured and can be further exploited using the current model. For all others (27 out of 30 samples) dominating evidences then confirm the filtering power of asymmetric DCC models on conditional returns, volatility and correlation.

Moreover, here there is also another important result worth noting. That is, before returns are standardized, only stock and bond data show clear evidence of dynamic correlation and these dynamics can be steadily captured after DCC model is put, step-by-step, onto a more sophisticated level, while others accept the null, suggesting a stable correlation evolving process. This result has important implications because it finally explains the puzzle of 'why a dynamic correlation model does not always produce dynamic correlation', which is in several occasions presented in the previous sections. Given this result, the non-dynamic (or sometimes constant) correlation evolving process generated for exchange rate data, simulated ADCC-MGM data and simulated ADCC-MTM data are then no longer unexpected.

Now, since this finding clearly contradicts our initial conjecture (correlation is dynamic), we highlight it in broader terms and reveal the necessity of revisiting the basic motivation of proposing a dynamic process for modelling correlation. Indeed, massive empirical evidences have documented that volatility is time-varying and will change dynamically, and that correlation also follows a similar process. However, the extent to which correlation might change as dynamically as volatility is still unknown. Although correlation can be manually modelled as if it follows a dynamic process (just like the cases discussed in our sample), empirical observations sometimes still support the evidence of a constant correlation especially when the sample size is not large enough to include any significant events affecting both assets. Such events for correlation are especially important because they can lead to the identification of a potential structural change (recall that, in credit portfolio, a small increase of correlation can result in a significant tilt in distribution). Thus, for the time being, unless such events are observed, correlation in a relative term is often considered to be following a stable process. To this end, the presence of constant correlation in our results then can be partly explained by a pure coincidence or just a result of relatively small-sized sample data because, if we now re-examine, say, the correlation between S&P500 and FTSE100 using 12 years' data (previous: T=1000; now: T≈3000), evidence then clearly supports a strong dynamic process.[104] The time invariant correlation is no longer the case in this larger timescale. Besides, given this finding, it is also natural to expect that using a jump diffusion process to fit correlation could potentially yield more desirable results because persistence and structural changes in this process can be simultaneously accounted.

7.4 Summary

In this chapter we have illustrated the estimation result of two mixture models and examined their performance from a range of perspectives including asset allocation and risk management. In simulation studies, we found that, for most parameters, empirical moments calculated from posterior draws are good estimators to approximate their corresponding true values. ADCC-MTM outperforms ADCC-MGM in terms of being able to generate a lower portfolio variance in optimization and a more sensible VaR result. However, in several cases, undesirable results are also documented. That is, in simulating some particular volatility parameters and correlation parameters, their generated posterior moments fail to approach corresponding theoretical values closely enough although statistical convergence of their resultant chains can be confirmed. Concerning the empirical investigations, in this research we analyze three different correlation scenarios (positive, negative and zero) and portfolios with assets of different classes and assets in different markets. After simulation, we found that, for foreign exchange data, appeals of modelling unconditional return using two-component mixture is not very significant because weight parameter and degree of freedom parameter which respectively govern the proportion of the mixture and tails behaviour of each component is roughly the same. However, for stock and bond data, a good fitting result is reported. Concerning the asymmetric factor, only when stock index data are fitted have time-varying correlations

[104] Indeed, we hope to choose a very large dataset for empirical analysis. However, it cannot be denied a balance always has to be made between computational costs and efficiency in estimation. As has been highlighted already, since a major deficiency of our proposed mixture model is their extremely high computational cost, it is then preferable to choose a relatively small sample size for analysis. However, to ensure the asymptotical normality in estimating GARCH using QML, this sample size in the meantime cannot be allowed too small. Thus, in every case we let our sample include at least 1000 observations.

shown a different response to negative news and positive news. Besides, another major topic in this chapter is to compare the model performance of mixture models with a variety of alternative DCCs. Here, it is especially worth noting the ADCC-*skew t* and AGDCC-*skew t* proposed and estimated in our paper. This is because these models are so generalized that can nest a range of standard conditional correlation models and also for the first time analyzed in financial literature. Concerning their results, strong evidences are found that, except for mixture models, they are the best among all alternatives on account of the flexibility and economic benefits (being able to generate the lowest portfolio variance among DCC variants). Finally, in this research we also prove that 'whether the correlation is a dynamic process' is actually an empirical problem, depending on the sample to be analyzed.

Chapter 8

Conclusion

8.1 Summary of findings and discussion

In this research, we focus on modelling the time-varying correlation using a variety of techniques. The whole thesis, in terms of the empirical analyses performed, can be divided into two parts.

In the first part, after reviewing a large amount of literature on covariance modeling, we use a variety of existing time series tools including historical correlation models, EWMA and GARCH variants to forecast the conditional correlation in two currency trios over the next week, next month and next quarter. Then, these forecasts are compared to implied correlations generated by using option prices as information processor. Here, for calculating implied correlation, contrary to most early researchers, who used implied volatility collected from a particular market participant, we utilize an index provided by British Bankers Association (BBA). The benefits of choosing this data are massive. Most importantly, it is because different opinions on how future volatility will move can be synthesized and we can now, through this index, obtain a thorough market view. After empirical analyses, we find the 'best' model to forecast future realized correlations is actually very sensitive to the loss functions used to evaluate them. Although implied correlation is able to consistently convey valuable information to accurately forecast the 'true' correlation in different trios, its cross-horizon performance is not uniform. Among the time series tools, simple forecasts calculated from the historical correlation model and EWMA can frequently produce an unbiased estimator for approximating realized correlation. However, in terms of the forecasting accuracy, these models are overwhelmingly outperformed by other competitors. Besides, our findings suggest that using GARCH models can generate information not obtainable from option prices. However, its advantage of capturing the time-varying characteristics of correlation is not fully exploited. For instance, in our sample a subtle tendency is to favour the flat-term-structure model, i.e., CCC of Bollerslev (1990). From the encompassing test results, we find the combination of historical information source and option-derived information source can produce a more accurate correlation forecast than using any single technique.

Furthermore, the explanatory power of the regression, after applying this strategy, can substantially arise.

Besides, in the same analysis, some interesting results are also worth mentioning here. For example, we find that long-term correlation can be more accurately forecasted than short-term correlation. This result is as expected because the former process usually tends to show more stable distributional characteristics than the latter one. Meanwhile, in several occasions, unconditional distribution of realized correlation is found presenting multi-modality, suggesting that market views may have diverged on ‘how future correlations will move’. Given this feature, it is then implied that, by adopting a mixture technique, correlation probably can be more accurately estimated and forecast. To test this hypothesis, we devote the second part of this thesis specifically to developing two new conditional heteroskedastic correlation mixture models.

As before, firstly we review a variety of mixture modelling techniques and their associated inferential methods (both classical and Bayesian) so that the questions of ‘how to construct our target models and how to estimate them’ can be answered. Then, after assuming the innovations of multiple returns to be respectively multivariate Gaussian mixture-distributed and multivariate T mixture-distributed and the correlation evolving process modelled using ADCC of Hafner and Franses (2003), specifications of ADCC-MGM and ADCC-MTM are then given. For estimating these models, we use Griddy-Gibbs sampler of MCMC to calculate their Bayesian inferences. And their model potentials are examined in two simulation studies as well as through three empirical investigations.

After posterior simulation, we find inferential results generated for simulation studies are generally good but not quite uniform. For most parameters, their resultant chains are found to be converged and can produce useful distributional information. Posterior means (or modes in the case of an asymmetric posterior density) of most chains are very close to their corresponding theoretical values set in either DGP1 or DGP2. However, in some cases, non-convergence is also documented, with a large gap observed between calculated posterior mean and true values. Concerning the empirical investigations, the usefulness of

ADCC mixture models for estimating correlation dynamics is demonstrated in four out of six cases. Only when exchange rate data are fitted, using one-component ADCC is found economically more beneficial. This is because posterior results of these data estimated on a two-component model now support equal mixing, and tail behaviours of two Gaussian/T components assumed in the mixture are also found roughly the same. Thus, for this particular case, by substituting the proposed models with either ADCC-Gaussian or ADCC-T, we can obtain the same quality of inferential results while saving a substantial amount of computational costs.

Besides, in this research we also confirm the superiority of our correlation models over a variety of alternatives such as CCC, DCC, ADCC, AGDCC and their variants. From a range of perspectives (both statistical and economical), we compare these models' performances in forecasting future correlation, generating VaR estimates and minimizing portfolio variance. Among competing models, here it is especially worth noting ADCC-*skew t* and AGDCC-*skew t,* proposed and estimated in this paper, because these models are so generalized that, except for ADCC-MGM and ADCC-MTM, they can nest all other conditional correlation models mentioned above. As a response to their parsimonious specification and great flexibility, they are also found, in the majority of cases, to outperform their competitors.[105] Only when mixture models are included, they become the second best. Now, as far as the generality and economic benefits of a model are concerned, unquestionably, in this research it is still the mixture ones that perform the best. Strong evidences have been found to confirm their superiority, on a consistent basis, over all other alternatives. And ADCC-MTM can outperform ADCC-MGM in terms of being able to generate a comparatively even lower portfolio variance in optimization and a more sensible VaR result on account of the extreme events.

Now, leaving aside temporarily the aforementioned posterior results, it is important to mention that all correlation dynamics modelled, calculated and forecast above are theoretically valid only when certain hypotheses are realized. Recall from Chapter two that

[105] For ADCC-*skew t* and AGDCC *skew t*, high moments of multivariate distributions can be accounted using only a moderate number of parameters.

these assumptions are respectively, the existence of a realistic causality between assets being modelled, univariate and jointly multivariate normality for their distributions and financial variables supposed to be only linearly associated with each other. The first and third conditions can be easily satisfied if a proper interpretation of the result is given. However, concerning the second, univariate normality and multivariate normality for most data used in this research are then firmly rejected. Given this feature, arguably, the validity of our results is then open to challenge. However, needless-to-say, in financial literature invalid results due to the violation of normality is nothing unusual. As has been confirmed by countless researchers, returns, even after being fitted by a heteroskedastic model which can capture the volatility clustering and fat tails and standardized by its calculated volatility, would still not, in most cases, show Gaussian characteristics. Since non-normality is a matter of common sense for financial data, this challenge to our sample is not massive, though undoubtedly it exists.

8.2 Contributions and Implications

Concerning our contributions to the current literature, they are threefold. First, on the theoretical side, we extend the existing framework for (covaraince) correlation modelling by incorporating advanced distributional techniques so that excess skewness and fat tails can be more flexibly accounted. Specifically, in the parametric framework we use a skewed version of symmetric *t*, and in the semi-parametric framework apply a mixture modelling techniques. Here, it is especially worth noting the correlation mixture models. A major advantage of these models, not shared by others, is their capability to allow for multi-modality. If, say, multiple opinions on ‘how future market will move’ are now formed among different investment groups and these opinions are sufficiently different from each other, we can then use mixture models to reveal the heterogeneity of investors and extract their expectations. Applied to a correlation model, since the changes in market behavior (due to the involvement of different investors, either geographically or psychologically) will eventually be transmitted to parameters of covariance equation and reflected through correlation dynamics, we can then, by plotting the kernel density of calculated correlation, to obtain an indicator for any potential divergence on market expectation. Besides, on the theoretical side, it is also worth noting that not only volatility and correlation, skewness of asset returns, given a correlation mixture model, is now also allowed to be time-varying. Since dynamic feedback between different components is permitted, there is no need to impose a specific evolving process for this conditional moment. The time-varying property for skewness is inherently given by the mixture model.[106] Second, on the computational side, in this research we demonstrate Griddy Gibbs sampler is a valid and easy-to-implement MCMC technique for estimating parameters of mixture models, although its associated computational cost is massive. Finally, on the empirical side, we confirm that, compared to a variety of alternatives, both ADCC-MGM and ADCC-MTM are better time series tools for forecasting future correlation, generating optimal portfolio and deriving sensible VaR results. Besides, since parameter uncertainty is allowed, we can also use them to obtain distributional information of, say, the next day's returns.

With respect to the implications of this research, two things need to be noted. First, our initial conjecture of modelling correlation as a dynamic process has been proved as an empirical issue. Various evidence found in this research supported a constant correlation between financial returns. However, when the sample size is enlarged, test results then favour the dynamic correlation again. Given this feature, 'to what extent correlation is a dynamic process' then become a sample-specific question and it needs to be put into a broader framework for analysis. Second, as for the asymmetric correlation, only when stock index data are fitted, we have confirmed its existence. Concerning all others, conditional returns then tend to give similar responses to both positive news and negative news. This result is not surprising because similar findings have already been documented by other researchers. For example, Baur (2003) found little evidence for correlation increasing with jointly negative shocks. This result has important implications for portfolio selection. Since it is usually expected correlation of various assets will rise when the overall market is going down and portfolio diversification in this scenario usually loses its

[106] Concerning this property, it can be numerically proved by writing conditional skewness as a function of conditional mean and conditional variance.

appeals when it is needed the most, theoretically, if this is proved not to be the case one can then utilize the traditional approach to buy assets having negative correlation with the one which is currently held, to hedge the overall market risk.

Besides, for portfolio manager, using our correlation mixture model can also bring other benefits. On one hand, since it has been confirmed that in this research the new proposed model can provide an economically better performance than the traditional DCC in terms of being able to generate the lowest portfolio variance, given that the portfolio returns of all competing models are set to be equal, it is then fair to say the ADCC mixture model is actually a very suitable tool for calculating the optimal weights of each asset to invest. On the other hand, since the inference of mixture model is now calculated using Bayesian method through the implementation of a stochastic simulation technique. Parameter uncertainty is obtainable after the inference calculation. That is to say one can now know more about the parameter risk concerning the model when it is applied to the real financial data. This information is valuable because it cannot be easily obtained through any other classical inferential approaches such as maximum likelihood or EM algorithm and can help a portfolio manager to obtain a more objective view on his model's performance. Least but not last, as have been mentioned in the start of this section, using ADCC-mixture model can also help a portfolio manager to gauge the market sentiment more accurately and make right decisions. Through modelling return distribution using a mixed way, any modes appearing in this distribution then can be regarded as representing the views of a group of investor on how the future market will move. Therefore, if the overall market sentiment does diverge, one can then easily detect this trend through the multi-modality.

8.3 Limitations and suggestions for future study

Apart from the positive contributions, undesirable and unexpected results were also generated in this research. For example, in the first empirical analysis, cross-horizon forecasting performance of implied correlation was found to be very confusing. In the Yen trio, these correlations are reported, in only one case, overwhelmingly dominating all other time series forecasts across two horizons analyzed. However, in the GBP trio no such evidence is reported again and indeterminacy results are documented in the majority of cases. This result clearly contradicts to those documented in cross-trio forecast comparison. According to the evidences provided in partial optimal test (See Table 3.3) and encompassing regression test (See Table 3.5), implied correlations are found, in both trios, to be able to outperform other forecasts on a consistent basis, and the usefulness of option-driven information is confirmed in most cases. Here, although we can attribute the non-uniform cross-horizon performance of implied correlation to the different dynamic processes being modelled (one short-term correlation and one long-term correlation), a more plausible explanation needs to be found in further research.

Concerning the second empirical analysis, disadvantages of using correlation mixture models are evident. Since ADCC-MGM and ADCC-MTM have both assumed a very complex specification and estimation of their specification needs extremely high numerical efforts, the empirical potential of these mixture models are then quite limited. As a response to the stringent demands of practical asset allocation and practical risk management on parsimony, they then clearly cannot be applied to solve any system of medium or large size (a portfolio with many assets).

However, this does not mean that our mixture models cannot be improved. For example, we can, by exploiting a result from Anderson *et al.* (2003) and Andersen *et al.* (2005), enhance our models' parsimony. Since, in their research, it was proved that true volatility of low frequency can be closely approximated using realized volatility of higher frequency, based on the DCC modelling virtue we can then use these volatilities to standardized conditional returns so that resultant innovations can be input to a heteroskedastic model to re-estimate correlation. In so doing, the GARCH specifications, along with their parameters previously assumed in ADCC-MGM and ADCC-MTM for fitting univariate volatility can then be eliminated. Besides, in Chapter 7 we have already reported that, in a number of cases, mean parameters are also frequently found to be close to zero. Thus, by eliminating these parameters altogether, numerical difficulties of estimating mixture models are then expected to be substantially relieved.

In addition, a new direction of research can also be proposed if mixture model is now extended using *skew-t* or asymmetric Laplace as components. If our task here is only to increase the generality, one may argue that a potential solution is to add another component to ADCC-MGM or ADCC-MTM, and there is no need to propose a new mixture. Indeed, given a series of highly asymmetric observations showing multiple modes, one is very easily tempted to use a multi-component (larger than two) mixture for modelling financial data. However, Wang, Ng and Mclachlan (2009), in a recent study, pointed out that "*... increased number of pseudo-components could lead to difficulties and inefficiencies in computations. Also, the contour of the fitted mixture components may be distorted....*" Thus, it is preferable to keep the number of components relatively low (or unchanged) while using a more flexible density as base distribution to construct the mixture. Based on this virtue, *skew-t,* which can allow for both skewness and fat tails in a multivariate distribution, is then an ideal choice. To author's knowledge, little work has been done in this direction. Here, only two papers are worth mentioning. One is a recent study by Lin (2009), who developed the maximum likelihood estimators for multivariate skew normal mixture. The other is by Wang, Ng and McLachlan (2009), who proposed the multivariate *skew t* mixture using EM as an inferential method.

Table 7.1 Summary Statistics and hypothesis test results of data simulated using DGP1 and DGP2

Panel A. Simulated bivariate Two-Component MGM distributed innovations with ADCC(1,1) covariance incorporated. (Sample size: 2000)

Unconditional correlation: 0.8023

	DGP1: (MGM)	
Mean	0.0003	0.0014
Median	0.0013	0.0009
Maximum	0.2856	1.0270
Minimum	-0.2861	-0.8712
StandardDeviation	0.0900	0.2611
Skewness	-0.0213	0.0447
Kurtosis	2.7542	3.9294
Uni-Normality (*p-val*)	0.4121	0.0005
Multi-Normality (*p-val*)	0.1143	

This table presents seven descriptive statistics and results of two hypothesis tests for data simulated using first Data generating process (DGP1) which corresponds to ADCC (1, 1) model with two-Component Gaussian Mixture distributed disturbances

Panel B. Simulated bivariate Two-Component MTM distributed innovations with ADCC(1,1) covariance incorporated. (Sample size: 2000)

Unconditional correlation: 0.7943

	DGP2: (MTM)	
Mean	-0.0052	-0.0079
Median	-0.0049	-0.0059
Maximum	0.8121	1.2022
Minimum	-0.4848	-1.5683
StandardDeviation	0.1086	0.3053
Skewness	0.1270	-0.0334
Kurtosis	5.6068	5.3456
Uni-Normality (*p-val*)	0.0005	0.0000
Multi-Normality (*p-val*)	0.0000	

This table presents seven descriptive statistics and results of two hypothesis tests for data simulated using second Data generating process (DGP2) which corresponds to ADCC (1, 1) model with Two-Component multivariate T Mixture distributed disturbances

Table 7.2 Posterior estimation result of simulation studies (sample size: 2000; number of iterations: 10000 (Burn-in) and 5000 (In equilibrium))

Panel A. ADCC-MGM model estimated on simulated data based on DGP1

	Mean	Median	Mode	S.t.d	Max	Min
π_1	0.6864	0.6855	0.6910	0.0329	0.7918	0.5678
μ_{a1}	0.0009	0.0010	0.0010	0.0026	0.0081	-0.0077
μ_{b1}	0.0052	0.0055	0.0086	0.0079	0.0247	-0.0214
ω_{a1}	0.0056	0.0057	0.0060	0.0012	0.0081	0.0022
ω_{b1}	0.0494	0.0508	0.0529	0.0119	0.0681	0.0071
ω_{a2}	0.0042	0.0042	0.0044	0.0014	0.0077	0.0003
ω_{b2}	0.0035	0.0034	0.0035	0.0018	0.0091	0.0005
α_{a1}	0.0266	0.0231	0.0066	0.0202	0.1325	0.0000
α_{b1}	0.0323	0.0271	0.0064	0.0237	0.1278	0.0000
α_{a2}	0.0774	0.0619	0.0204	0.0631	0.4086	0.0000
α_{b2}	0.0409	0.0333	0.0112	0.0318	0.2236	0.0000
β_{a1}	0.2573	0.2413	0.2448	0.1553	0.6988	0.0003
β_{b1}	0.3959	0.3816	0.3433	0.1452	0.8956	0.0459
β_{a2}	0.2997	0.2853	0.3231	0.1936	0.9230	0.0001
β_{b2}	0.1838	0.1443	0.0428	0.1661	0.8558	0.0000
η_1	0.0780	0.0650	0.0165	0.0590	0.3273	0.0001
η_2	0.1852	0.1601	0.0306	0.1366	0.6095	0.0001
ζ_1	0.7299	0.7884	0.9310	0.2088	0.9799	0.0007
ζ_2	0.5782	0.6317	0.7344	0.2444	0.9792	0.0002
ι_1	0.3021	0.3070	0.2972	0.1111	0.6586	0.0014
ι_2	0.2633	0.2403	0.0451	0.1795	0.9011	0.0001

Panel B. ADCC-MTM model estimated on simulated data based on DGP2

	Mean	Median	Mode	S.t.d	Max	Min
π_1	0.6964	0.6964	0.6894	0.0345	0.7897	0.5669
μ_{a1}	-0.0050	-0.0050	-0.0058	0.0031	0.0039	-0.0137
μ_{b1}	-0.0090	-0.0091	-0.0079	0.0091	0.0145	-0.0353
ω_{a1}	0.0043	0.0043	0.0044	0.0019	0.0096	0.0001
ω_{b1}	0.0401	0.0367	0.0258	0.0211	0.0925	0.0035
ω_{a2}	0.0050	0.0048	0.0044	0.0013	0.0102	0.0013
ω_{b2}	0.0045	0.0042	0.0036	0.0025	0.0124	0.0006
α_{a1}	0.0411	0.0379	0.0081	0.0276	0.1606	0.0001
α_{b1}	0.0374	0.0315	0.0093	0.0270	0.1857	0.0000
α_{a2}	0.1114	0.0912	0.0289	0.0889	0.5732	0.0002
α_{b2}	0.0368	0.0317	0.0092	0.0271	0.1830	0.0000
β_{a1}	0.3842	0.3537	0.3427	0.2369	0.9789	0.0001
β_{b1}	0.4857	0.5180	0.7128	0.2469	0.9497	0.0018
β_{a2}	0.2192	0.2094	0.0361	0.1615	0.7082	0.0007
β_{b2}	0.1559	0.1035	0.0413	0.1593	0.8240	0.0001
η_1	0.1903	0.1876	0.1607	0.1065	0.6428	0.0001
η_2	0.1136	0.0829	0.0347	0.1041	0.6915	0.0001
ζ_1	0.3253	0.2907	0.0486	0.2274	0.9652	0.0004
ζ_2	0.5116	0.5279	0.8331	0.2962	0.9800	0.0005
ι_1	0.2989	0.3038	0.3384	0.1456	0.7507	0.0011
ι_2	0.2004	0.1758	0.0401	0.1513	0.7993	0.0002
ν_1	9.7546	9.0451	5.8775	6.4977	97.4777	1.0565
ν_2	8.3224	6.8628	5.8985	8.7192	98.5901	1.0200

* ‘a’, ’b’ denotes the first and second series of bivariate sample data; ‘1’, ‘2’ represents the first and second mixture component included in either MGM or MTM.

Table 7.3 Convergence diagnostic results of simulation studies (sample size: 2000; number of iterations: 10000 (Burn-in) and 5000 (In equilibrium))

	DGP1			**DGP2**		
	Z-test	PSRF	IPSRF	*Z*-test	PSRF	IPSRF
π_1	0.9231	0.9997	1.0028	0.2910	0.9998	0.9941
μ_{a1}	0.7580	0.9997	0.9986	0.6442	0.9998	1.0000
μ_{b1}	0.5551	0.9997	0.9982	0.6195	0.9997	0.9944
ω_{a1}	0.0499	1.0001	0.9993	0.1791	1.0036	0.9965
ω_{b1}	0.1746	1.0001	1.0000	0.0520	1.0009	1.0003
ω_{a2}	0.7045	1.0001	0.9999	0.7983	1.0008	1.0036
ω_{b2}	0.2962	1.0000	0.9980	0.9637	0.9998	0.9999
α_{a1}	0.8685	0.9997	0.9992	0.5578	1.0000	0.9966
α_{b1}	0.8804	0.9997	1.0006	0.4620	0.9997	1.0021
α_{a2}	0.3406	0.9999	1.0010	0.4349	0.9998	1.0000
α_{b2}	0.0566	1.0007	1.0018	0.7863	0.9997	0.9948
β_{a1}	0.0430	1.0001	1.0009	0.1387	1.0036	0.9906
β_{b1}	0.1661	1.0001	0.9968	0.0264	1.0013	1.0015
β_{a2}	0.8311	1.0005	1.0004	0.9439	1.0014	0.9989
β_{b2}	0.7689	0.9998	0.9991	0.8831	1.0003	1.0075
η_1	0.6213	0.9998	1.0005	0.9294	0.9997	1.0000
η_2	0.5774	0.9997	0.9994	0.6906	0.9997	1.0028
ζ_1	0.6468	0.9997	0.9976	0.6132	0.9997	1.0000
ζ_2	0.7423	0.9997	0.9998	0.8621	0.9998	1.0008
ι_1	0.7382	0.9998	1.0015	0.7019	0.9997	1.0067
ι_2	0.8348	0.9997	1.0002	0.9421	0.9999	0.9994
ν_1				0.3542	0.9998	1.0011
ν_2				0.8201	0.9998	1.0034

This table reports the convergence diagnostic results of Markov chains simulated from the first and second simulation study. Specifically, Geweke (1992)'s partial mean test (or called *Z*-test), Gelman and Rubin (1992)'s PSRF test and Brooks and Gelman(1997)'s IPSRF test are carried out here. For Z-test, we report *p*-value of test statistic and set the significance level to be 95%. Therefore, any values lower than 0.05 is interpreted as casting doubts on the null 'Markov chain has converged'. This test is performed here to test whether the posterior means of first N_a draws and last N_b draws of Markov chain are the same. In this research, N_a and N_b are respectively set as first 1500 and final 1500 of the total 5000 equilibrium draws in Markov chains. For PSRF and IPSRF, Gelman and Rubin (1992) argued a value close to one is enough to claim convergence.

Table 7.4 Posterior correlation matrix of the simulated parameter values of ADCC-MTM

	π_1	μ_{a1}	μ_{b1}	ω_{a1}	ω_{b1}	ω_{a2}	ω_{b2}	α_{a1}	α_{b1}	α_{a2}	α_{b2}	β_{a1}	β_{b1}	β_{a2}	β_{b2}	η_1	η_2	ζ_1	ζ_2	ι_1	ι_2	ν_1
μ_{a1}	0.032																					
μ_{b1}	0.022	0.785																				
ω_{a1}	0.073	0.054	0.047																			
ω_{b1}	0.017	0.025	0.016	0.420																		
ω_{a2}	0.056	-0.042	-0.029	-0.021	0.062																	
ω_{b2}	-0.029	-0.066	-0.035	0.001	0.034	0.166																
α_{a1}	-0.008	-0.031	-0.009	-0.076	-0.072	-0.008	0.006															
α_{b1}	-0.074	0.013	0.018	-0.155	-0.008	-0.022	0.031	0.395														
α_{a2}	0.146	-0.015	-0.006	0.065	0.056	0.021	0.352	-0.036	-0.043													
α_{b2}	-0.074	0.038	0.020	-0.032	-0.059	-0.041	-0.324	-0.005	-0.024	-0.155												
β_{a1}	-0.140	-0.043	-0.035	-0.950	-0.388	-0.013	-0.032	-0.046	0.128	-0.065	0.034											
β_{b1}	-0.094	-0.027	-0.015	-0.391	-0.964	-0.085	-0.046	0.047	-0.062	-0.069	0.050	0.413										
β_{a2}	0.009	0.009	0.009	0.017	-0.047	-0.596	0.345	0.021	0.038	-0.035	0.083	-0.034	0.040									
β_{b2}	-0.015	0.029	0.013	0.005	-0.032	0.009	-0.718	-0.010	-0.014	-0.250	-0.043	0.017	0.046	-0.355								
η_1	0.019	-0.011	0.012	-0.031	0.031	0.027	0.101	0.162	0.379	0.037	-0.076	0.011	-0.058	0.037	-0.084							
η_2	0.080	0.014	0.020	0.038	0.035	-0.050	-0.031	0.012	0.006	0.128	0.083	-0.036	-0.038	0.054	-0.044	0.011						
ζ_1	0.001	0.008	0.001	-0.096	-0.061	-0.065	-0.073	0.108	-0.055	-0.007	0.020	0.120	0.092	-0.004	0.037	-0.110	0.017					
ζ_2	0.034	0.020	0.010	0.015	0.004	0.078	-0.037	-0.005	-0.008	-0.075	0.184	-0.024	-0.011	0.019	-0.026	0.027	0.041	-0.027				
ι_1	-0.134	0.041	0.043	-0.071	0.031	0.052	0.106	0.174	0.298	-0.032	-0.010	0.090	-0.018	0.031	-0.024	-0.139	-0.020	-0.186	0.016			
ι_2	0.045	-0.030	-0.018	0.040	0.032	0.164	0.289	0.003	-0.017	0.068	0.070	-0.057	-0.041	0.181	-0.230	0.048	0.047	0.002	-0.017	-0.064		
ν_1	-0.159	-0.007	0.015	-0.019	-0.029	-0.077	-0.022	-0.011	0.038	-0.024	-0.010	0.149	0.138	-0.024	-0.003	0.051	0.020	0.075	-0.048	0.097	-0.047	
ν_2	0.077	-0.066	-0.055	-0.013	-0.016	0.225	0.397	-0.031	-0.021	0.218	0.006	-0.031	-0.025	0.145	-0.227	0.067	0.047	-0.066	0.097	0.135	0.210	-0.101

* Big triangle denotes the correlation matrix of posterior values drawn for volatility (ω, α, β) parameters. Small triangle denotes the correlation matrix of posterior values drawn for correlation parameters (η, ζ, ι).

Table 7.5 Summary Statistics and hypothesis test results of empirical data

Panel A: Foreign exchange data (US/UK and EU/JP)

Sample size: 1689 Unconditional correlation: -0.3182

	US/UK	EU/JP
Mean	0.0000	0.0000
Median	-0.0001	0.0003
Maximum	0.0200	0.0448
Minimum	-0.0251	-0.0304
StandardDeviation	0.0050	0.0073
Skewness	-0.0551	0.0135
Kurtosis	3.7677	5.0402
Uni-Normality (*p-val*)	0.0092	0.0081
Multi-Normality (*p-val*)	0.0000	

Panel B: S&P500 and 10y US Bond

Sample size: 3000 Unconditional correlation: -0.0863

	S&P500	US Bond
Mean	0.0004	0.0001
Median	0.0002	0.0000
Maximum	0.0573	0.0143
Minimum	-0.0687	-0.0282
StandardDeviation	0.0108	0.0038
Skewness	-0.0171	-0.5718
Kurtosis	6.5493	6.2270
Uni-Normality (*p-val*)	0.0425	0.0009
Multi-Normality (*p-val*)	0.0000	

Panel C: S&P500 and FTSE100

Sample size: 1000 Unconditional correlation: 0.4259

	S&P500	FTSE100
Mean	0.0005	0.0004
Median	0.0004	0.0006
Maximum	0.0227	0.0222
Minimum	-0.0293	-0.0402
StandardDeviation	0.0067	0.0067
Skewness	-0.3360	-0.3370
Kurtosis	4.4311	4.6782
Uni-Normality (*p-val*)	0.0000	0.0000
Multi-Normality (*p-val*)	0.0000	

Table 7.6 Posterior estimation result of exchange rate data (sample size: 1689; number of iterations: 10000 (Burn-in) and 5000 (In equilibrium))

Panel A. ADCC-MGM model estimated on exchange rate data

	Mean	Median	Mode	S.t.d	Max	Min
π_1	0.66763	0.63396	0.56639	0.12723	0.98930	0.42542
μ_{a1}	-0.00017	-0.00018	-0.00019	0.00018	0.00045	-0.00053
μ_{b1}	0.00011	0.00014	0.00023	0.00030	0.00073	-0.00071
ω_{a1}	0.00000	0.00000	0.00000	0.00000	0.00003	0.00000
ω_{b1}	0.00001	0.00001	0.00000	0.00001	0.00005	0.00000
ω_{a2}	0.00001	0.00001	0.00001	0.00001	0.00003	0.00000
ω_{b2}	0.00002	0.00001	0.00000	0.00002	0.00005	0.00000
α_{a1}	0.09610	0.09145	0.09111	0.05715	0.36439	0.00002
α_{b1}	0.14116	0.13791	0.11527	0.07480	0.46065	0.00015
α_{a2}	0.10163	0.06466	0.04576	0.10780	0.91424	0.00005
α_{b2}	0.16393	0.12609	0.03829	0.13050	0.76556	0.00001
β_{a1}	0.75720	0.78348	0.76911	0.13582	0.99823	0.08173
β_{b1}	0.73353	0.74894	0.79258	0.11579	0.99662	0.18047
β_{a2}	0.48547	0.49655	0.73559	0.25019	0.98079	0.00002
β_{b2}	0.55495	0.57853	0.63718	0.18793	0.97998	0.00056
η_1	0.11129	0.10245	0.01979	0.07573	0.39517	0.00003
η_2	0.17305	0.15557	0.03653	0.11895	0.72746	0.00017
ζ_1	0.49141	0.47069	0.93089	0.29997	0.97987	0.00022
ζ_2	0.45871	0.43057	0.04915	0.28691	0.97947	0.00018
ι_1	0.12571	0.10460	0.04605	0.09859	0.92061	0.00002
ι_2	0.19058	0.15550	0.04712	0.15448	0.94237	0.00000

Panel B. ADCC-MTM model estimated on exchange rate data

	Mean	Median	Mode	S.t.d	Max	Min
π_1	0.69473	0.67187	0.54227	0.13598	0.98985	0.46328
μ_{a1}	-0.00014	-0.00015	-0.00019	0.00019	0.00045	-0.00053
μ_{b1}	0.00016	0.00018	0.00023	0.00030	0.00073	-0.00071
ω_{a1}	0.00001	0.00000	0.00000	0.00001	0.00003	0.00000
ω_{b1}	0.00001	0.00001	0.00000	0.00001	0.00005	0.00000
ω_{a2}	0.00001	0.00000	0.00000	0.00001	0.00003	0.00000
ω_{b2}	0.00001	0.00001	0.00000	0.00001	0.00005	0.00000
α_{a1}	0.06847	0.06215	0.04398	0.04519	0.29307	0.00002
α_{b1}	0.11398	0.10266	0.06494	0.07439	0.43260	0.00005
α_{a2}	0.07780	0.05813	0.04481	0.07727	0.89616	0.00000
α_{b2}	0.16348	0.12603	0.04582	0.14206	0.91608	0.00002
β_{a1}	0.66496	0.73800	0.84452	0.22471	0.99334	0.00121
β_{b1}	0.68275	0.72461	0.74095	0.18876	0.98778	0.00048
β_{a2}	0.46525	0.44623	0.84789	0.26989	0.99744	0.00044
β_{b2}	0.50143	0.54626	0.64042	0.23888	0.98481	0.00083
η_1	0.09092	0.07996	0.01964	0.06581	0.39168	0.00006
η_2	0.17171	0.13469	0.04304	0.14381	0.86041	0.00002
ζ_1	0.48540	0.45644	0.93100	0.30530	0.97999	0.00022
ζ_2	0.45786	0.43512	0.14697	0.29056	0.97958	0.00004
ι_1	0.11268	0.09613	0.03008	0.08512	0.59922	0.00013
ι_2	0.18292	0.14296	0.04866	0.15824	0.97281	0.00002
ν_1	21.46468	11.83584	5.95098	22.88304	99.99257	1.00143
ν_2	36.27728	25.13340	5.95147	31.02248	99.99957	1.00157

* ‘a’, ’b’ denotes the first and second series of bivariate sample data; ‘1’, ‘2’ represents the first and second mixture component included in either MGM or MTM.

Table 7.7 Posterior estimation result of stock and bond data (sample size: 3000; number of iterations: 10000 (Burn-in) and 5000 (In equilibrium))

Panel A. ADCC-MGM model estimated on stock and bond data

	Mean	Median	Mode	S.t.d	Max	Min
π_1	0.80700	0.81881	0.81775	0.08378	0.98987	0.49810
μ_{a1}	0.00092	0.00093	0.00099	0.00016	0.00119	-0.00016
μ_{b1}	0.00017	0.00018	0.00019	0.00007	0.00034	-0.00010
ω_{a1}	0.00001	0.00001	0.00000	0.00000	0.00003	0.00000
ω_{b1}	0.00000	0.00000	0.00000	0.00000	0.00001	0.00000
ω_{a2}	0.00004	0.00004	0.00003	0.00002	0.00012	0.00000
ω_{b2}	0.00001	0.00001	0.00001	0.00000	0.00002	0.00000
α_{a1}	0.12798	0.12441	0.12398	0.06091	0.35387	0.00019
α_{b1}	0.07849	0.07480	0.06931	0.04540	0.27722	0.00001
α_{a2}	0.24493	0.24403	0.28331	0.09558	0.62930	0.00023
α_{b2}	0.19038	0.19071	0.19137	0.08824	0.54656	0.00011
β_{a1}	0.77376	0.77717	0.77866	0.08082	0.99508	0.37672
β_{b1}	0.71608	0.73939	0.74603	0.13764	0.99432	0.00115
β_{a2}	0.70953	0.71022	0.70173	0.10086	0.99488	0.34343
β_{b2}	0.73905	0.73805	0.73848	0.09042	0.99601	0.42373
η_1	0.38367	0.38426	0.39028	0.04928	0.54562	0.20042
η_2	0.28313	0.28350	0.26854	0.07514	0.59409	0.00218
ζ_1	0.87912	0.88288	0.88259	0.03318	0.96369	0.73198
ζ_2	0.86659	0.91791	0.93091	0.16324	0.97986	0.00094
ι_1	0.09734	0.06572	0.02192	0.08882	0.43758	0.00005
ι_2	0.10539	0.07975	0.03920	0.09268	0.78362	0.00002

Panel B. ADCC-MTM model estimated on stock and bond data

	Mean	Median	Mode	S.t.d	Max	Min
π_1	0.81267	0.84556	0.85859	0.11949	0.98998	0.46440
μ_{a1}	0.00078	0.00080	0.00084	0.00021	0.00119	-0.00021
μ_{b1}	0.00017	0.00017	0.00020	0.00009	0.00034	-0.00021
ω_{a1}	0.00001	0.00001	0.00000	0.00000	0.00004	0.00000
ω_{b1}	0.00000	0.00000	0.00000	0.00000	0.00001	0.00000
ω_{a2}	0.00001	0.00001	0.00001	0.00001	0.00010	0.00000
ω_{b2}	0.00000	0.00000	0.00000	0.00000	0.00002	0.00000
α_{a1}	0.12770	0.12617	0.12810	0.05751	0.36595	0.00003
α_{b1}	0.08249	0.08083	0.09250	0.04633	0.26423	0.00003
α_{a2}	0.07983	0.05632	0.04239	0.08258	0.84726	0.00003
α_{b2}	0.08148	0.05365	0.04951	0.09838	0.98919	0.00005
β_{a1}	0.78487	0.78966	0.80131	0.07820	0.98780	0.24181
β_{b1}	0.75131	0.77503	0.84424	0.13124	0.99199	0.00701
β_{a2}	0.35458	0.30014	0.04862	0.26664	0.97195	0.00003
β_{b2}	0.30622	0.25441	0.04660	0.23925	0.93140	0.00003
η_1	0.28956	0.29155	0.29340	0.05435	0.45615	0.09449
η_2	0.23313	0.21638	0.23318	0.15227	0.93261	0.00003
ζ_1	0.91842	0.92095	0.92186	0.02831	0.97965	0.81452
ζ_2	0.60763	0.70024	0.93094	0.30751	0.97993	0.00009
ι_1	0.06447	0.04285	0.01983	0.06451	0.39594	0.00004
ι_2	0.25179	0.18085	0.04889	0.22446	0.97598	0.00009
ν_1	8.74881	8.19857	5.79482	5.71238	96.69125	1.01080
ν_2	17.92063	7.57842	5.94400	23.28419	99.87698	1.00016

* 'a', 'b' denotes the first and second series of bivariate sample data; '1', '2' represents the first and second mixture component included in either MGM or MTM.

Table 7.8 Posterior estimation result of stock index data (sample size: 1000; number of iterations: 10000 (Burn-in) and 5000 (In equilibrium))

Panel A. ADCC-MGM model estimated on stock index data

	Mean	Median	Mode	S.t.d	Max	Min
π_1	0.61552	0.58935	0.52824	0.09690	0.95197	0.45346
μ_{a1}	0.00068	0.00069	0.00073	0.00030	0.00132	-0.00036
μ_{b1}	0.00094	0.00098	0.00108	0.00022	0.00126	0.00005
ω_{a1}	0.00001	0.00001	0.00001	0.00001	0.00003	0.00000
ω_{b1}	0.00001	0.00000	0.00000	0.00000	0.00002	0.00000
ω_{a2}	0.00001	0.00001	0.00001	0.00001	0.00005	0.00000
ω_{b2}	0.00003	0.00003	0.00004	0.00001	0.00004	0.00000
α_{a1}	0.09283	0.08798	0.09182	0.05474	0.36727	0.00000
α_{b1}	0.07465	0.06822	0.06894	0.04761	0.27573	0.00001
α_{a2}	0.20181	0.18874	0.16774	0.10056	0.66873	0.00074
α_{b2}	0.07684	0.06079	0.02255	0.06418	0.45043	0.00003
β_{a1}	0.52776	0.57611	0.64683	0.22644	0.99506	0.00012
β_{b1}	0.62850	0.66643	0.73891	0.18600	0.98472	0.00150
β_{a2}	0.68617	0.70829	0.70542	0.13323	0.99539	0.16690
β_{b2}	0.56835	0.56018	0.51088	0.15632	0.96632	0.13825
η_1	0.20501	0.20952	0.25321	0.11282	0.56255	0.00012
η_2	0.08024	0.06593	0.02472	0.06316	0.49400	0.00002
ζ_1	0.41443	0.39252	0.24493	0.25332	0.97938	0.00011
ζ_2	0.48850	0.49012	0.63695	0.27664	0.97980	0.00023
ι_1	0.13052	0.10453	0.04203	0.10522	0.83934	0.00007
ι_2	0.07534	0.06028	0.02928	0.06426	0.58540	0.00002

Panel B. ADCC-MTM model estimated on stock index data

	Mean	Median	Mode	S.t.d	Max	Min
π_1	0.70948	0.72626	0.75044	0.10564	0.99000	0.45766
μ_{a1}	0.00068	0.00070	0.00072	0.00032	0.00132	-0.00039
μ_{b1}	0.00055	0.00054	0.00049	0.00035	0.00126	-0.00044
ω_{a1}	0.00001	0.00001	0.00001	0.00001	0.00003	0.00000
ω_{b1}	0.00001	0.00001	0.00001	0.00001	0.00004	0.00000
ω_{a2}	0.00001	0.00001	0.00001	0.00001	0.00005	0.00000
ω_{b2}	0.00001	0.00000	0.00000	0.00001	0.00004	0.00000
α_{a1}	0.11115	0.10801	0.08536	0.05404	0.34138	0.00003
α_{b1}	0.06365	0.05771	0.04778	0.04131	0.31852	0.00000
α_{a2}	0.11722	0.09825	0.03790	0.09306	0.75642	0.00008
α_{b2}	0.07569	0.06032	0.04722	0.06999	0.94426	0.00000
β_{a1}	0.65050	0.70842	0.74801	0.21244	0.99735	0.00000
β_{b1}	0.59779	0.63521	0.73631	0.19897	0.98142	0.00100
β_{a2}	0.40869	0.39779	0.04846	0.25591	0.96897	0.00001
β_{b2}	0.30938	0.28967	0.04922	0.21528	0.98368	0.00004
η_1	0.08685	0.06829	0.02332	0.07424	0.46481	0.00008
η_2	0.20578	0.18249	0.03952	0.14900	0.78960	0.00004
ζ_1	0.46045	0.45120	0.04909	0.27490	0.97976	0.00010
ζ_2	0.42253	0.40235	0.14704	0.26339	0.97910	0.00021
ι_1	0.06433	0.04991	0.02776	0.05849	0.55500	0.00001
ι_2	0.18637	0.12763	0.04753	0.17606	0.95054	0.00000
ν_1	23.30019	15.89370	15.83946	21.12179	99.92278	1.00123
ν_2	40.55162	34.58961	5.95334	29.87513	99.95393	1.00594

* ‘a’, ’b’ denotes the first and second series of bivariate sample data; ‘1’, ‘2’ represents the first and second mixture component included in either MGM or MTM.

Table 7.9 Summary statistics of predictive densities of correlation forecasts, VaR forecast, return forecast and minimized variance

		Estimated using ADCC-MGM						Estimated using ADCC-MTM					
Simulated		Mean	Median	Mode	S.t.d	Max.	Min	Mean	Median	Mode	S.t.d	Max.	Min
Data	ρ_{t+1}	0.7751	0.7869	0.7997	0.0422	0.8402	0.5701	0.7573	0.7713	0.7948	0.0380	0.8062	0.5777
	minimized δ^2	0.4802	0.4558	0.4435	0.0327	3.2181	0.0492	0.5166	0.4539	0.4641	0.0698	11.531	0.0430
	y_{t+1}(1st series)	0.0023	0.0022	-0.019	0.0882	0.3329	-0.306	0.0510	0.0385	0.4817	1.9697	103.18	-17.64
	y_{t+1}(2rd series)	0.0059	0.0113	-0.057	0.2979	1.1785	-1.068	0.0467	0.0121	5.8371	2.3562	136.39	-17.20
	VaR-99% (1st series)	-0.203	-0.202	-0.200	0.0103	-0.184	-0.230	-0.334	-0.3347	-0.333	0.0325	-0.226	-0.454
	VaR-99% (2rd series)	-0.684	-0.685	-0.681	0.0360	-0.616	-0.760	-0.331	-0.3282	-0.321	0.0291	-0.276	-0.405
Exchange rate	ρ_{t+1}	-0.326	-0.328	-0.327	0.012	-0.226	-0.382	-0.326	-0.327	-0.335	0.009	-0.244	-0.447
US/UK	minimized δ^2	0.0006	0.0010	0.0006	0.0001	0.0153	0.0000	0.0009	0.0009	0.0008	0.0001	0.0141	0.0000
EU/JP	y_{t+1}(1st series)	0.000	0.000	0.000	0.0049	0.0163	-0.020	0.0021	0.0161	-1.108	1.1767	11.615	-11.52
	y_{t+1}(2rd series)	0.0002	0.0002	0.0021	0.0058	0.027	-0.029	0.0332	0.0155	0.2680	1.1136	5.6050	-6.255
	VaR-99% (1st series)	-0.012	-0.012	-0.012	0.0005	-0.011	-0.014	-0.2807	-0.2816	-0.2842	0.0207	-0.244	-0.334
	VaR-99% (2rd series)	-0.015	-0.014	-0.014	0.0011	-0.013	-0.018	-0.2824	-0.2829	-0.2842	0.0176	-0.253	-0.342
Stock and bond	ρ_{t+1}	-0.034	-0.044	-0.050	0.050	0.235	-0.145	-0.005	-0.013	-0.002	0.071	0.258	-0.214
S&P500	minimized δ^2	0.0012	0.0008	0.0008	0.0002	0.0279	0.0001	0.0009	0.0008	0.0008	0.0001	0.0238	0.0000
10y Bond	y_{t+1}(1st series)	0.0009	0.0010	-0.001	0.0072	0.0420	-0.036	-0.013	-0.014	-0.793	1.4187	32.201	-11.79
	y_{t+1}(2rd series)	0.0002	0.0003	0.0002	0.0027	0.0171	-0.014	-0.003	0.000	-2.307	1.4714	35.223	-14.82
	VaR-99% (1st series)	-0.017	-0.017	-0.018	0.0013	-0.014	-0.020	-0.326	-0.3219	-0.317	0.0305	-0.273	-0.399
	VaR-99% (2rd series)	-0.007	-0.007	-0.007	0.0005	-0.006	-0.008	-0.330	-0.3224	-0.315	0.0374	-0.256	-0.415
Stock index	ρ_{t+1}	0.4395	0.4293	0.4241	0.0295	0.7034	0.3748	0.4334	0.4256	0.4111	0.0313	0.7194	0.3083
FTSE100	minimized δ^2	0.0010	0.0021	0.0017	0.0003	0.0352	0.0002	0.0011	0.0026	0.0029	0.0003	0.0224	0.0000
S&P500	y_{t+1}(1st series)	0.0007	0.0007	-0.001	0.0056	0.0253	-0.023	0.0122	0.0080	-0.038	1.2103	10.164	-18.98
	y_{t+1}(2rd series)	0.0007	0.0008	-0.001	0.0057	0.0282	-0.026	0.000	-0.010	-1.150	1.1827	18.913	-7.84
	VaR-99% (1st series)	-0.013	-0.013	-0.013	0.0009	-0.011	-0.015	-0.2611	-0.261	-0.262	0.0184	-0.226	-0.329
	VaR-99% (2rd series)	-0.014	-0.014	-0.014	0.0010	-0.013	-0.017	-0.2590	-0.255	-0.245	0.0147	-0.237	-0.295

This Table presents the summary statistics of predictive densities of one-step-ahead correlation forecast, minimized variance of a portfolio constructed using bivariate sample data, next day's return forecast and next days VaR forecast at 99% level. For the last two evaluation criteria, statistics are reported for both individual time series.

Table 7.10 Estimation results of fitting conditional correlation models (with correlation targeting included) to simulated data of DGP1

Panel A: Parameter estimation results

	Volatility parameters					
	ω_1	α_1	β_1	ω_2	α_2	β_2
value	0.002032	0	0.74988	0.012386	0	0.81852
s.t.d	-1.50E-06	6.51E-06	0.023554	8.9E-06	1.65E-05	0.001978
significance	**		**	**		**

	Correlation parameters					
DCC	η	ς				
value	0.016276	0.68358				
s.t.d	0.017999	0.30922				
significance		*				
ADCC	η	ς	ι			
value	0.12511	0.78794	0.18027			
s.t.d	0.068968	0.17291	0.13962			
significance	*	**				
AGDCC	η_{11}	η_{22}	ς_{11}	ς_{22}	ι_{11}	ι_{22}
value	-0.17979	-0.10409	0.82884	0.908	0.24497	0.34566
s.t.d	0.062966	0.044598	0.11034	0.047127	0.097752	0.24804
significance	**	**	**	**	**	*

This panel reports the parameter results of four conditional correlation models (CCC, DCC, ADCC and AGDCC) estimated on simulated ADCC-MGM (Gaussian mixture) data. Since in estimation volatility part and correlation part of these models' logliklihood functions are to be optimized separately, we report their corresponding parameters also using different ways. For volatility parameters, since in above models bivariate time series are all estimated using same GARCH(1,1), we only report their result once. For CCC model, since correlation is assumed to be fixed, it does not have any correlation parameters to be reported. Above, ** and * respectively represent the statistical significance level 1% and 5%.

Panel B: In-sample and Out-of-sample analysis

	CCC	DCC	ADCC	AGDCC
LogLikelihood	-2825.3	-2825.8	-2826.1	-2827.4
Optimal weights	(1.2802, -0.28016)	(1.2803, -0.28027)	(1.283, -0.28304)	(1.2799, -0.27988)
Minimized portfolio variance	0.5096	0.50933	0.50439	0.50979
One-step-ahead correlation	0.80337	0.77688	0.77439	0.77639

This panel reports the logliklihood of CCC, DCC, ADCC and AGDCC estimated on simulated ADCC-MGM data. Also presented are optimal weight of each asset and minimized portfolio variance when these models are used to construct an unconstrained optimal portfolio for given bivariate data where short selling is allowed. Besides, we also report one-step-ahead correlation forecasts generated by these dynamic models.

Table 7.11 Estimation results of fitting conditional correlation models (with correlation targeting included) to simulated data of DGP2

Panel A. Parameter results

	Volatility parameters					
	ω_1	α_1	β_1	ω_2	α_2	β_2
value	0.0076356	0.059162	0.29828	0.040792	0.03106	0.5334
s.t.d	6.64E-06	0.0009413	0.043573	0.00018028	0.0003057	0.022555
significance	**	**	**	**	**	**

	Correlation parameters					
DCC	η	ς				
value	0.0052583	0.98586				
s.t.d	0.0032101	0.011458				
significance	**	*				
ADCC	η	ς	ι			
value	0.21806	0.60002	0			
s.t.d	0.049317	0.19576	0.17293			
significance	**	**				
AGDCC	η_{11}	η_{22}	ς_{11}	ς_{22}	ι_{11}	ι_{22}
value	-0.30901	-0.20011	0.45805	0.81032	0.26045	0.066761
s.t.d	0.06447	0.038426	0.15918	0.059163	0.39742	0.2336
significance	**	**	**	**		

This panel reports the parameter results of four conditional correlation models (CCC, DCC, ADCC and AGDCC) estimated on simulated ADCC-MTM (T mixture) data. Since in estimation volatility part and correlation part of these models' logliklihood functions are to be optimized separately, we report their corresponding parameters also using different ways. For volatility parameters, since in above models bivariate time series are all estimated using same GARCH(1,1), we only report their result once. For CCC model, since correlation is assumed to be fixed, it does not have any correlation parameters to be reported. Above, ** and * respectively represent the statistical significance level 1% and 5%.

Panel B. In-sample and Out-of-sample analysis

	CCC	DCC	ADCC	AGDCC
LogLikelihood	-2109.3	-2112.7	-2112.3	-2115
Optimal weights	(1.2789, -0.27886)	(1.2802, -0.28021)	(1.2791, -0.27915)	(1.2788, -0.27882)
Minimized portfolio variance	0.7801	0.77564	0.77589	0.77515
One-step-ahead correlation	0.79565	0.75986	0.74572	0.70448

This panel reports the logliklihood of CCC, DCC, ADCC and AGDCC estimated on simulated ADCC-MTM data. Also presented are optimal weight of each asset and minimized portfolio variance when these models are used to construct an unconstrained optimal portfolio for given bivariate data where short selling is allowed. Besides, we also report one-step-ahead correlation forecasts generated by these dynamic models.

Table 7.12 Estimation results of fitting conditional correlation models (with correlation targeting included) to foreign exchange data

Panel A. Parameter results

	Volatility parameters					
	ω_1	α_1	β_1	ω_2	A_2	β_2
value	1.16E-06	0.057516	0.89652	1.58E-06	0.089938	0.88657
s.t.d	1.63E-13	0.0001673	0.0005385	2.20E-12	0.0011619	0.0032396
significance	**	**	**	**	**	**

	Correlation parameters					
DCC	η	ς				
value	0.0076061	0.96835				
s.t.d	0.0060627	0.026908				
significance		**				
ADCC	η	ς	ι			
value	-1.69E-06	1	-2.24E-05			
s.t.d	0.11122	3.06	0.098457			
significance						
AGDCC	η_{11}	η_{22}	ς_{11}	ς_{22}	ι_{11}	ι_{22}
value	0.071969	-0.067336	0.14603	0.99315	-0.3829	0.33361
s.t.d	0.17998	0.070772	0.14131	0.0037751	0.23456	0.063156
significance				**	**	**

This panel reports the parameter results of four conditional correlation models (CCC, DCC, ADCC and AGDCC) estimated on exchange rate data. Since in estimation volatility part and correlation part of these models' logliklihood functions are to be optimized separately, we report their corresponding parameters also using different ways. For volatility parameters, since in above models bivariate time series are all estimated using same GARCH(1,1), we only report their result once. For CCC model, since correlation is assumed to be fixed, it does not have any correlation parameters to be reported. Above, ** and * respectively represent the statistical significance level 1% and 5%.

Panel B. In-sample and out-of-sample analysis

	CCC	DCC	ADCC	AGDCC
LogLikelihood	-12420	-12423	-12420	-12423
Optimal weights	(0.6322, 0.3678)	(0.6325, 0.3675)	(0.63258, 0.36742)	(0.63283, 0.36717)
Minimized portfolio variance	0.0011051	0.0011022	0.001112	0.001114
One-step-ahead correlation	-0.33538	-0.29894	-0.33107	-0.38924

This panel reports the logliklihood of CCC, DCC, ADCC and AGDCC estimated on exchange rate data. Also presented are optimal weight of each asset and minimized portfolio variance when these models are used to construct an unconstrained optimal portfolio for given bivariate data where short selling is allowed. Besides, we also report one-step-ahead correlation forecasts generated by these dynamic models.

Table 7.13 Estimation results of fitting conditional correlation models (with correlation targeting included) to stock and bond data

Panel A. Parameter results

	Volatility parameters					
	ω_1	α_1	β_1	ω_2	A_2	B_2
value	5.36E-06	0.104	0.77434	1.23E-05	0.059367	0.66185
s.t.d	4.90E-12	0.000667	0.0043416	1.32E-10	0.0006688	0.065029
significance	**	**	**	**	**	**

	Correlation parameters					
DCC	η	ς				
value	0.0055277	0.99167				
s.t.d	0.0029946	0.0044185				
significance	**	**				
ADCC	η	ς	ι			
value	0.00E+00	0.27992	0.00E+00			
s.t.d	0.67629	1.7926	0.75038			
significance						
AGDCC	η_{11}	η_{22}	ς_{11}	ς_{22}	ι_{11}	ι_{22}
value	0.12441	0.59338	0.58538	0.36226	1	0.075441
s.t.d	0.25814	0.49452	2.42	0.27654	5.8121	1.3297
significance						

This panel reports the parameter results of four conditional correlation models (CCC, DCC, ADCC and AGDCC) estimated on stock and bond data. Since in estimation volatility part and correlation part of these models' logliklihood functions are to be optimized separately, we report their corresponding parameters also using different ways. For volatility parameters, since in above models bivariate time series are all estimated using same GARCH(1,1), we only report their result once. For CCC model, since correlation is assumed to be fixed, it does not have any correlation parameters to be reported. Above, ** and * respectively represent the statistical significance level 1% and 5%.

Panel B. In-sample and Out-of-sample analysis

	CCC	DCC	ADCC	AGDCC
LogLikelihood	-7144.5	-7146.7	-7144.5	-7147.9
Optimal weights	(0.51855, 0.48145)	(0.51504, 0.48496)	(0.51866, 0.48134)	(0.51747, 0.48253)
Minimized portfolio variance	0.003039	0.0030308	0.0030455	0.003098
One-step-ahead correlation	0.42282	0.50198	0.42598	0.48386

This panel reports the logliklihood of CCC, DCC, ADCC and AGDCC estimated on stock and bond data. Also presented are optimal weight of each asset and minimized portfolio variance when these models are used to construct an unconstrained optimal portfolio for given bivariate data where short selling is allowed. Besides, we also report one-step-ahead correlation forecasts generated by these dynamic models.

Table 7.14 Estimation results of fitting conditional correlation models (with correlation targeting included) to stock index data

Panel A. Parameter results

	Volatility parameters					
	ω_1	α_1	β_1	ω_2	α_2	B_2
value	6.59E-07	0.060165	0.9357	2.10E-07	0.051779	0.93658
s.t.d	6.11E-14	0.000138	0.0001384	2.72E-15	0.0001032	0.0001023
significance	**	**	**	**	**	**

	Correlation parameters					
DCC	η	ς				
value	0.030176	0.96574				
s.t.d	0.0048694	0.0059456				
significance	**	**				
ADCC	η	ς	ι			
value	1.74E-01	0.9827	2.76E-02			
s.t.d	0.013746	0.0031766	0.19103			
significance	**	**				
AGDCC	η_{11}	η_{22}	ς_{11}	ς_{22}	ι_{11}	ι_{22}
value	0.11676	0.24767	0.99764	0.96936	0.068501	0.02111
s.t.d	0.017064	0.039426	0.0027736	0.0057436	0.10377	0.12207
significance	**	**	**	**		

This panel reports the parameter results of four conditional correlation models (CCC, DCC, ADCC and AGDCC) estimated on stock index data. Since in estimation volatility part and correlation part of these models' logliklihood functions are to be optimized separately, we report their corresponding parameters also using different ways. For volatility parameters, since in above models bivariate time series are all estimated using same GARCH(1,1), we only report their result once. For CCC model, since correlation is assumed to be fixed, it does not have any correlation parameters to be reported. Above, ** and * respectively represent the statistical significance level 1% and 5%.

Panel B. In-sample and Out-of-sample analysis

	CCC	DCC	ADCC	AGDCC
LogLikelihood	-22067	-22241	-22241	-22244
Optimal weights	(0.15229, 0.84771)	(0.13061, 0.86939)	(0.1305, 0.8695)	(0.12351, 0.87649)
Minimized portfolio variance	0.0012705	0.0011758	0.0011763	0.0011909
One-step-ahead correlation	0.014602	0.23692	0.23852	0.29701

This panel reports the logliklihood of CCC, DCC, ADCC and AGDCC estimated on stock index data. Also presented are optimal weight of each asset and minimized portfolio variance when these models are used to construct an unconstrained optimal portfolio for given bivariate data where short selling is allowed. Besides, we also report one-step-ahead correlation forecasts generated by these dynamic models.

Table 7.15 Parameter estimation results of ADCC-Gaussian with *constant term* ω included for five datasets

		Correlation parameters							
Simulated		ω	η	ς	ι	LogF	ρ_{t+1}	variance	weights
ADCC-MGM	value	2.11E-07	2.11E-07	1	2.11E-07	-2864.5	0.80373	0.5082	(1.2804,
	s.t.d	7.56E-01	10.908	0.54184	1.41E+00				-0.28041)
				**					
Simulated		ω	η	ς	ι				
ADCC-MTM	value	2.11E-07	9.77E-06	0.99305	2.11E-07	-2148.3	0.73216	0.78952	(1.2734,
	s.t.d	2.46E-05	0.0001932	0.0039958	5.14E-05				-0.2734)
				**					
Exchange rate		ω	η	ς	ι				
US/UK	value	0.0006645	3.97E-04	0.99775	2.11E-07	-12609	-0.1114	0.001172	(0.63393,
EU/JP	s.t.d	0.033082	0.50388	0.040684	0.9269				0.36607)
				**					
Stock index		ω	η	ς	ι				
FTSE100	value	2.11E-07	5.52E-06	0.99851	1.37E-02	-7324.3	0.56152	0.0031808	(0.51552,
S&P500	s.t.d	6.67E-03	1.76E-02	0.0011262	0.0092976				0.48448)
			*	**					
Stock and bond		ω	η	ς	ι				
S&P500	value	0.023966	1.65E-01	0.98599	2.11E-07	-22433	0.1408	0.0011629	(0.11679,
10y Bond	s.t.d	0.27829	1.5183	0.10799	0.31455				0.88321)
				**					

This panel reports the estimation result of ADCC model on five datasets. Along with the parameter estimates and their corresponding standard errors, also presented are calculated logliklihood function value, optimal weight of each asset and minimized portfolio variance when these models are used to construct an unconstrained optimal portfolio where short selling is allowed. Besides, we also report one-step-ahead correlation forecasts generated by this dynamic correlation model.

Table 7.16 Parameter estimation results of AGDCC-Gaussian model with *constant term* C included for five datasets

		Correlation parameters												
Simulated		C_1	C_2	C_3	η_{11}	η_{22}	ς_{11}	ς_{22}	ι_{11}	ι_{22}	LogF	ρ_{t+1}	variance	weights
ADCC-MGM	value	0.30887	1.7567	1.2578	0.11815	0.46322	0.85041	0.80263	0.14749	0.75778	-2866.2	0.80244	0.50832	(1.2797,
	s.t.d	0.22202	2.4062	1.7049	0.094067	0.64514	0.081374	0.11766	0.15699	1.0457				-0.27974)
	significance	*			*		**	**						
Simulated														
ADCC-MTM	value	0.040033	0.035434	2.11E-07	0.068827	0.10742	0.98946	0.99177	1.41E-01	7.99E-02	-2151.4	0.7772	0.77912	(1.2782,
	s.t.d	0.14464	0.084234	0.040419	0.079697	0.13834	7.44E-03	3.77E-03	0.149	0.089565				-0.27819)
	significance						**	**						
Exchange rate														
US/UK	value	2.11E-07	2.11E-07	2.11E-07	0.023326	0.094221	0.99466	0.99775	2.11E-07	0.000141	-12607	-0.10874	0.001136	(0.63389,
EU/JP	s.t.d	0.066926	9.75E-02	2.71E-02	0.376884	0.09936	0.068649	0.26094	9.75E-02	2.71E-02				0.36611)
	significance						**	**						
Stock index														
FTSE100	value	6.85E-07	0.65732	0.8037	2.11E-07	0.80992	0.004682	0.36374	0.000228	0.58033	-7337.1	0.36775	0.003168	(0.51542,
S&P500	s.t.d	9.79E-06	1.40E-04	7.62E-06	4.33E-06	1.44E-07	0.003366	9.61E-06	3.47E-05	0.00438				0.48458)
	significance		**	**		**	*	**	**	**				
Stock and bond														
S&P500	value	0.044617	0.007893	2.11E-07	0.5256	0.06624	0.99099	0.97396	6.50E-05	1.81E-06	-22439	0.10944	0.001179	(0.12603,
10y Bond	s.t.d	0.066926	0.01184	3.17E-04	0.7884	0.09936	0.486485	0.26094	9.75E-02	2.71E-02				0.87397)
	significance						**	**						

This panel reports the estimation result of AGDCC model on five datasets. Along with the parameter estimates and their corresponding standard errors, also presented are calculated logliklihood function value, optimal weight of each asset and minimized portfolio variance when these models are used to construct an unconstrained optimal portfolio where short selling is allowed. Besides, we also report one-step-ahead correlation forecasts generated by this dynamic correlation model.

Table 7.17 Parameter estimation results of ADCC-*t* with *constant term* ω included for five datasets

		Correlation parameters							
Simulated		ω	η	ς	ι	v	ρ_{t+1}	minimized var	optimal weight
ADCC-MGM	value	0.0044	0.0000	1.0000	0.0000	44.6050	0.8111	0.5019	(1.2839
	s.t.d	72.1120	83.5540	16.3740	17.4000	487.0000			-0.28394)
Simulated		ω	η	ς	ι	v			
ADCC-MTM	value	0.0295	0.0549	0.9981	0.0000	8.8297	0.8013	0.6567	(1.3276
	s.t.d	5.5271	11.0590	1.2162	3.7633	0.9647			-0.32759)
						***			
Exchange rate		ω	η	ς	ι	v			
US/UK	value	0.0000	0.0152	0.9961	0.0000	18.0980	-0.1146	0.00108	(0.6308
EU/JP	s.t.d	0.0057	0.0208	0.0019	0.0042	3.5302			0.36924)
				***		***			
Stock index		ω	η	ς	ι	v			
FTSE100	value	0.0116	0.0000	0.9989	0.0000	15.3770	0.6126	0.0032	(0.5159
S&P500	s.t.d	0.0059	0.0376	0.0009	0.0158	3.4069			0.48409)
		**		***		***			
Stock and bond		ω	η	ς	ι	v			
S&P500	value	0.0195	0.1926	0.9811	0.0000	12.2470	0.1149	0.0011	(0.1169
10y Bond	s.t.d	0.0169	0.0714	0.0033	0.0740	0.9925			0.88309)
			***	***		***			

This panel reports the estimation result of ADCC-*t* model on five datasets. Along with the parameter estimates and their corresponding standard errors, also presented are optimal weight of each asset and minimized portfolio variance when these models are used to construct an unconstrained optimal portfolio where short selling is allowed. Besides, we also report one-step-ahead correlation forecasts generated by this dynamic correlation model.***,** and * respectively represents the significance of parameter, different from zero, at 99%,95% and 90% level. 0.000 here denotes a very small number.

Table 7.18 Parameter estimation results of AGDCC-*t* model with *constant term C* included for five datasets

		Correlation parameters												
Simulated		C_1	C_2	C_3	η_{11}	η_{22}	ς_{11}	ς_{22}	ι_{11}	ι_{22}	v	ρ_{t+1}	variance	weights
ADCC-MGM	value	0.3301	2.0000	0.6824	0.1374	0.4090	0.9150	0.7823	0.2102	0.8434	42.7180	0.8103	0.4949	(1.2873
	s.t.d	0.1896	2.1811	1.3183	0.0918	0.5275	0.0374	0.1200	0.1469	1.1387	17.4410			-0.2873)
		*			*		***	***	*		***			
Simulated														
ADCC-MTM	value	0.9565	1.2268	0.7121	0.2510	0.4343	0.6217	0.6871	0.2079	0.0000	8.6092	0.8336	0.6430	(1.3324
	s.t.d	0.5547	0.5595	0.3549	0.1609	0.1878	0.1784	0.1121	0.1559	0.0244	0.8547			-0.3324)
		**	**	**	*	**	***	***			***			
Exchange rate														
US/UK	value	0.0000	1.1125	0.3162	0.0559	2.0000	0.3358	0.9592	0.0000	0.2624	11.9860	0.7193	0.00144	(0.66064
EU/JP	s.t.d	-	-	-	-	-	-	-	-	-	-			0.33936)
Stock index														
FTSE100	value	0.0000	1.5115	1.8096	0.0000	1.7782	0.0002	0.3016	0.2513	1.4300	16.3460	0.3513	0.0032	(0.51582
S&P500	s.t.d	0.0000	0.0003	0.1661	0.0000	0.0012	0.0001	0.0001	0.0011	0.0001	0.0008			0.48418)
		***	***	***	**	***	***	***	***	***	***			
Stock and bond														
S&P500	value	0.0000	0.0000	0.0000	0.3189	0.1161	0.9841	0.9769	0.0653	0.0251	12.3090	0.1153	0.0012	(0.118
10y Bond	s.t.d	0.3871	5.3968	23.2450	0.0927	0.0581	0.4258	7.2284	0.7659	15.5810	1.0926			0.882)
							***	**						

This panel reports the estimation result of AGDCC-*t* model on five datasets. Along with the parameter estimates and their corresponding standard errors, also presented are logliklihood value, optimal weight of each asset and minimized portfolio variance when these models are used to construct an unconstrained optimal portfolio where short selling is allowed. Besides, we also report one-step-ahead correlation forecasts generated by this dynamic correlation model. ***, ** and * respectively represents the significance of parameter, different from zero, at 99%,95% and 90% level

Table 7.19 Parameter estimation results of ADCC-*skew-t* model with *constant term* ω included for five datasets

		Correlation parameters									
Simulated		ω	η	ς	ι	v	ζ_1	ζ_2	ρ_{t+1}	minimized var	optimal weight
ADCC-MGM	value	0.0000	0.0000	1.0000	0.0066	44.9440	0.9859	0.9947	0.8109	0.5022	(1.2838
	s.t.d	0.0496	0.3158	0.0370	0.2715	143.0200	0.1572	0.0854			-0.28378)
				***			***	***			
Simulated		ω	η	ς	ι	v	ζ_1	ζ_2			
ADCC-MTM	value	0.1692	0.1405	0.9691	0.1120	6.1408	2.2064	2.4523	0.8319	0.5671	(1.3535
	s.t.d	6.4389	6.4068	1.0165	1.5793	95.8280	112.8500	132.3800			-0.35353)
Exchange rate		ω	η	ς	ι	v	ζ_1	ζ_2			
US/UK	value	0.0000	0.0206	0.9966	0.0000	17.9170	1.0506	0.9261	-0.1527	0.00108	(0.6306
EU/JP	s.t.d	0.0032	0.0252	0.0019	0.0074	3.0195	0.0113	0.0243			0.36936)
				***		***	***	***			
Stock index		ω	η	ς	ι	v	ζ_1	ζ_2			
FTSE100	value	0.0000	0.0000	0.9994	0.0323	13.8210	0.8204	0.8412	0.6164	0.0033	(0.5162
S&P500	s.t.d	0.0094	0.0291	0.0009	0.0142	2.5294	0.0391	0.0265			0.48379)
				***	***	***	***	***			
Stock and bond		ω	η	ς	ι	v	ζ_1	ζ_2			
S&P500	value	0.0203	0.1973	0.9801	0.0000	12.2590	0.9084	0.9734	0.1380	0.0011	(0.1160
10y Bond	s.t.d	0.3159	2.9493	0.0359	0.1898	4.9665	0.0419	0.2380			0.884)
				***		***	***	***			

This panel reports the estimation result of ADCC-*skew-t* model on five datasets. Along with the parameter estimates and their corresponding standard errors, also presented are loglikelihood value, optimal weight of each asset and minimized portfolio variance when these models are used to construct an unconstrained optimal portfolio where short selling is allowed. Besides, we also report one-step-ahead correlation forecasts generated by this dynamic correlation model. ***, ** and * respectively represents the significance of parameter, different from zero, at 99%,95% and 90% level.

Table 7.20 Parameter estimation results of AGDCC-*skew-t* model with *constant term C* included for five datasets

		Correlation parameters														
Simulated		C_1	C_2	C_3	η_{11}	η_{22}	ς_{11}	ς_{22}	ι_{11}	ι_{22}	ν	ζ_1	ζ_2	ρ_{t+1}	variance	weights
ADCC-MGM	value	0.1988	0.3155	0.1893	0.0046	0.0031	0.9906	0.9795	0.1388	0.2036	41.1420	0.9848	0.9935	0.8114	0.4947	(1.2879
	s.t.d	0.0774	0.1026	0.0350	0.0031	0.0533	0.0053	0.0122	0.0554	0.0594	13.9440	0.0230	0.0175			-0.2879)
		**	***	**	*		***	***	***	***	***	***	***			
Simulated																
ADCC-MTM	value	0.2397	0.2681	0.1403	0.0744	0.1240	0.8206	0.7238	0.1910	0.5172	8.6184	1.0170	1.0051	0.8397	0.6432	(1.3326
	s.t.d	0.3637	0.5744	0.2910	0.1206	0.2723	0.1033	0.0907	0.3371	1.1808	0.7623	0.0218	0.0181			-0.3326)
							***	***			***	***	***			
Exchange rate																
US/UK	value	0.0000	0.0318	0.2060	0.1554	1.0568	0.9748	0.8318	0.2444	0.0000	22.6940	1.0910	0.8899	0.2532	0.00133	(0.6497
EU/JP	s.t.d	22.5	7.9	32.8	19.8	35.4	0.9	5.1	13.4	1.0	463.8	1.6	2.6			0.3503)
Stock index																
FTSE100	value	0.0097	0.6599	0.9006	0.0033	0.7971	0.1405	0.0000	0.0472	0.3942	16.3030	0.8613	0.8853	0.4759	0.0032	(0.5170
S&P500	s.t.d	0.0724	0.0468	0.0570	0.0245	0.2291	0.0872	0.1165	0.4013	0.0780	1.9671	0.0365	0.0333			0.4829)
			***	***		***	*			***	***	***	***			
Stock and bond																
S&P500	value	0.0315	0.0109	0.0000	0.3323	0.1167	0.9835	0.9773	0.0000	0.0000	12.2450	1.0186	0.9967	0.1331	0.0012	(0.8822
10y Bond	s.t.d	0.0519	0.0221	0.0334	0.2984	0.1097	0.0043	0.0063	0.4601	0.1770	1.0926	0.0120	0.0122			0.1178)
							***	***			***	***	***			

This panel reports the estimation result of AGDCC-*skew-t* model on five datasets. Along with the parameter estimates and their corresponding standard errors, also presented are logliklihood value, optimal weight of each asset and minimized portfolio variance when these models are used to construct an unconstrained optimal portfolio where short selling is allowed. Besides, we also report one-step-ahead correlation forecasts generated by this dynamic correlation model. ***, ** and * respectively represents the significance of parameter, different from zero, at 99%,95% and 90% level.

Table 7.21 Minimized portfolio variance generated by applying various correlation models to two simulated data and three empirical data

	ADCC-G	AGDCC-G	ADCC-t	AGDCC-t	ADCC-skew-t	AGDCC-skew-t	ADCC-MGM	ADCC-MTM
DGP1	0.50820	0.50832	0.50190	0.49490	0.50220	0.49490	0.48020	-
DGP2	0.78920	0.77912	0.65670	0.64300	0.56710	0.64320	-	0.51660
Exchange rate data	0.00117	0.00114	0.00108	0.00144	0.00108	0.00133	0.00060	0.00090
Stock and bond data	0.00318	0.00317	0.00320	0.00320	0.00330	0.00320	0.00120	0.00090
Stock index data	0.00116	0.00118	0.00110	0.00120	0.00110	0.00120	0.00100	0.00110

This panel reports the minimized portfolio variance generated by fitting eight different correlation models to two simulated data and three empirical data that analyzed in this research. Here, ADCC-G and AGDCC-G respectively represent the ADCC model and AGDCC model whose innovations are assumed to (one-component) be Gaussian distributed. ADCC-t and AGDCC-t assume the innovations to be t distributed whilst the asymmetric DCC structure for modeling correlation dynamics is kept the same as previously. Other variants of asymmetric DCC are given based on the similar rules where the mechanism for updating correlation is retained but the distributional assumption substituted. For example, the last two represents the ADCC model respectively associated with two component Gaussian mixture distribution and two component T mixture distribution. Here, one thing needs to be noted is for the first six models they are estimated by maximum likelihood and portfolio variance is generated by applying point estimate of covariance to equations illustrated in Section 6.5.4. However, for the last two models, their values are then reported by using posterior mean since Bayesian inferential method is now implemented.

Table 7.22 Normality and autocorrelation test results for standardized residuals generated from the fitting of, ADCC-*Gaussian/T* model, AGDCC- *Gaussian/T* model, ADCC-*skew-t* and AGDCC-*skew-t* all with *constant term* ω (or *C*) included, to simulated- and empirical- data.

	ADCC-Gaussian			ADCC-*t*			AGDCC-Gaussian			AGDCC-*t*			ADCC-skew-t			AGDCC-skew-t		
simulated ADCC-MGM	$\chi2$	Q(20)	$Q^2(20)$	$\chi2$	Q(20)	$Q^2(20)$	$\chi2$	Q(20)	$Q^2(20)$	$\chi2$	Q(20)	$Q^2(20)$	$\chi2$	Q(20)	$Q^2(20)$	$\chi2$	Q(20)	$Q^2(20)$
1st series	0.150	0.328	0.565	0.143	0.335	0.546	0.318	0.329	0.998	0.280	0.328	0.864	0.144	0.335	0.547	0.267	0.302	0.548
2rd series	0.000	0.209	0.239	0.000	0.210	0.238	0.000	0.208	0.296	0.000	0.211	0.272	0.000	0.210	0.238	0.000	0.208	0.237
simulated ADCC-MTM																		
1st series	0.000	0.347	0.930	0.000	0.434	0.864	0.000	0.422	0.920	0.000	0.411	0.996	0.000	0.925	0.102	0.000	0.493	0.802
2rd series	0.000	0.729	0.733	0.000	0.676	0.704	0.000	0.702	0.745	0.000	0.645	0.852	0.000	0.731	0.821	0.000	0.679	0.732
exchange rate data																		
US/UK	0.000	0.174	0.409	0.000	0.143	0.504	0.000	0.140	0.474	0.000	0.155	0.669	0.000	0.160	0.463	0.000	0.155	0.476
EU/JP	0.000	0.241	0.873	0.000	0.229	0.876	0.000	0.224	0.846	0.000	0.163	0.292	0.000	0.244	0.879	0.000	0.238	0.879
stock index data																		
S&P500	0.000	0.000	0.470	0.000	0.000	0.500	0.000	0.000	0.122	0.001	0.000	0.176	0.000	0.000	0.477	0.000	0.000	0.148
FTSE100	0.000	0.001	0.635	0.000	0.001	0.500	0.000	0.020	0.532	0.000	0.014	0.490	0.000	0.001	0.655	0.000	0.001	0.885
stock and bond																		
S&P500	0.000	0.928	0.957	0.000	0.933	0.964	0.000	0.949	0.987	0.000	0.942	0.966	0.000	0.972	0.910	0.000	0.981	0.918
10y Bond	0.000	0.148	0.935	0.000	0.140	0.988	0.000	0.144	0.991	0.000	0.146	0.991	0.000	0.162	0.718	0.000	0.167	0.736

This panel reports normality and autocorrelation results for standardized residuals generated from fitting ADCC-Gaussian/t, AGDCC-Gaussian/t, ADCC-*skew-t* and AGDCC-*skew-t* models (all with constant term included) to two simulated-data and three empirical data. The first column reports the *p*-values of Jarque-Bera normality test whose statistic follow a chi-square distribution. The next two present *p*-values of two autocorrelation tests. Q(20) denotes the *p*-values of Box-Pierce test of order 20 on the standardized residuals and $Q^2(20)$ reports the *p*-values of the same statistic calculated on the squared residuals. Significance level is set to be 95%. Thus any value below 0.05 is an indication for rejecting the null hypothesis for which in this case are univaraite time series is normal distributed and there is no autocorrelation in either standardized residual or its squared products.

Table 7.23 Constant correlation test results and unconditional correlation calculated from standardized residuals generated from the fitting of, ADCC-*Gaussian*/*T* model, AGDCC- *Gaussian*/*T* model, ADCC-*skew-t* and AGDCC-*skew-t* all with *constant term* ω (or *C*) included, to simulated- and empirical data.

		ADCC-normal	ADCC-*t*	AGDCC-normal	AGDCC-*t*	ADCC-*skew-t*	AGDCC-*skew-t*
	Before-	After-	After-	After-	After-	After-	After-
simulated ADCC-MGM							
unconditional corr.	0.804	-0.002	-0.011	0.001	-0.019	-0.011	-0.018
χ^2	3.638	3.559	3.553	5.176	5.227	3.554	4.073
p-values	0.457	0.469	0.470	0.270	0.265	0.470	0.396
simulated ADCC-MTM							
unconditional corr.	0.796	0.014	-0.123	0.003	-0.135	-0.353	-0.138
χ^2	3.080	3.653	3.442	2.393	1.766	25.590	1.617
p-values	0.545	0.455	0.487	0.664	0.779	0.000	0.806
exchange rate data							
unconditional corr.	-0.318	-0.033	0.022	-0.015	-0.213	0.021	0.022
χ^2	1.129	0.991	1.142	1.245	40.606	0.786	0.852
p-values	0.890	0.911	0.888	0.871	0.000	0.940	0.931
stock index data							
unconditional corr.	0.426	-0.052	-0.092	-0.021	-0.060	-0.046	-0.107
χ^2	1.673	1.783	1.821	4.338	5.526	1.660	4.310
p-values	0.796	0.776	0.769	0.362	0.237	0.798	0.366
stock and bond							
unconditional corr.	-0.086	-0.026	-0.014	-0.009	-0.011	-0.016	-0.014
χ^2	62.748	9.103	5.950	5.414	4.636	4.250	3.514
p-values	0.000	0.059	0.203	0.247	0.327	0.203	0.461

This panel reports Engle's constant correlation test result and unconditional correlation of standardized residuals calculated from fitting ADCC-Gaussian/t, AGDCC-Gaussian/t, ADCC-*skew-t* and AGDCC-*skew-t* models (all with constant term included) to five simulated- and empirical data. The first row reports the unconditional correlation, while the next two present the Engle's test statistic (chi-square distributed) and its associated *p*-value (with three lags). The first column reports the results for un-standardized return. The remaining then present those after return is standardized by various GARCH volatilities. Significance level here is set to be 95%.

Appendix I. Dependence measures.

A. Linear Correlation

The most popular way to calculate the relationship between two variables is to use the linear correlation. Let $(X,Y)^T$ be a vector of random variables with nonzero finite variance. The linear correlation coefficient for $(X,Y)^T$ is defined as

$$\rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} \tag{I.1}$$

where Cov(X,Y)=E(XY)-E(X)E(Y) is the covariance of $(X,Y)^T$, and Var(X) and Var(Y) are the variance of X and Y. Here, note that this correlation coefficient can only be used to measure the linear dependence. While it possesses invariant property under strictly increasing linear transformation, e.g., $\rho(\alpha X+\beta,\gamma Y+\sigma)=sign(\alpha\gamma)\rho(X,Y)$, the results are sometimes misleading due to the massive evidences observed in financial market rejecting its assumption of X and Y both being univariate normal distributed and (X, Y) being jointly multivariate normal distributed.

B. Copular function

To obtain a more reliable and accurate dependence measure in a multivariate distribution, copular function provides a nature alternative. It models the relationship between two or more variables by splitting the definition of marginal distributions from their joint distribution. For example, in the credit market, a typical use of copular is to price the portfolio-based product such as CDO. Since the major task here is to determine the joint default probability distribution function for multiple credits which does not usually follow normal distribution, then, the use of linear correlation may cause misleading result and this task does not have an explicit solution. As a result, an efficient numerical procedure is then required. Here, copular function can provide an ideal solution to link multiple single-credit (or unidimensional) survive curve to one multi-credit (or multidimensional) survival curve.

Consider a joint distribution function $F(x_1, x_2,\ldots, x_n)$ of random variables $(x_1, x_2,\ldots, x_n)$, according to Theorem 3 of Sklar (1959), this $F(x_1, x_2,\ldots, x_n)$ then can be decomposed into a composition of individual marginal distributions $F_i(x_i)$ and a copular function C(.). That is,

$$F(x_1,x_2,\cdots x_n)=C(F_1(x_1),F_2(x_2),\cdots F_n(x_n)) \tag{I.2}$$

If we replace $F_i(x_i)$ with a new uniform random variate u in [0,1] and invert the above function, the copular function C(.) then can be written as

$$\begin{aligned} C(u_1,u_2,\cdots u_n) &= F(F_1^{-1}(u_1),F_2^{-1}(u_2),\cdots F_n^{-1}(u_n)) \\ &= p(U_1\le u_1,U_2\le u_2,\cdots U_n\le u_n) \end{aligned} \tag{I.3}$$

where $F_i^{-1}(\cdot)$ is the quasi-inverse function of $F_i(\cdot)$.

Here, an important feature of C(.) is this measure is invariant under strictly increasing transformation of the marginal distributions. Note, this transformation function is now only required to be an increasing function; and it can be either linear or nonlinear. Therefore, compared to explicitly linear association considered in the above correlation, copular's advantage of relaxing the restrictions is then obvious (See Embrechts, Lindskog and McNeil, 2001 p6, theorem 2.6 for proofs)

B1. Copular measure

Next, we describe two copular-based dependence measures known as Kendall's *Tau* and Spearman's *rho*. They are also usually referred to as the ranking statistics since the random variables needs to be sorted before calculation.

B1.1. Kendall's Tau

Consider a random vector $(X,Y)^T$, Kendall's ranking correlation (*Tau*) of this vector is defined as

$$\rho_\tau(X,Y) = p\{(X-\tilde{X})(Y-\tilde{Y})>0\} - p\{(X-\tilde{X})(Y-\tilde{Y})<0\} \tag{I.4}$$

where $(\tilde{X},\tilde{Y})^T$ is an independent realization of joint distribution of $(X,Y)^T$. Here, it is clear that this correlation is actually the probability difference between the concordance and discordance of $(X,Y)^T$. To write it in a copular form, *Tau* then can be defined as

$$\rho_\tau(X,Y) = 4\int_0^1\int_0^1 C(u_x,u_y)dC(u_x,u_y)-1 \tag{I.5}$$

or simply, $\rho_\tau(X,Y) = 4\mathrm{E}[C(U_x,U_y)]-1$, where $U_x,U_y \sim U(0,1)$

B1.2. Spearman's rho

Spearman's *rho* of the same random vector $(X,Y)^T$ is given as

$$\rho_s(X,Y) = 3\{p[(X-\tilde{X})(Y-Y')>0] - p[(X-\tilde{X})(Y-Y')<0]\} \tag{I.6}$$

where $(\tilde{X},\tilde{Y})^T$ and $(X',Y')^T$ are independent realizations of joint distribution of $(X,Y)^T$. And its copular form given that the random variables of $(X, Y)^T$ are all continuous can be written as

$$\rho_s(X,Y) = 12\int_0^1\int_0^1 C(u_x,u_y)du_x du_y - 3 \tag{I.7}$$

Here, sine *Tau* $\rho_\tau(X,Y)$ and *rho* $\rho_s(X,Y)$ both can be expressed as a function of copular, they are invariant under monotonic transformations. For a more detailed illustration of these two correlation coefficients, see Kendall and Stuart (1977) and Lehmann (1975).

B2. Copular with dependence structure

Above, the dependence in a multivariate distribution is all depicted through a scalar measure where no particular correlating-structure is assumed. However, a more popular way to calculate the relationship between two or more variables is to assume an inherent distribution, usually the same, for both marginal distributions and joint distribution, and then derive a copular based on this assumption to suit the empirical multivariate data observed. Depending on the correlating structure assumed for random variables, copular functions can vary from the simple (*independence* or *Gaussian* copular) to more complex (*Gumbel* , *Clayton* or *Student-t* copular). In the following, we describe two most popularly used copular models in finance.

B2.1 Gaussian Copular

Consider *n* random variables whose marginal distributions follow standard univariate normal distribution denoted by Φ and their joint distribution follows multivariate normal distribution denoted by Φ_R , the Gaussian copular function of this random vector is then defined as

$$C_R^{Gaussian}(u_1,u_2,\cdots u_n)=\Phi_R^n(\Phi_1^{-1}(u_1),\Phi_2^{-1}(u_2),\cdots,\Phi_n^{-1}(u_n)) \tag{I.8}$$

where R is the linear correlation matrix of multivariate normal and $\Phi^{-1}(\cdot)$ is the inverse of cumulative function of Gaussian. For bivariate random variates, the above function then can be written as

$$C_R^{Gaussian}(u_1,u_2)=\int_{-\infty}^{\Phi^{-1}(u_1)}\int_{-\infty}^{\Phi^{-1}(u_2)}\frac{1}{2\pi\sqrt{1-R^2}}\exp\left\{\frac{x_1^2-2Rx_1x_2+x_2^2}{2(1-R^2)}\right\}dx_1dx_2 \tag{I.9}$$

Here, it is important to note two things. First, while it is a tradition to apply the same distributional type to both marginal distribution and joint distribution of a copular, there is no inherent linkage between two. For example, when Gaussian copular is used to determine the dependency in a credit portfolio, the assumed normally distributed joint relationship is independent of the actual distributions of each individual credit returns (although they are also assumed to be normal). Second, Gaussian copular can be only used to capture the dependence around the mean; it however does not incorporate the dependence around the tails. Although the fat tail is a stylised feature presented in most asset return distributions, the application of Gaussian copular in finance especially in the credit market is still massive compared to other alternatives. Due to the numerical tractability and small number of parameters required, this model has gained substantial popularity among market participants for the risk management purpose. And, nowadays, it nearly becomes a standard framework to model the default correlation just like the similar importance observed for B-S in the modelling of time-varying volatility.

B2.2 Student-t Copular

In the real credit market if correlation is used for trading purpose, more reliable models than Gaussian copular is then required to provide an accurate dependence measure. For example, *Student-t* copular is often suggested as such an alternative. It provides a significant improvement compared to the Gaussian copular on capturing the tail dependence between various credit instruments. And this improvement is essential to capture an important feature of the market, that is, if one name of a credit portfolio tends to default, the probability of another name to default will also increase. By simply putting, the dependence of different credits now tends to increase at the extreme events (at the tails of credit return distribution)

To take into account this tail dependence, consider again an *n*-element random vector (x_1, x_2,…, x_n) whose marginal distributions now follow univariate *t* distributions t_v and their joint distribution follows multivariate *t* distribution $t_{R,v}$, the *student-t* copular function of this random vector then can be defined as

$$C_R^t(u_1,u_2,\cdots u_n)=t_{R,v}^n(t_v^{-1}(u_1),t_v^{-1}(u_2),\cdots,t_v^{-1}(u_v)) \tag{I.10}$$

where *R* is the linear correlation, *v* denotes the degree of freedom parameter, t_v is defined only for *v*>2 and $t_v^{-1}(\cdot)$ is the inverse cumulative distribution function of univariate *t*. For a bivariate data, the above *t* copular then can be rewritten to, (See Picone, 2005)

$$C_{R,v}^t(u_1,u_2)=\int_{-\infty}^{t^{-1}(u_1)}\int_{-\infty}^{t^{-1}(u_2)}\frac{1}{2\pi\sqrt{1-R^2}}\exp\left\{1+\frac{x_1^2-2Rx_1x_2+x_2^2}{v(1-R^2)}\right\}^{-(v+2)/2}dx_1dx_2 \tag{I.11}$$

Appendix II. DCC type modeling of Conditional Covariance

Since DCC type modelling technique is a major content of this research and we will use it throughout the thesis to modelling conditional covariance and correlation, it is then necessary to dedicate a separate part specifically illustrate the modelling structures, statistical characteristics of this type of models. Here, in this appendix we start from presenting the features of Engle (2002)'s standard DCC and then elaborating some variants of it.

Consider a *D-variate* random variable y_t which follows a unknown multivariate distribution Φ after information filtration, if the first central moment of this variable is assume to equal zero and its covariance matrix Σ_t modelled by a dynamic conditional correlation model, for example standard DCC of Engle (2002), Σ_t then can be estimated by firstly writing its specification as $D_tR_tD_t$ and then using independant univariate GARCH processes to model D_t and R_t respectively.

Here, note that in a standard DCC D_t is $d \times d$ diagonal matrix with $\sqrt{\Sigma_{it}}$ on its i^{th} diagonal denoting the *s.t.d* of i^{th} time series. This variable is easy to estimate using traditional optimization process of univaraite GARCH such as BHHH. However, to calculate R_t, the time varying correlation matrix, one then needs to introduce a new auxiliary function so that this matrix can be formed as a by-product of the auxiliary function. For example, in Engle (2002) this auxiliary function, called Q_t, is modeled by another univariate GARCH

$$Q_t = (1-\eta-\varsigma)\bar{Q} + \eta\varepsilon_{t-1}\varepsilon_{t-1}^{'} + \varsigma Q_{t-1} \tag{II.1}$$

where η, ς are scalar vectors denoting the ARCH and GARCH parameter, $\bar{Q}$ represents the unconditional covariance of standard error ε_t, that is $\bar{Q} = E[\varepsilon_t\varepsilon_t^{'}]$. And the general DCC(p,q) model then can be defined as

$$\begin{aligned}
& y_t \mid F_{t-1} \sim \Phi(0,\Sigma_t) \\
& y_t = \varepsilon_t \qquad \Sigma_t = D_tR_tD_t \\
& \underbrace{D_t = \varpi + \sum_{i=1}^{p} \alpha_i y_{t-i} y_{t-i}^{'} + \sum_{j=1}^{q} \beta_j D_{t-j}}_{Volatiilty \quad estimation} \\
& \varepsilon_t = y_t / \sqrt{\Sigma_t} = y_t D_t^{-1} \\
& Q_t = (1 - \sum_{i=1}^{p} \eta_i - \sum_{j=1}^{q} \varsigma_j)\bar{Q} + \sum_{i=1}^{p} \eta_i \varepsilon_{t-i}\varepsilon_{t-i}^{'} + \sum_{j=1}^{q} \varsigma_j Q_{t-j} \\
& \underbrace{R_t = diag(Q_t)^{-1/2} Q_t diag(Q_t)^{-1/2} \qquad\qquad\qquad}_{Correlation \quad estimation}
\end{aligned} \tag{II.2}$$

Above, if we use other processes than GARCH to model auxiliary function, R_t will change and another forms of DCCs can be derived. For example, the simplest variant of standard DCC is the CCC of Bollerslev (1990) where author assumed the conditional correlation no longer a time varying variable but a deterministic constant. For this particular case, dynamic property of correlation matrix is now scarified although estimation cost becomes much lower.

In other cases, variants of DCC are then proposed in more sophisticated ways. Take ADCC (1,1) of Hafner and Franses (2003) for example, auxiliary function Q_t of correlation matrix is assumed to be

$$Q_t = (1-\eta^2-\varsigma^2)\bar{Q} - \iota^2\bar{N} + \eta^2\varepsilon_{t-1}\varepsilon_{t-1}^{'} + \varsigma^2 Q_{t-1} + \iota^2\vartheta_{t-1}\vartheta_{t-1}^{'} \qquad \text{(II.3)}$$

where two new variables are now introduced to account for the asymmetric effects. One is $\vartheta_t = I\left[\varepsilon_t < 0\right]\Theta\varepsilon_t$ which denotes the observations whose values of different time series involved in empirical data at the same date are all negative. The other is $\bar{N}$ that represents the unconditional covariance of ϑ_t , that is $\bar{N} = E[\vartheta_t\vartheta_t^{'}]$. Here, it is worth noting in (II.1) parameter η and ς are both set to be scalar products for simplicity. However, in Hafner and Franses (2003) they are proposed in squared forms so that positive definitiveness of covariance matrix can be ensured. As for stationarity of covariance, this condition is met if we restrict $\eta^2+\varsigma^2+\iota^2<1$

Similarly, Capiello *et al*., (2004) developed another modification of standard DCC by using a set of diagonal matrixes. In their AGDCC (*p,q*) model Q_t is specified, using the same way as Σ_t in Diagonal-BEKK,

$$Q_t = \left(\bar{Q} - \eta^{'}\bar{Q}\eta - \varsigma^{'}\bar{Q}\varsigma - \iota^{'}\bar{N}\iota\right) + \eta^{'}\varepsilon_{t-1}\varepsilon_{t-1}^{'}\eta + \varsigma^{'}Q_{t-1}\varsigma + \iota^{'}\vartheta_{t-1}\vartheta_{t-1}^{'}\iota \qquad \text{(II.4)}$$

where parameters are all defined to be diagonal matrixes so that positive definitiveness of covariance is ensured from the start of modeling and the correlation targeting (or stationarity) is allowed after a nonlinear restriction on parameters η,ς is imposed to constraint the eigenvalues of $\eta+\varsigma$ lies within the unit circle.

Meanwhile, to increase the model flexibility and yield more benefits, Cajigas and Urga (2005) combined ADCC and AGDCC to propose a new dynamic correlation model. By assuming standard error ε_t to be multivariate asymmetric Laplace distributed, they let their Q_t to follow a new hybrid updating process. That is

$$Q_t = \left(1-\overline{\eta}^2-\overline{\varsigma}^2\right)\overline{Q}-\overline{\iota}^2\overline{N}+\eta^{'}\varepsilon_{t-1}\varepsilon_{t-1}^{'}\eta+\varsigma^{'}Q_{t-1}\varsigma+\iota^{'}\vartheta_{t-1}\vartheta_{t-1}^{'}\iota \qquad \text{(II.5)}$$

Surely, apart from changing the form of auxiliary function, we have other ways to propose a generalized DCC. As just illustrated, unknown distribution assumed for standard error can be modified so that probability associated with the dynamic feedback may change. And this way of increasing generality is usually cheaper than just changing Q_t because less parameter will be involved. ADCC-MGM and ADCC-MTM model to be proposed in Chapter 5 in this thesis is just a case of it.

Appendix III. Hierarchical form of multivariate T distribution (MTM)

Specification of multivariate T distribution can be written in several ways. If this density is compounded into a standard mixture, surely, the resulting mixture distribution can also be expressed in similar ways. For example, if a *D*-variate random process y_t whose observations is now assumed to follow a standard *M*-component *t* mixture distribution (MTM), its density in its most common form then can be written as

$$y_t \mid F_{t-1} \sim \sum_{m=1}^{M} \pi_m t\left(y_t, \varphi_m; \mu_m, \Sigma_m, \nu_m\right) \tag{IV.1}$$

where π_m denotes the mixing probability, μ_m, Σ_m and ν_m represent the mean, variance and degree of freedom parameter. After augmenting each observation with a label variable, say z_t, likelihood function of (IV.1) can be defined as

$$\begin{aligned} L\left(\varphi \mid F_{t-1}\right) &\propto \prod_{i \in \{z_i = m\}} f\left(\varphi_m\right) \\ &= \prod_{i \in (z_i = m)} \frac{\pi_m}{2\pi \left|\Sigma_{mt}\right|^{D/2}} \left(1 + \frac{\left(\mathrm{y_t} - \mu_m\right)' \Sigma_{mt}^{-1} \left(y_t - \mu_m\right)}{\nu_\mathrm{m}}\right)^{-\frac{\nu_m + D}{2}} \end{aligned} \tag{IV.2}$$

Above, if we rewrite all component distributions (*student t*) in a hierarchical way, (IV.2) can also be obtained in another form. To illustrate this modification in more details, consider now a new random variable *X* drawn from an *i.i.d* multivariate *t*, say $X \sim t\left(X, \varphi; \mu, \Sigma, \nu\right)$, if its density is now written using (IV.2), we can easily obtain

$$p(X) = \frac{1}{2\pi \left|\Sigma_t\right|^{D/2}} \left(1 + \frac{\left(\mathrm{y_t} - \mu\right)' \Sigma_t^{-1} \left(y_t - \mu\right)}{\nu}\right)^{-\frac{\nu + D}{2}} \tag{IV.3}$$

However, this equation can be re-organized if two hierarchical forms of *standard t* are adopted. One is combination of Normal and Gamma

$$\begin{aligned} X \mid \tau &\sim N(\mu, \Sigma / \tau) \\ \tau &\sim \Gamma(\nu/2, \nu/2) \end{aligned} \tag{IV.4}$$

the other is combination of Normal and Chi square

$$\begin{aligned} X \mid \tau &\sim N(\mu, \Sigma / \tau) \\ \tau \cdot \nu &\sim \chi^2(\nu) \end{aligned} \tag{IV.5}$$

Here, τ denotes the missing weight vector of observable data *X*; $N(\mu,\Sigma/\tau)$ is a normal distribution with mean μ and covariance matrix Σ/τ ; $\Gamma(v/2,v/2)$ and $\chi^2(v)$ respectively represents a *Gamma* distribution and a *Chi square* distribution.

Given above equations, for a *t* mixture model, if *Normal*-*Gamma* is used to define each component *t* distribution, probability of an observation, say y_t , drawn from m^{th} mixture component is then just,

$$\begin{aligned} y_t \mid (\tau_t, z_t = m) &\sim N(\mu_m, \Sigma_m / \tau_t) \\ \tau_t \mid (z_t = m) &\sim \Gamma(v_m/2, v_m/2) \end{aligned} \qquad \text{(IV.6)}$$

And we can obtain its corresponding likelihood function by

$$L(\varphi \mid F_{t-1}) \propto \prod_{i\in\{z_i=m\}} \pi_m N(y_t \mid \mu_m, \Sigma_m/\tau_t)\Gamma(v_m/2, v_m/2) \qquad \text{(IV.7)}$$

after all y_t, z_t, τ are known.

Since in this thesis the purpose is to derive sampling kernels for all parameters in *t* mixture model so that random draws of these parameters can be simulated and their empirical moments estimated, given (IV.4) and (IV.5) we can find a new way different from those depicted in Chapter 6 for simulation. However, to avoid any duplicative task, we do not use it in our analysis. To see its application in similar Bayesian statistics, Lee *et al.* (2004) provided an example.

Appendix IV. Use EM algorithm to estimate M-component Gaussian Mixture distribution

Consider a *D-variate* random variable y_t and a multivariate Gaussian mixture distribution Φ. Presume each observation of this variable is now drawn from a component (Gaussian) of this mixture and associated with a specific indicator variable (or label variable) z_t, which is assumed to be multinomially distributed and having value one for element corresponding to the selected mixture component, and zeros for all others. That is,

$$z_t = \underbrace{\left[0,\, 0, \cdots 0, 1, 0, \cdots 0, 0\right]^T}_{M\ componet} \tag{V.1}$$

Then, after all observations are labelled, we can form a complete information set (y,z) for mixture model. And using EM algorithm to estimate this model is just to maximize the likelihood function of the joint density $f(y,z\,|\,\varphi)$. Here, it is worth noting that this optimization step is different from our traditional task of maximizing the likelihood of $f(y\,|\,\varphi)$ using only observed data y. This is because only after the component label is updated we can know which component generate a specific observation. However, since z_t is now unobservable, often we need to continuously update its information and iterate this procedure with maximization (or optimization) step until the convergence of parameter values can be finally confirmed.

Below, we illustrate an example of this estimation process. Say the probability of m^{th} Gaussian component being selected to generate the t^{th} observation is denoted by π_m. That is $f(z_{mt}=1)=\pi_m$. Log-likelihood function of the complete data (y, z) is

$$\begin{aligned} \ell_c(\varphi) &= \log p\left(y, z\,|\,\varphi\right) \\ &= \log \prod\nolimits_{t=1}^{T} p(y_t, z_t\,|\,\varphi) \\ &= \log \prod\nolimits_{t=1}^{T} \prod\nolimits_{m=1}^{M} \left[p(y_t\,|\,z_{mt}=1;\varphi)p(z_{mt}=1)\right]^{z_{mt}} \\ &= \sum\nolimits_{t=1}^{T} \sum\nolimits_{m=1}^{M} [z_{mt} \log p(y_t\,|\,z_{mt}=1;\varphi) + z_{mt} \log \pi_m] \end{aligned} \tag{V.2}$$

where $\varphi=\left(\mu_m, \Sigma_m, \pi_m\right)$ denotes the parameter set of interest, $p(y_t\,|\,z_{mt}=1;\varphi)$ represents the likelihood function of m^{th} Gaussian component $N(\mu_m, \sigma_m)$.

We now take the expectation of (V.2)

$$\left\langle \ell_c(\varphi) \right\rangle = \sum\nolimits_{t=1}^{T} \sum\nolimits_{m=1}^{M} [\left\langle z_{mt} \right\rangle \log p(y_t\,|\,z_{mt}=1;\varphi) + \left\langle z_{mt} \right\rangle \log \pi_m] \tag{V.3}$$

and maximizing $\langle \ell_c(\varphi) \rangle$ with respect to φ, then **M-step** (Maximization) of EM algorithm can be formulized as

$$\begin{aligned} \frac{\partial \langle \ell_c(\varphi) \rangle}{\partial \mu_m} = 0 \Rightarrow \mu_m = \frac{\sum_{t=1}^{T} \langle z_{mt} \rangle y_t}{\sum_{t=1}^{T} \langle z_{mt} \rangle} \\ \frac{\partial \langle \ell_c(\varphi) \rangle}{\partial \Sigma_m} = 0 \Rightarrow \Sigma_m = \frac{\sum_{t=1}^{T} \langle z_{mt} \rangle (y_t - \mu_m)(y_t - \mu_m)^T}{\sum_{t=1}^{T} \langle z_{mt} \rangle} \end{aligned} \tag{V.4}$$

Here, note that in (V.3) while the weight parameter is updated, usually it is beneficial to impose a Lagrange multiplier to the target derivative function $\partial \langle \ell_c(\varphi) \rangle / \partial \pi_m$.

$$\frac{\partial \langle \ell_c(\varphi) \rangle}{\partial \pi_m} - \lambda = 0 \Rightarrow \pi_m = \frac{\sum_{t=1}^{T} \langle z_{mt} \rangle}{\lambda} \tag{V.5}$$

This is because π_m is now a probability measure that needs to satisfy $\sum \pi_m = 1$. Besides we can also rewrite (V.5) to

$$\pi_m = \frac{\sum_{t=1}^{T} \langle z_{mt} \rangle}{T} \tag{V.6}$$

, since $\lambda = \sum_{t=1}^{T} \sum_{m=1}^{M} \langle z_{mt} \rangle = T$.

Once the expected complete log-likelihood function $\langle \ell_c(\varphi) \rangle$ has been maximized and elements in φ have all been updated, to ensure the incomplete log-likelihood is also maximized, each of the expectation of the latent variable $\langle z_{mt} \rangle$ then needs to be computed. And this step is just the **E-step** of EM algorithm. That is

$$\begin{aligned} \langle z_{mt} \rangle &= p(z_{mt} = 1 \mid y_t; \varphi) \\ &= \frac{p(y_t \mid z_{mt} = 1; \varphi) f(z_{mt} = 1)}{\sum_{m=1}^{M} p(y_t \mid z_{mt} = 1; \varphi) f(z_{mt} = 1)} \\ &= \frac{p(y_t \mid z_{mt} = 1; \varphi) \pi_m}{\sum_{m=1}^{M} p(y_t \mid z_{mt} = 1; \varphi) \pi_m} \end{aligned} \tag{V.7}$$

And the whole iterations will just alternate between these E-s and M-s until the convergence of MLE is finally proved.

To provide a more straightforward illustration of above estimation procedure, we now present the pseudo-code of implementing EM algorithm in a Gaussian mixture model.

1. Initialisation: $\pi_m, \mu_m, \Sigma_m, \langle z_{tm} \rangle$

```
2. E-steps:
            for t=1 to T
              for m=1 to M; Calculate
                $p(y_t \mid z_{mt} = 1, \varphi) = (2\pi)^{-d/2} |\Sigma_m|^{-1/2} \exp\left\{-(y_t - \mu_m)^T \Sigma_m^{-1} (y_t - \mu_m)/2\right\}$
                      $\langle z_{mt} \rangle = \frac{p(y_t \mid z_{mt} = 1; \varphi)\pi_m}{\sum_{m=1}^{M} p(y_t \mid z_{mt} = 1; \varphi)\pi_m}$
                end
                    end
3. M-steps:
            for m=1 to M; Calculate
                    $\pi_m = \frac{\sum_{t=1}^{T} \langle z_{mt} \rangle}{T}$
                    $\mu_m = \frac{\sum_{t=1}^{T} \langle z_{mt} \rangle y_t}{\sum_{t=1}^{T} \langle z_{mt} \rangle}$
                    $\Sigma_m = \frac{\sum_{t=1}^{T} \langle z_{mt} \rangle (y_t - \mu_m)(y_t - \mu_m)^T}{\sum_{t=1}^{T} \langle z_{mt} \rangle}$
            end
4. Convergence
```

Appendix V. Definition and Statistical properties of Markov Chains

Definition of Markov chains

We define a Markov Chain starting from the concept of a stochastic process. A stochastic process, say $\{\varphi^{(m)}\}$, is a consecutive set of random quantities defined on some given state space Θ and indexed so that the order is known. Here, the state space refers to the range of possible values for φ; it could be either discrete or continuous depending on how the variable of interest is measured. Given this definition, $\{\varphi^{(m)}\}$ is then said to be a Markov Chain if its sampling sequence on state space Θ satisfies the condition, $E(\varphi^{(m+1)} \mid \varphi^{(1)}, \varphi^{(2)}, ..., \varphi^{(m)}) = E(\varphi^{(m+1)} \mid \varphi^{(m)}),$ for all $m \geq 0$. That is, the conditional expectation of $\varphi^{(m+1)}$ only depends on the preceding value $\varphi^{(m)}$ and independent of all earlier information. Thus, the current state is the only information source that determines the nature of the next, and all earlier memory will be forgotten. This characteristic is called the 'local property' of Markov Chain. It turns out to be enormously useful when generating samples from the limiting distributions of interest because when the chain eventually finds the region of the state space with the highest density, it will only produce a sample that is mildly dependent on the value of the last state.

Before proceeding, it is also important to know two basic elements for constructing Markov Chains. That is the initial probability P_0 and the transition probability T. P_0 tells us how Markov chain starts; T maps the potential transition events to the probability of occurrence. If the state space is discrete, usually a Markov chain will present a matrix mapping, however while the state space becomes continuous, T is then usually represented by a probability density function (*p.d.f*). For a more detailed illustration on this issue, see for example Doob (1990) and Gamerman (1997).

Statistical properties

Markov Chain can show many different characteristics depending on the way it is constructed. However, basically there are only few that are central to the Bayesian statistics. For example, a major theorem of Markov Chain called *convergence theorem* states that, *"...under certain regularity conditions, an irreducible, aperiodic and positive recurrent homogeneous chain will converge to a limiting probability distribution after the initial burn-in period is thrown away..."* In practice, it means if one looks at the values generated by a given chain sufficiently far from its simulation origin, the successive values will be distributed with stable frequencies stemming from a stationary probability distribution. This is one of the most important results of this stochastic process, and it explained why MCMC algorithms are feasible for inference calculation. Since in Bayesian statistics the major task is to simulate such a sampling sequence that can converge to the posterior density, it then equals to know how to generate this chain so that it has the properties of irreducibility, aperiodicity, positive recurrence and homogeneity. Concerning this task, we present in the following a brief description of these statistical properties.

a. Homogeneity

A Markov Chain is said to be *homogeneous* or *stationary* if at any step m, the transition probability T does not depends on the value of m, or by a similar putting, T does not evolve with time. Therefore, given the initial distribution P_0, the state of a homogenous chain after m-step is

$$P_m = P_0 T^m \tag{VI.1}$$

b. Irreducibility

Irreducible, loosely speaking, is the property that any state of a Markov chain can be reached from all other states. To make this illustration more clearly, consider a discrete Markov chain $\{\varphi^{(m)}\}$ with finite state space $S=\{s_1,...s_k\}$. We say a state s_i will communicate with another state s_j, $s_i \rightarrow s_j$, if the chain has positive probability of ever reaching s_j from s_i. And these two states are said to be intercommunicating, $s_i \longleftrightarrow s_j$, if the conditions of $s_i \rightarrow s_j$ and $s_j \rightarrow s_i$ are both satisfied. Given these definitions, an irreducible Markov chain can be defined if for all s_i, $s_j \in S$, we have $s_i \longleftrightarrow s_j$, or we can find an n such that $(T^n)_{i,\, j}>0$ where $T_{i,\, j}$ denotes the transition probability from state i to state j.

c. Aperiodicity

Now, we move onto illustrating the *aperiodicity* property of Markov chains. And we start from defining the *period* of a state. First, for a finite or infinite set $\{a_1, a_2, \ldots\}$ of positive integers, we write $gcd\{a_1, a_2, \ldots\}$ as the greatest common divisor of a_1, a_2, …. The *period*, $d(s_i)$, of a state $s_i \in S$ is then defined as the length of time to repeat an identical cycle of chain values. That is,

$$d(s_i) = \gcd\{n \geq 1 : (T^n)_{i,j} > 0\} \tag{VI.2}$$

For example, if we now start from s_i, $d(s_i)$ is then the greatest common divisor of the set of times that the chain can return (i.e., has positive probability of returning) to s_i,. If $d(s_i)$ equals to one, then we can say the state s_i is aperiodic. And the whole Markov chain is *aperiodic*, if all its states are aperiodic.

d. Positive recurrence

Apart from the above characteristics, recurrence is also an important property of Markov Chain. It has a close relationship to irreducibility. And the linkage between these two concepts is important for defining a subspace that captures the Markov Chain and simultaneously assures this Markov chain will explore the entire subspace. An irreducible Markov Chain is said to be *recurrent* with respect to a given state *A* which is a single point or a defined collection of points, if the probability that the chain occupies *A* infinitely often over unbounded time is nonzero. And a Markov chain is said to be *positive recurrent*, if the average time to return to *A* is bounded.

e. Markov Chain Convergence Theorem

Given a Markov chain which possesses all statistical properties described above including homogenous, irreducible, aperiodic and recurrent, an important theorem frequently referred to as the '*Existence of stationary distribution*' states that there always exists for this chain at least one stationary distribution that over the states *S* will persist forever once it is reached. Formally, this *stationary distribution* (also called *invariant distribution*, *equilibrium distribution* or *limiting distribution*), say $\pi = (\pi_1, \cdots \pi_k)$ can be identified, if for a Markov chain, it satisfies the conditions

(i) $\pi_i \geq 0$ for *i=1,...,k*, and $\sum_{i=1}^{k} \pi_i = 1$, and

(ii) $\pi = \pi \cdot T$, meaning that $\sum_{i=1}^{k} \pi_i T_{i,j} = \pi_j$ for *j=1,...,k*.

Besides, another part of this theorem called '*Uniqueness of the stationary distribution*' states that for any irreducible, aperiodic and homogeneous Markov chain, it will converges to *one* and *only one* stationary distribution. Thus, in Bayesian statistics once this stationary distribution is obtained, it will correspond to only posterior density of interest.

f. Erogdicity

Now, it is necessary to introduce a new concept which can encompasses all statistical properties illustrated above. That is *ergodicity*. Formally, we say a Markov Chain is *ergodic* if this chain have all properties of irreducibility, aperiodicity, positive recurrence and homogeneity, and

$$\lim_{n \to \infty} (T^n)_{i,j} = \pi_j \qquad \text{(VI.3)}$$

for all s_i and s_j in *S.* Since an ergodic Markov Chain can now fulfill all conditions mentioned in '*Markov Chain convergence theorem*', it is easy to find one and only one stationary distribution for its sampling sequence. And in Bayesian statistics, if a specific chain is found to have reached its ergoic state, then we say it will behave as a pseudo sample from the posterior density.

Since it is already known the state of a chain at time *m* will be nearly independent of the state at time *n* if $m >> n$, different states in this chain although by their very definition are serial dependent; their empirical moments can be used to approximate the distributional characteristics of the density of interest. For example, suppose now we have a sampling sequence $\{\varphi^{(m)}\}$ with *M* simulated values and an arbitrary burn-in period (to eliminate the effect from P_0) with length of *N*, the conditional mean of this chain

$$E(\hat{\varphi}) \approx \frac{1}{M-N}\sum_{N+1}^{M}\varphi^{(m)} \tag{VI.4}$$

then can be used to approximate the true parameter value $\hat{\varphi}$.

g. Reversibility

Apart from the erogidicty, another important property of Markov chain is also worth noting here although it is not a necessary condition for chains to converge. Countless researchers found that Markov chain simulated by applying a MCMC algorithm is usually reversible to the state where it is generated from. Concretely, a probability distribution π on *S* is said to be *reversible* for the chain (or for the transition matrix T) if for all $i, j \in \{1,...,k\}$ we have

$$\pi_i T_{i,j} = \pi_j T_{j,i} \tag{VI.5}$$

And a Markov Chain is said to be reversible if there exists a reversible distribution for it. Here, although this property is not a necessary condition for convergence, it can be imposed as a restrictive condition when simulating chains. This is because in most nontrivial situations, the easiest way to construct a chain with a given stationary distribution π is just to make sure this reversibility condition holds (See Robert and Casella 1999 p. 235 for the proof).

Appendix VI. How Gibbs sampler is related to Metropolis Hastings algorithm (MH)

A lot of statistics textbooks have referred the Gibbs sampler as a special case of MH algorithm. To obtain a practical view of how these two MCMC techniques are closely related to each other, we provide in the following the proof. For a more detailed illustration, see Robert and Casella (1999).

To prove Gibbs sampler is a special case of MH algorithm, first it is necessary to start from defining the jumping density of MH algorithm equivalent to the full conditional of Gibbs sampler so that $q(\varphi_{(k)}\,|\,\varphi_{-(k)}) = p(\varphi_{(k)}\,|\,\varphi_{-(k)})$. Then, by letting $\varphi^{(m-1)} = (\varphi_1^{(m-1)}, \varphi_2^{(m-1)}, ..., \varphi_K^{(m-1)})$ be the current state and $\varphi^{*} = (\varphi_1^{(m)}, \varphi_2^{(m-1)}, ..., \varphi_K^{(m-1)})$ be a candidate value for m^{th} simulated value of φ, the acceptance probability $D(\cdot)$ of φ^{*} in MH algorithm can be calculated by,

$$
\begin{aligned}
D(\cdot) &= \frac{p(\varphi^{*})/q\left(\varphi^{*}\,|\,\varphi^{(m-1)}\right)}{p(\varphi^{(m-1)})/q\left(\varphi^{(m-1)}\,|\,\varphi^{*}\right)} \\
&= \frac{p(\varphi_1^{(m)}, \varphi_2^{(m-1)}, ..., \varphi_K^{(m-1)})/p\left(\varphi_1^{(m)}\,|\,\varphi_2^{(m-1)}, ..., \varphi_K^{(m-1)}\right)}{p(\varphi_1^{(m-1)}, \varphi_2^{(m-1)}, ..., \varphi_K^{(m-1)})/p\left(\varphi_1^{(m-1)}\,|\,\varphi_2^{(m-1)}, ..., \varphi_K^{(m-1)}\right)} \\
&= \frac{p(\varphi_1^{(m)}, \varphi_2^{(m-1)}, ..., \varphi_K^{(m-1)}) * p\left(\varphi_1^{(m-1)}, \varphi_2^{(m-1)}, ..., \varphi_K^{(m-1)}\right) * p(\varphi_2^{(m-1)}, ..., \varphi_K^{(m-1)})}{p(\varphi_1^{(m-1)}, \varphi_2^{(m-1)}, ..., \varphi_K^{(m-1)}) * p\left(\varphi_1^{(m)}, \varphi_2^{(m-1)}, ..., \varphi_K^{(m-1)}\right) * p(\varphi_2^{(m-1)}, ..., \varphi_K^{(m-1)})} \\
&= 1
\end{aligned}
$$

Here, note that whenever the full conditionals of Gibbs sampler are set equal to the jumping densities of MH, $D(\cdot)$ is always equal to one. Thus, every candidate values drawn from the jumping density will be accepted for sure in MH algorithm, and no rejection will occur. Thus, we can say Gibbs sampler is a special case of MH algorithm.

Appendix VII. BIS's Triennial Central Bank Survey on trading volume of four major currencies

This appendix reports the daily trading volume of four major currencies *USD EUR GBP* and *JPY*. Data is collected from Bank of International Settlement's (BIS) annual survey published in Apr. 2004. And three panels are presented below to show the amounts and shares of how these currencies are traded in both spot and OTC market. First, *Panel A* presents the currency distribution of reported foreign exchange market turnover. Then, given the cross pairs of above currencies, *Panel B* reports the daily turnover of these pairs. Finally, *Panel C* gives the daily trading volume of OTC derivatives traded on these pairs. Note that the data below are all documented in quantity of billions of US dollar.

***Panel A.* Currency distribution of reported foreign exchange market turnover**

Currency	1989	1992	1995	1998	2001	2004
USD	90	82	83.3	87.3	90.3	88.7
EUR	-	-	-	-	37.6	37.2
JPY	27	23.4	24.1	20.2	22.7	20.3
GBP	15	13.6	9.4	11	13.2	16.9

***Panel B.* Foreign exchange turnover by currency pairs**

Currency pairs	1992		1995		1998		2001		2004	
	Vol.	%.	Vol.	%.	Vol.	%.	Vol.	%.	Vol.	%.
USD/EUR	-	-	-	-	-	-	354	30	501	28
USD/JPY	155	20	242	21	256	18	231	20	296	17
USD/GBP	77	10	78	7	117	8	125	11	245	14
EUR/JPY	-	-	-	-	-	-	30	3	51	3
EUR/GBP	-	-	-	-	-	-	24	2	43	2

***Panel C.* OTC foreign exchange derivatives turnover by currency pairs**

Currency pairs	Total				Currency options			
	1995	1998	2001	2004	1995	1998	2001	2004
USD vs. others	34	77	54	110	31	68	48	92
EUR	-	-	17	38	-	-	16	31
JPY	14	36	19	30	13	33	17	27
GBP	3	5	4	12	3	4	3	9
Euro vs. others	-	-	10	23	-	-	9	20
JPY	-	-	6	10	-	-	6	10
GBP	-	-	2	4	-	-	2	3

Bibliography

Aitkin, M., and I. Aitkin (1996), "A Hybrid EM/Gauss-Newton Algorithm for Maximum likelihood in mixture distributions" *Statistics and Computing*: 6, 127-130.

Aitkin, M., and D., B. Rubin (1985), "Estimation and hypothesis testing in finite mixture models" *Journal of Royal Statistics Society*: B47, 67-75.

Alizadeh, S., M. Brandt and F. Diebold (2002), "Range based estimation stochastic volatility models" *Journal of Finance*: 57, 1047-1091.

Akaike, H. (1973), "Information theory and an extension of the maximum likelihood principle", In Second International Symposium on Information Theory, B. N. Petrov and F. Csaki (eds.). Budapest: Akadémiai Kiadó, pp. 267-281. (Reproduced in 1992 in Breakthroughs in Statistics 1, S. Kotz and N. L. Johnson (eds.). *New York: Springer- Verlag*, 610-624.)

Amin, Ki. and V.K. Ng (1997), "Inferring future volatility from the information in implied volatility in Eurodollar options: a new approach", *Review of Financial Studies*: 10, 333-367.

Andersen, T.G., Bollerslev, T., Diebold, F.X., and Labys, P. (2000), "Exchange rate returns standardized by realised volatility are (nearly) Gaussian," *Multinational Finance Journal*: 4, 159-179.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and P.F., Christoffersen (2005), "Practical Volatility and correlation modeling financial market risk management," *Risk of Financial Institutions*, NBER, 53-548.

Ausín, M. and P. Galeano (2005), "Bayesian Estimation of the Gaussian Mixture GARCH model", working paper 05-36, *University of Carlos de Madrid*.

Azzalini, A. (1985), "A class of distributions which includes the normal ones", *Scandinavian Journal of Statistics*: 12, 171-178.

Azzalini, A., and A. Capitanio (1996), "The multivariate skew normal distribution", *Biometrika:* 83, 715-726.

Azzalini, A., and A. Capitanio (2003), "Distribution generated by perturbation of symmetry with emphasis on a multivariate skew t distribution", *Journal of royal statistics society:* B65, 367-389.

Bai, X., Rusell, J. R., Tiao, G. C. (2003), "Kurtosis of GARCH and stochastic volatility models with non-normal innovations", *Journal of Econometrics:* 114, 349-360.

Ball, C. A. and W. N. Torous (1983), "A simplified Jump Process for common stock returns", *Journal of financial and quantitative analysis*: 18, 53-65.

Bank of International Settlements (2004), "Central bank survey of foreign exchange and derivatives market activity" *International Bank of Settlements*: Basel.

Barndorff-Nielsen, O. (1965), "Identifiablility of Mixtures of Exponential families", *Journal of Mathematical Analysis and Applications:* 12, 115-121.

Barndorff-Nielsen, O. (1977), "Exponentially decreasing distributions for the logarithm of particle size", *Proceedings of the Royal Society London* A: 353, 401-419.

Barndorff-Nielsen, O. and Shephard, N. (2001), "Normal modified stable processes", *Theory of Probability and Mathematics Statistics*: 65, 1–19.

Basle Committee on Banking Supervision (1998), "*International Convergence of Capital Measurement and Capital standard*" July.

Baur, D. (2003), "A flexible dynamic correlation model", European commission, Joint research center: *Ispra*.

Baum, L. E. and Petrie, T (1966), "Statistical Inference for probabilistic functions of finite Markov chains", *Annals of Mathematical Statistics*: 37, 1554-1563.

Bauwens, L., and Lubrano, M (1998), "Bayesian inference on GARCH models using Gibbs sampler", *Journal of Econometrics*: 1, 23-46.

Bauwens, L., and Laurent, S (2002), "A new class of multivariate skew densities, with application to GARCH models", *Journal of Business and Economic Statistics*.

Bauwens, L., Laurent S., Rombouts J. V. K. (2004), "Multivariate GARCH models: a survey", *CORE DP*: 2003/31.

Bauwens, L., Hafner C., and Romboust J.V.K., (2006), "Multivariate Mixed Conditional Heteroskedasticity", *Computational Statistics and Data Analysis*.

Bayes, T. (1763), "An Essay Towards Solving a Problem in the Doctrine of Chances", *Philosophical Transactions of the Royal Society of London*: 53, 370-418.

Beale, E, M, L. and Little, R. J. A. (1975), "Missing values in multivariate analysis", *Journal of Royal Statistics Society*: B37, 12-146.

Beckers, S. (1981), "A Note on Estimating the Parameters of the Diffusion-Jump model of Stock returns", *Journal of financial and quantitative analysis*: 16, 127-140.

Benhamou, E. (2000), "Option Pricing with Lévy process", *LSE*

Berger, J. O. (1985), *Statistical Decision Theory and Bayesian analysis*, 2nd Ed. New York: Springer-Verlag.

Berger, J. O. (1986), "*Bayesian Salesmanship*", In Bayesian inference and Decision Techniques with applications: Essays in Honor of Bruno de Finetti, Arnold Zellner (ed.). Amsterdam: North Holland: 473-488.

Berkowitz, J. and J. O'Brien (2002), "How accurate are Value at risk models at commercial banks?" *Journal of Finance:* 57, 1093-1112.

Berndt, E., Hall, B., Hall, R and J. Hausman (1974),"Estimation and Inference in Nonlinear Structural Models," *Annals of Economic and Social Measurement:* 4, 653–65.

Besag, J. (1975), "Statistical analysis of non-lattice data", *The Statistician*: 24, 179-195.

Besag, J., Green, P. J. Higdon, D. M., and Mengersen, K. L. (1995), "Bayesian computation and stochastic systems (with discussion)", *Statistical Science:* 10 , 2-66.

Bevan B., Poon, S.H., and Taylor, S. (1999), "Forecasting S&P 100 volatility: The Incremental Information Content of Implied Volatility and High Frequency index returns", *Lancaster University Management School*, Working Paper 99/014.

Bibby, B. M., and Sørensen, M., (2003) "*Hyperbolic Processes in Finance*": In Rachev, S. (Ed.) Handbook of Heavy Tailed Distributions in Finance. Elsevier Science.

Billio, M., Caporin, M. and Gobbo, M. (2006), "Flexible Dynamic Conditional Correlation multivariate GARCH models for asset allocation", *Applied Financial Economics Letters*, vol. 2(2), pages 123-130, March.

Billio M. and Caporin, M (2007), "A generalised Dynamic Conditional Correlation model for portfolio risk evaluation", forthcoming Mathematics and Computers in Simulation.

Black, F. and Scholes, M. (1973), “The pricing of options and corporate liabilities”, *Journal of Political Economy*, Jun: 637-654.

Branco, M., and D. Dey (2001), “A general class of multivariate elliptical distributions”, *Journal of Multivariate Analysis* 79: 99-113.

Brannas, K., and N. Nordman (2001), “An Alternative Conditional Asymmetry Specification for Stock Returns," *Economic Studies*.

Brenner, M., and D. Galiai (1993), “Hedging Volatility in foreign currencies”, *Journal of derivatives*, 1993, 53-9.

Brooks, S. P. and A. Gelman (1998a), “General Methods for Monitoring Convergence of Iterative simulations”, *Journal of Computational and Graphical Statistics* 7: 434-455.

Brooks, C. and J. Chong (2001), “The cross-currency hedging performance of implied versus statistical forecasting models”, *Journal of Futures Markets* 21: 1043-1069.

Brooks, C. and G. Persand (2003), “Volatility forecasting for risk management”, *Journal of Forecasting* 22: 1-22.

Brooks, S. P. and G. O. Roberts (1998b), “Convergence Assessment Techniques for Markov Chain Monte Carlo”, *Statistics and Computing* 8: 319-335.

Brown, B. W. and S. Maital (1981), “What do economists know? An empirical study of experts expectations” *Econometrica* 49: pp. 491–504.

Böhning, D. (1999), C.A.MAN-Computer Assisted Analysis of Mixtures and Applications, Chapman & Hall, London.

Bodurtha, J.N. and Shen, Q., (1995), “Historical and Implied Measures of Value at Risk: The DM and Yen Case,” Manuscript, *Georgetown University.*

Bollerslev T. (1986), “Generalized autoregressive conditional heteroskedasticity”, *Journal of Econometrics* 31: 307–327.

Bollerslev T. (1990), “Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model”, *Review of Economics and Statistics* 72: 498–505.

Bollerslev, T., Engle, R., and Wooldridge, J.M. (1988), “A capital asset pricing model with time varying covariances”, *Journal of Political Economy* 96: 116–131.

Bones, A., Chen, A. and S. Jatusipitak (1974), “Investigation of Nonstationary Prices”, *Journal of Business* 47: 518-537.

Bookstaber, R., and J. MacDonald (1987), “A general distribution for describing security price returns”, *Journal of Business* 60: 401-424.

Box, G. E. P. and D. R. Cox (1964), “An analysis of transformations”, *Journal of Royal Statistical Society* B 26: 211-252.

Box, G. E. P. and TIAO, G. C. (1973), “*Bayesian Inference in Statistical Analysis*”, Reading, Mass: Addison-Wesley.

Buckley, I., Comezaña, G., Djerroud, B. and L. Seco (2002), “Portfolio optimisation for alternative investments”, *Imperial College*, London.

Bye, B. V. and E. S. Schechter (1986), “A latent Markov model approach to the estimation of response errors in multiwave panel data”, *Journal of American Statistical Association* 81: 375-380.

Campbell, B. and Dufour, J.M., (1994), “*Excat Nonparametric Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter*”, Universite de Montreal, Departement de sciences economiques, Cahiers de recherche 9407.

Campa, J.M. and Chang, K.P.H. (1998), “The forecasting ability of correlations implied in foreign exchange options”, *Journal of International Money and Finance* 17: 855-880.

Cappiello, L., Engle, R. and K. Sheppard (2004), “Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns”, *ECB working paper*: No. 204.

Cappuccio, N., Lubian, D. and D. Raggi (2004), "MCMC Bayesian Estimationof a Skew-GED stochastic volatility model”, *Studies in Nonlinear Dynamics & Econometrics*, 8(2): pp 1211-1211.

Cajigas, P. J. and Urga, G. (2005), “Dynamic Conditional Correlation models with Asymmetric multivariate Laplace innovations”, *Cass Business School*, London.

Carlin, B. P. and S. Chib (1995), “Bayesian Model Choice via Markov Chain Monte Carlo Methods”, *Journal of the Royal Statistical Society Series* B 57: 473-484.

Carlin, B. P. and A. E. Gelfand (1991), “An iterative Monte Carlo method for nonconjugate Bayesian analysis”, *Statistics Computing* 1: 119-128.

Carlin, B. P. and Louis, T. A. (2001), *Bayes and Empirical Bayes Methods for Data analysis*, 2nd ed. New York: Chapman & Hall.

Carlin, B. P., Polson, N. G. and D. S. Stoffer (1992), “A Monte Carlo approach to non-normal and nonlinear state-space modeling”, *Journal of the American Statistical Association* 87(418): 493-500.

Casarin, R. (2003), “*Bayesian inference for Mixture of Stable Distribution*”, PhD thesis, CEREMADE, University of Paris IX.

Casella, G. and R. L. Berger (2002), *Statistical Inference,* 2nd ed. Pacific Grove, CA: Duxbury.

Casella, G. and George, E. I. (1992), “Explaining the Gibbs Sampler”, *American Statistician* 46: 167-174.

Charlier, C. (1906), “Researches into the theory of probability”, *Acta Univ. Lund* (NeueFolge. Abt. 2) 1: 33-38.

Chatfield, C. (1996), *The Analysis of Time series*, 5th ed, New York: Chapman & Hall.

Chib, S. (1995), “Marginal Likelihood from the Gibbs output”, *Journal of American Statistical Association* 90: 1313-1321.

Chib, S. (1996), “Calculating Posterior Distributions and Model Estimates in Markov Mixture models”, *Journal of Econometrics* 95: 79-97.

Chib, S. and Greenberg, E. (1995), “Understanding the Metropolis-Hasting Algorithm”, *American Statistician* 49: 327-335.

Chib, S., Nardari, F. and Shephard, N. (2002), “Analysis of high dimensional multivariate stochastic volatility models”, *working paper, Nuffield college, Oxford*.

Choi, K. and W. G. Bulgren (1968), “An estimation procedure for mixtures of distributions”, *Journal of Royal Statistics Society* B 30: 444-60.

Christensen, B.J. and Prabhala, N.R. (1998), “The relation between implied and realized volatility”, *Journal of Financial Economics* 50: 125-150.

Clements, M.P. and Hendry, D.F. (1993), “On the limitations of comparing mean square forecast errors, with discussion”, *Journal of Forecasting*, Vol. 12, pp. 617-637.

Colacito, R. and R. Engle and E. Ghysels (2009), "A component model for dynamic correlations", Department of Finance, *Kenan-Flagler Business School*, University of North Carolina at Chapel Hill, Unpublished.

Cowles, M. K. and B. P. Carlin (1996), "Markov Chain Monte Carlo Convergence Diagnostics: A comparative Review", *Journal of American Statistical Association* 91: 883-904.

Cox, J.C. and S. A. Ross (1976), "The Valuation of Options for Alternative Stochastic Processes", *Journal of Financial Economics*, 145-166.

Dagum, P., Karp, R., Luby, M. and S. Ross (1995), "*An optimal algorithm for Monte Carlo estimation*", In Proceedings of the 36th IEEE Symposium on Foundations of Computer Science, pp 142-149, Portland, Oregon.

Damien, P., Wakefield, J. and S. Walker (1999), "Gibbs sampling for Bayesian non-conjugate and hierarchical models using auxiliary variable", *Journal of Royal statistics society* B 61: 331-344.

Davis, P. J. and Rabinowitz, P. (1975), *Methods of Numerical Integration*, 2nd ed. San Diego: Academic Press.

Day, N. E. (1969), "Estimating the components of a mixture of two normal distributions", *Biometrika* 56: 463-474.

Day, T.E. and Lewis, C.M. (1992), "Stock market volatility and the information content of stock index options", *Journal of Econometrics* 52: 267-287.

DeGroot, M. *Optimal Statistical Decisions*, McGraw-Hill, (1970).

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977), "Maximum Likelihood from incomplete data via the EM algorithm (with discussion)", *Journal of Royal Statistical Society* B 39: 1-38.

Dennis, P. and S. Mayhew (2002), "Risk-Neutral Skewness: Evidence from Stock options", *Journal of Financial and Quantitative Analysis* 37, 471-493.

Dennis, P., S. Mayhew and C. Stivers (2005), "Stock Returns, Implied Volatility Innovations, and the Asymmetric Volatility Phenomenon", *Journal of Financial and Quantitative Analysis*.

Dias, J. G. (2004), "*Finite Mixture Models –Review, Application and Computer-intensive Methods*", PhD thesis, University of Groningen, Netherlands.

Diebolt, J. L. and C. P. Robert (1994), "Estimation of Finite Mixture Distributions through Bayesian Sampling", *Biometrika* 56: 363-375.

Diebold, F., and J. Lopez (1996): "*Forecast evaluation and combination*," in Handbook of Statistics, ed. by G. G.S. Maddala,and C. Rao, pp. 214–268. Amsterdam: North Holland.

Ding, Z.X., and C.W. J. Granger and R. Engle (1993), "A long memory property of stock market returns and a new model", *Journal of empirical finance*, 83-106.

Doob, J. L. (1990), *Stochastic Processes*. New York: Wiley.

DuMouchel, W. (1973), "Stable Distribution in Statistical inference: 1. Symmetric Stable Distribution Compared to Other Symmetric Long-Tailed Distribution", *Journal of American Statistical Association* 68: 469-477.

Dupire, B. (1993), "Model Art", *Risk*, 1993: pp118-120.

Efron, B. (1979), "Bootstrap Methods: Another look at the Jack knife", *Annals of Statistics* 7: 1-26.

Embrechts, P., Lindskog, P. and A. McNeil (2001),"Modelling Dependence with Copular and Applications to Risk management", *working paper, ETH Zurich*.

Engle, R. (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation", *Econometrica* 50: 987–1007.

Engle, R. (2002), "Dynamic conditional correlation—a simple class of multivariate GARCH models", *Journal of Business and Economic Statistics* 20: 339–350.

Engle, R. and Bollerslev, T. (1986), "Modelling the persistence of conditional variances", *Econometric Reviews* 5: 1–50.

Engle, R and E. Ghysels and B. Sohn (2006), "On the economic sources of stock market volatility", *NYU and UNC* unpublished manuscript.

Engle, R. and Gonzalez-Rivera G. (1991), "Semiparametric ARCH model", *Journal of Business and Economic Statistics* 9: 345–360.

Engle, R. and G. Lee (1999), "*A permanent and transitory component model stock return volatility*", in ed. R. F. Engle and H. White, Co-integration, Causality, and Forecasting: A Festschrift in Honor of Clive W.J. Grander, (Oxford University Press), 475-497.

Engle, R. and Kroner, F. K. (1995), "Multivariate simultaneous generalized ARCH", *Econometric Theory* 11: 122–150.

Engle, R. and Sheppard K. (2001), "Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH", *Mimeo, UCSD.*

Eugenie, H. and S. J. Koopman (2000), "Forecasting the Variability of Stock Index Returns with Stochastic Volatility Models and Implied Volatility", *Tinbergen Institute Discussion Papers 00-104/4, Tinbergen Institute*.

Evans, M. (1991), "Chaining via annealing", *Annals of Statistics* 19: 382-393.

Everitt, B. S. and Merette, C. (1990), *Finite Mixture Distributions*, London: Chapman & Halls.

Everitt, B. S (1996), "An Introduction to the finite mixture distributions", *Statistical Methods in Medical Research* 5: 107-127.

Fabio Trojani & Francesco Audrino (2005), "A general multivariate threshold GARCH model with dynamic conditional correlations", University of St. Gallen Department of Economics working paper series 2005 2005-04, Department of Economics, *University of St. Gallen.*

Fama, E. F. (1965), "The Behaviour of Stock Market Prices", *Journal of Business* 38: 34-105.

Fernandez, C., Osiewalski, J. and Steel, M.F.J. (1995), "Modeling and Inference with v-Spherical Distributions", *Journal of the American Statistical Association:* 90, 1331-1340.

Fernandez, C., and M. Steel (1998), "On Bayesian modelling of fat tails and skewness", *Journal of American statistical association*: 93, 359-371.

Fisher, R. A. (1956), *Statistical Methods and Scientific Inference*: Edinburgh, Oliver and Boyd.

Fiorentini, G.., Calzolari, G. and Panattoni, L. (1996), "Analytic derivatives and the computation of GARCH estimate" *Journal of Applied Econometrics*: 11, 399–417.

Fleming, J. (1998), "The quality of market volatility forecasts implied by S&P 100 index option prices", *Journal of Empirical Finance* 5: 317-345.

Fleming, J. and B. Ostdiek and R. Whaley (1993), "Predicting stock market volatility: A new measure", *Duke university working paper*.

Ftse Global Market (2005), "CDOs give hedge funds a headache", *FTSE Global Market*, issue 8, Jul/Aug 2005: 74-77.

Furman, D. W. and B. G. Lindsay (1994), “Measuring the relative effectiveness of moment estimators as starting values in maximizing likelihoods”, *Computational Statistics and Data Analysis* 17: 493-507.

Gallant, A.R., C. T. Hsu, and G. E. Tauchen (1999), “Using daily range data to calibrate volatility diffusions and extract the forward integrated variance”, *Review of economics and statistics*, 81, 617-631.

Gamerman, D. (1997), *Markov Chain Monte Carlo*, New York: Chapman & Hall.

Gander, M. P and D. A. Stephens (2005), “Inference for stochastic volatility models driven by Lévy process”, Department of Mathematics, *Imperial college*, London.

Gelfand, A. E. and A. F. M. Smith (1990), “Sampling Based Approaches to Calculating Marginal Densities”, *Journal of American Statistical Association* 85: 398-409.

Gelman, A. (1996), “*Inference and Monitoring Convergence*”, in W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, eds., Markov Chain Monte Carlo in Practice. Chapman and Hall, London.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995), *Bayesian Data analysis*. London: Chapman and Hall.

Geman, S. and D. Geman (1984), “Stochastic Relaxation, Gibbs Distributions and Bayesian Restoration of Images”, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6: 721-741.

Gelman, A. and D. B. Rubin (1992), “Inference from Iterative Simulation Using Multiple Sequences”, *Statistical Science* 7: 457-472.

Gelman, A. and D.B. Rubin (1996), “Markov chain Monte Carlo methods in biostatistics”, *Statistical Methods in Medical Research* 5: 339—355.

Geweke, J. (1989a), “Bayesian Inference in Econometric Models Using Monte Carlo Integration”, *Econometrica* 57: 1317-1340

Geweke, J. (1989b), “Exact Predictive Densities in Linear Models with ARCH Distribution”, *Journal of Econometrics* 40: 63-86.

Geweke, J. (1992), “*Evaluating the Accuracy of Sampling Based approaches to the calculation of Posterior moments*”, in J. M. Bernardo et al., eds. Bayesian Statistics, Vol. 4. Oxford, UK: Clarendon Press.

Geweke, J. (1993), “Bayesian treatment of the independent student-t linear model”, *Journal of Applied Econometrics* 8: 19-40.

Geweke, J. (1999), “Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication” (with discussion and rejoinder), *Econometric Review* 18: 1-126

Geweke, J. (2005), *Contemporary Bayesian Econometrics and Statistics*, New Jersey: John Wiley & Sons, Inc.

Geyer, C. J. (1992), “Practical Markov Chain Monte Carlo”, *Statistical Science* 7: 473-511.

Geyer, C. J. and Thompson, E. A. (1992), “Constrained Monte Carlo maximum likelihood for dependent data (with discussion)”, *Journal of Royal Statistical Society*: B54, 657-699.

Gibson, M.S and Boyer, B.H. (1998), “Evaluating forecasts of correlation using option pricing”, *Journal of Derivatives*, winter: 18-38.

Gill, J. (2002), *Bayesian Methods for the Social and Behavioural Sciences, London*: Chapman & Hall/ CRC.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996), *Markov Chain Monte Carlo in Practice:* London, Chapman & Hall.

Glosten, L.R. and R. Jagannathan and D. E. Runkle (1993), "On the relation between expected value and the volatility of the nominal excess return on stock", *Journal of Finance*: 48, 1779-1801.

Gnedenko, B. V. and H. Fahim (1969), "On a Transfer Theorem", *Dok. Akad. Nauk SSSR*, 187:15-17.

Gonzalez-Rivera, G. (1996), "Smooth Transition GARCH models", working paper at department of Economics, *University of California, Riverside*.

Good, I. J. (1950), *Probability and the Weighting of Evidence*. London: Griffin.

Gould, P., and Bos, C. (2007), "Dynamic correlations and optimal hedge ratios", *Tinbergen Institute Discussion Paper*: TI 2007-025-4.

Gourieroux C. (1997), *ARCH Models and Financial Applications*, New York: Springer-Verlag.

Granger, C. W. J. and Z. Ding (1995a), "Some Properties of Absolute Returns and Alternative Measure of Risk", *Annales d' Economie et de Statistique* 40: 67-91.

Greene, W. H. (2003), *Econometric Analysis*, 5th ed. Upper Saddle River, NJ: Prentice-Hall.

Haas M., Mittnik S. and Paolella M.S. (2004), "Mixed normal conditional heteroskedasticity", *Journal of Financial Econometrics* 2: 211-250.

Haas, M., Mittnik, S., Paolella, M. S and Steude, S. C. (2005), "Stable Mixture GARCH models", *working paper*.

Hafner, C. and Franses, H. P. (2003), "A Generalized DCC model for Many Asset returns", Mimeo Econometric Institute, *Erasmus University Rotterdam.*

Hafner, C. and Rombouts, J. V. K. (2004), "Semiparametric multivariate volatility models", Econometric Institute Report 21. *Erasmus University Rotterdam.*

Häggström, O. (2003), *Finite Markov Chains and Algorithmic Applications*, 2nd ed., Cambridge: Cambridge University Press.

Hagerud, G.E. (1996), "A smooth transition GARCH models", *working paper Stockholm school of Economics.*

Hamilton, J. D. (1994), *Time series Analysis. Princeton*: Princeton University Press.

Hammersly, J. M. and D. C. Handscomb (1964), *Monte Carlo Methods*, London: Methuen & Co.

Hanson, F. B. and Zhu, Z. (2004), "Comparison of Market parameters for jump-diffusion distributions using multinomial maximum likelihood estimation", *working paper.*

Harris, R., Stoja, E., and J. Tucker (2004), "A Simplified Approach to Modeling the Co-movement of Asset Returns October", *Finance and Investment Working Paper, University of Exeter.*

Harvey, A. C., E. Ruiz and N. Shephard (1994), "Multivariate stochastic variance models", *Review of Economics Studies* 61(2), 247-264.

Harvey, C. R. and A. Siddique (1999), "Autoregressive Conditional Skewness", *Journal of Financial and Quantitative Analysis*, Cambridge University Press, vol. 34(04), pages 465-487, December.

Hasselblad, V. (1966), "Estimation of parameters for a mixture of normal distributions", *Technometrics* 8: 431-444.

Hastings, W. K. (1970), “Monte Carlo Sampling Methods using Markov Chains and their applications”, *Biometrika*: 57, 97-109.

Heidelberger, P. and Welch, P. D. (1983), “Simulation Run Length Control in the Presence of an Initial Transient”, *Operations Research:* 31, 1109-1144.

Hansen, B. (1994), “Autoregressive conditional density estimation”, *International economics reviews*, 35, 705-730.

Henson, F. B. and J. J. Westman (2002), “Stochastic analysis of Jump-diffusion for financial log-return processes”, *Proceedings of stochastic theory and control workshop*, Springer-Verlag, New York 1-15.

Heyde, C. C. and Kou, S. G. (2004), “On the controversy over tail weight of distributions”, *Operations Research Letters:* 32, 399-408.

Higdon, D. M. (1998), “Auxiliary variable methods for Markov chain Monte Carlo with applications”, *Journal of American Statistics Association* 93: 585-595.

Hsieh, D. (1989), “Modeling Heteroskedasticity in Daily Foreign Exchange Rates”, *Journal of Business and Economics Statistics:* 7, 307-317.

Huber, P. J. (1964), “Robust Estimation of a Location Parameter”, *Annals of Mathematical Statistics* 35: 73-101.

Jacquier, E., N. G. Polson and P. Rossi (1994), “Bayesian analysis of stochastic volatility models”, *Journal of Business & Economic Statistics*: 12, 371–417.

Jacquier, E., N. G. Polson and P. Rossi (1999), “Models and priors for multivariate stochastic volatility”, *working paper, CIRANO.*

Jeantheau, T. (1998), “Strong consistency of estimator for multivariate GARCH models”, *Econometrics theory:* 14, 70-86.

Jeffreys, S. H. (1961), *Theory of Probability*, 3rd ed. Oxford: Claredon Press.

Jennison, C. (1993), “Discussion on the Meeting on the Gibbs Sampler and Other Markov Chain Monte Carlo Methods,” *Journal of the Royal Statistical Society*: B55, 54–56.

Johnson, N. L., S. Kotz and N. Balakrishnan (1995), *Continuous Univariate Distributions* Vol. 1,2nd ed. New York: Wiley.

Johnson, N. L. and S. Kotz (1972), *Distributions in statistics: Continuous Multivariate Distributions*. New York: Wiley.

Jones, P. N. and G. J. Mclachlan (1990), “Laplace-Normal mixtures fitted to wind shear data”, *Journal of Applied statistics*: 17, 271-276.

Jones, M., and M. Faddy (2000), “A skew extension of the t distribution, with application”, mimeo, Department of statistics, *Open university, Walton Hall, UK.*

Jordan, M. I. And Xu, L. (1995), “Convergence results for the EM approach to mixture of experts architectures”, *Neural Networks*: 8, 1409-1431.

Jorion, P. (1995), “Predicting volatility in the foreign exchange market”, *Journal of Finance*: 50, 507-508.

Jose, A.L. (1999), “Methods for evaluating value-at-risk estimates”, *Economic Review, Federal Reserve Bank of San Francisco:* 3-17.

J.P. Morgan (1996)*, RiskMetrics™ – Technical Document* (Fourth edition), New York.

Karian, Z. E., Dudewicz, E. J. and P. McDonald (1996), “The extended Generalized Lambda distribution system for fitting distributions to data: history, completion of theory, tables, applications, the “final word” on moment fits”, *Communications in Statistics - Computation and Simulation* 25(3): 611–642.

Kendall, M. G. (1949), “On Reconciliation of the Theories of Probability”, *Biometrika*: 36, 101-116.

Kendall, M. and A. Stuart, *Handbook of Statistics*: Griffin & Company, London.

Keynes, J. M. (1921), *A Treatise on probability*: Macmillan, London.

Kenji, G. K. (1985), “A mixture for wind shear data”, *Journal of Applied Statistics* 12: 49-58.

Kim, S., Shephard, N. and S. Chib (1998). “Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models,” *Review of Economic Studies* 65: 361-93.

Kleibergen and Van Dijk, H. K. (1993), “Non-stationary in GARCH models: a Bayesian analysis” *Journal of Applied Econometrics* 8: 41-61.

Kloek, T. and van Dijk, H. K. (1978), “Bayesian Estimates of Equations System parameters; An application integration by Monte Carlo”, *Econometrica* 46: 1-19.

Knight, J.L. and Stachell, S.E. and Tran, K.C., (1995), “Statistical Modeling of Asymetric Risk in Asset Returns,” Papers 95-3, *Saskatchewan - Department of Economics*.

Komunjer, I. (2005), “Quasi-Maximum likelihood estimation for conditional quintiles”, forthcoming *Journal of Econometrics.*

Koop, G. (2003), *Bayesian Econometrics*. West Sussex, UK: Wiley.

Kosinski, A. (1999), “A procedure for the detection of multivariate outliers”, *Computational Statistics and Data Analysis* 29: 145-161.

Kotz, S., N. Balakrishnan, and N. L. Johnson (2001), *Continuous Multivariate Distributions*, Vol 1. New York: Wiley.

Kotz, S., Kozubowski, T. J., and K. Podgorski (2003), “An asymmetric multivariate Laplace distribution”, *working paper.*

Kroner, K.F., Kneafsey, K.P., and S. Claessens (1995), “Forecasting volatility in commodity markets”, *Journal of Forecasting*: 14, 77-95.

Kroner, K.F., and V.K. Ng (1998), “Modeling asymmetric co-movements of asset returns”, *Review of Financial Studies:* 11, 817-844.

Kuechler, U., K. Neumann, M. Sørensen and A. Streller. (1999), “Stock returns and hyperbolic distributions”, *Mathematical & Computer Modelling* 29: 1-15.

Kuester, K., Mittnik, S., and M. S. Paolella (2005), “Value–at–Risk Prediction: A Comparison of Alternative Strategies”, *Journal of Financial Econometrics*.

Labidi, C. and T. An (2000), “Revisiting the finite mixture of gaussian distributions with applications to futures markets”, *Computing in Economics and Finance* no. 67.

Lambert, P., and S. Laurent (2000), “Modeling skewness dynamics in series of financial data”, Discussion paper, institute de statistique, *Louvain-La-Neuve.*

Laplace, P. S. (1814), *Essai Philosophique sur les la Probabilities*. Paris: Ve Courcier.

Lawrence, C.T. and A.L. Tits (2001), “A Computationally Efficient Feasible Sequential Quadratic Programming Algorithm”, *SIAM J. Optimization*, 11 (4): 1092-1118.

Lehmann, E. (1975), *Nonparametric: Statistical Methods Based on Ranks*, Holden-Day, Inc., San Francisco.

Lee, S., and B. Hansen (1994), “Asymptotic properties of the maximum likelihood estimator and test of the stability of parameters of the GARCH and IGARCH models”, *Econometric theory*: 10, 29-52.

Lin, T. I., Lee, J. C. and H. F. Ni (2004), “Bayesian analysis of mixture modelling using the multivariate t distribution”, *Statistics and Computing* 2(14): 119-130.

Lin, T. I. (2009), “Maximum likelihood estimation for multivariate skew normal mixture models”, *Journal of multivariate analysis*: 100 (2009), 257-265.

Lindley, D. (1961), “The Use of Prior probability distributions in Statistical inference and decision”, *Proceedings of the Fourth Berkeley Symposium on Mathematical statistics and probability*: Berkeley: University of California Press, pp. 453-468.

Lindley, D. and A. F. M. Smith (1972), “Bayes Estimates for the Linear Model”, *Journal of the Royal Statistical Society*: B34, 1-41.

Lindsay, B. G. and Basak, P. (1993), “Multivariate Normal Mixtures: a fast, consistent method of moments”, *Journal of American Statistical Association* 88: 468-476.

Lindsay, B. G. and K. Roeder (1992), “Residual diagnostics for mixture models” *Journal of American Statistical Association:* 87, 785-794.

Liu, C. (1997), “Comment on ‘The EM algorithm - an old folk song sung to fast new tune’ Yu Meng and van Dyk” *Journal of Royal statistics Society*: B59, 557-558.

Liu, C., Liu, J. S., and Rubin, D. B. (1992), “A Variational Control Variable for assessing the convergence of the Gibbs sampler”, *Proceedings of the American Statistical Association*, *Statistical Computing Section*: 74-78.

Liu, C. and D. B. Rubin (1994), “The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence”, *Biometrika* 85: 633-648.

Liu, C., Rubin, D. B. and Wu, Y. N. (1998), “Parameter expansion to accelerate EM: the PX-EM algorithm”, *Biometrika* 85:755-770.

Liu, J., Wong, W. and A. Kong (1991a), “Correlation structure and the convergence of the Gibbs sampler, I”, Technical Report 299, *Dept. of Statistics, University of Chicago.*

Liu, J., Wong, W. and A. Kong (1991b), “Correlation structure and the convergence of the Gibbs sampler, II: Applications to various scans”, Technical Report 304, *Dept. of Statistics, University of Chicago*.

Liu, S. M. and B. W. Brorsen (1995), “Maximum Likelihood Estimation of a GARCH-stable Model”, *Journal of Applied Econometrics*, Vol. 10, No. 3, 1995, pp. 273-285.

Longin, F. and B. Solnik (2001), “Extreme correlation of international equity market”, *Journal of Finance:* 56, 649-676.

Lochos, V.H., V.G., Cancho and R. Aoki (2008), “Bayesian analysis of skew-t multivariate null intercept measurement error model” *Statistics papers*.

Macdonald, P. D. M. (1971), “Comment on a paper by Choi and Bulgren” *Journal of Royal Statistics Society*: B33, 326–329.

Merton, R. C. (1973), “Theory of Rational Option Pricing”, *Bell Journal of Economics and Management science*: 4, 141-183.

Merton, R. C. (1976), “Option Pricing when Underlying Stock Returns are Discontinuous” *Journal of Financial Economics*: 3, 125-147.

McCurdy, T., and I. Morgan (1988) “Testing the Martingale Hypothesis in Deutsche Mark Futures with Models Specifying the Form of the Heteroscedasticity,” *Journal of Applied Econometrics:* 3, 187-202.

Mcdonald, J. B. and Newey, W. K. (1988), “Partially Adaptive Estimation of Regression Models Via the Generalized t Distribution”, *Econometric Theory*: 4, 428-457.

Mandelbrot, B. (1963), “The Variation of Certain Speculative Prices” *Journal of Business*: 36, 394-419.

Markowitz, H. (1952), “Portfolio selection”, *Journal of Finance* 7: 71-77.

Marron, J, S. and Wand, M. P (1992), “Exact mean integrated squared error”, *Annals of Statistics* 20: 712-736.

McGinty, L. and Beinstein, E. (2004), “Credit correlation: A Guide”, *Credit Derivate Strategy*, JP-Morgan, London.

McLachlan, G. J., and K. E. Basford (1988), *Mixture models: Inference and Applications to Clustering*. NewYork: Marcel Dekker.

McLachlan, G. J., and P. N. Jones (1988), “Fitting mixture models to grouped and truncated data via the EM algorithm”, *Biometrics* 44: 571-578

McLachlan, G. and T. Krishnan (1997), *The EM algorithm and extensions*, John Wiley & Sons.

McLachlan, G. J. and D. Peel (1998), “*Robust cluster analysis via mixtures of multivariate t-distributions*”. In Lecture Notes in Computer Science Vol. 1451, A. Amin, D. Dori, P. Pudil, and H. Freeman (Eds.). Berlin: Springer-Verlag, pp. 658-666.

McLachlan, G. J., and D. Peel (2000), *Finite Mixture Models*, New York: John Wiley & Sons.

McNeil, A. J., Nyfeler, M and R. Frey (2001), “Copulars and Credit models”, *Risk* 111-114.

Mencia, F. J. and Sentana, E. (2004), “Estimation and testing of dynamic models with generalized hyperbolic innovations”, *CEMFI working paper*: No 0411.

Mendoza-Blanco, J. R and M. T. Xin (1997), “An Algorithm for Sampling the Degrees of Freedom in Bayesian Analysis of Linear Regressions with t-Distributed Errors”, *Applied Statistics* 46(3): 383-388.

Meng, X. L. and van Dyk, D. A. (1997), “The EM algorithm -an old folk song sung to a fast new tune” *Journal of Royal Statistics Society*. B 59: 511-567.

Meng, X. L. and D. B. Rubin (1991), “Using EM to obtain asymptotic variance-covariance matrices: the SEM algorithm”, *Journal of American Statistics Association* 86: 899-909.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953), “Equation of State Calculations by Fast Computing Machines”, *The Journal of Chemical Physics* 21: 1087-1092.

Metropolis, N. and Ulam, S. (1949), “The Monte Carlo Method”, *Journal of American Statistical Association* 44: 335-341.

Mincer,J.A. and Zarnowitz, V. (1969), “The evaluation of economic forecasts In: J.A. Mincer, Editor, Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance”, *NBER*, New York, pp. 3–46.

Mittnik, S. and Rachev, S. (1993a), “Modeling Asset Returns with Alternative Stable Models”, *Econometric Reviews* 12: 261-330.

Mittnik, S. and Rachev, S. (1993b), “Reply to Comments on ‘Modelling Asset Returns with Alternative Stable Models’ and some extensions” *Econometric Reviews* 12: 347-389.

Mykland, P., L. Tierney, and B. Yu (1995), “Regeneration in Markov-Chain Samplers” *Journal of American statistical Association* 90: 233-241.

Neal, R. M. (1993), “Comments on ‘A theoretical analysis of Monte Carlo algorithms for the simulation of Gibbs random field images’”, *IEEE Transactions on Information Theory*, 39: 310.

Neal, R. M. (1996) ``Sampling from multimodal distributions using tempered transitions'', *Statistics and Computing*, vol. 6, pp. 353-366

Nelson, D. B. (1990), “Stationarity and persistence in the GARCH (1,1) model”, *Econometric Theory* 6: 318-334.

Nelson, D. B. (1991), “Conditional Heteroskedasticity in Asset Returns: A New Approach”, *Econometrica*, 59: 347-370.

Nelson, R. (1999), *An introduction to Copulars*, Springer, New York.

Neuberger, A. (1990), “Volatility trading”, *London business school working paper.*

Newbold, P., Harvey, D.I., and L. Stephen (1999), “*Ranking Competing Multi-Step Forecasts”, Cointegration, Forecasting and Causality*, Oxford University Press, New York, Robert F. Engle and Halbert White, editors.

Newey, W. and K. West (1987), “A Simple Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix” *Econometrica:* 55, 703-708.

Newey, W., Engle, R and McFadden, D. (1994) "Large Sample Estimation and Hypothesis Testing," with (eds.) *Handbook of Econometrics*, vol. IV, 2111-2245, North Holland: Amsterdam, 1994.

Nielsen, S. F. (2000), “On simulated EM algorithms” *Journal of Econometrics:* 2, 267-292.

Orchard, T. and Woodbury, M. A. (1972), “A missing information principle: theory and application”, *Proceeding 6th Berkeley Symp Math statist*, 1: 697-715.

Ord, J. K., Snyder, R. D., Koehler, A. B., Hyndman, R.J. and Leeds, M. (2005), “Time series forecasting: The case for the single source of error state space approach” Working paper 7/05 *Monash Econometrics and Business Statistics*

Patton, A. J. and Sheppard, K (2008), "Evaluating Volatility and Correlation Forecasts," OFRC Working Papers Series, *Oxford Financial Research Centre*

Pearson, K. (1894), “Contributions to the theory of mathematical evolution”, *Philosophical Transaction of the Royal Society of London* A185: 71-110.

Pearson, K. (1895), “Contributions to the theory of mathematical evolution: skew variation”, Philosophical Transaction of the Royal Society of London A 186: 343-414.

Pearson, K. (1906), “Walter Frank Raphael Galton, 1860-1906: A Memoir”, Biometrika 5: 1-52.

Pelagatti, M.M and S. Rondena (2004), “Dynamic conditional correlation with elliptical distribution”: *University of Milan*.

Pelletier, D. (2004), “Regime Switching for Dynamic Correlations”, Econometric Society 2004 North American Summer Meetings, *Econometric Society*.

Peskun, P. H. (1973), “Optimum Monte-Carlo Sampling Using Markov chains”, *Biometrika* 60: 607-612.

Picone, D. (2005), “Pricing and rating CDOs of equity default swaps with NGARCH-M copular”, Default risk, *Cass Business school, London.*

Placket, R. L. (1966), "Current Trends in Statistical Inference", *Journal of Royal Statistical Society* A129: 249-267.

Poon, S. H. and C. Granger (2003), "Forecasting volatility in financial markets: a review", *Journal of Economic Literature*.

Poon, S. H. and M. Rockinger, and J. Tawn (2004), "Extreme value dependence in Financial markets: Diagnostics, Models and financial implications", *Review of financial studies*, 17, 581-610.

Quandt, R. E. and J. B. Ramsey (1978), "Estimating mixtures of normal distributions and switching regressions" *Journal of the American Statistical Association* 73: 730– 738.

Rachev, S. T. (2003), *Handbook of Heavy Tailed Distributions in Finance*, North-Holland.

Rachev, S., Kim, J. R. and S. Mittnik (1999), "Stable Paretian econometrics part I and II" *The Mathematical Scientist* 24: 24-55 and 113-127.

Rachev, S. and A. SenGupta (1993), "Laplace-Weibull mixtures for modeling price changes", *Management Science* 8(39): 1029-1038.

Raftery, A. E. and Lewis, S. M. (1992), "How Many Iterations in the Gibbs sampler?", in *Bayesian Statistics*, 4, J. M. Bernardo, A. F. M. Smith, A. P. Dawid, and J. O. Berger (eds.). Oxford: Oxford University Press, pp. 763-773.

Rao, C. R. (1948), "The Utilization of multiple measurements in problems of biological classification" *Journal of Royal Statistical Society* B10: 159-203.

Ritter, C. and M. A. Tanner (1992), "Facilitating the Gibbs sampler: the Gibbs Stopper and the Griddy Gibbs sampler", *Journal of American Statistical Association* 87: 861-868.

Robbins, H. (1948), "On the Asymptotic Distribution of the Sum of a Random Number of Random Variable", *Proc. Nat. Acad. Sci:* 34, 162-163.

Robert, C. P. (1995), "Convergence Control Methods for Markov Chain Monte Carlo Algorithms" *Statistical Science*: 10, 231-253.

Robert, C. P. (1998), *Discretization and MCMC Convergence Assessment: Lecture notes in Statistics:* 135. New York: Springer-Verlag.

Robert, C. P. and Casella, G. (1999), *Monte Carlo Statistics Methods*, New York: Springer-Verlag.

Roberts, G. O. (1994), "Method for estimating L2 convergence of Markov chain Monte Carlo", in *Bayesian Analysis in Statistics and Econometrics*: Essay in Honour of Arnold Zellner, D. Berry, K. Chaloner and J. Geweke (eds.). New York: John Wiley & Sons, pp. 373-384.

Roberts, G. O. and A. F. M. Smith (1994), "Simple Conditions for the Convergence of the Gibbs Sampler and Metropolis-Hastings Algorithms" *Stochastic Processes and Their applications* 49: 207-216.

Roberts, G. O. and Sahu, S. K. (1999), "On convergence of the EM algorithm and the Gibbs sampler" *Statistics and Computing* 9: 55-64.

Rocke, D. M. and Woodruff, D. L. (1997), "Robust estimation of multivariate location and shape" *Journal of Statistical Planning and Inference*: 57, 245-255.

Roeder, K (1994), "A graphical technique for determining the number of components in a mixture of normal", *Journal of American Statistical Association* 89: 487-495.

Roeder, K. and L. Wasserman (1997), "Practical density estimation using mixtures of normals" *Journal of American Statistical Association*: 89, 487-495.

Rubin, D. B. and Wu, Y. N. (1997), “Modelling schizophrenic behaviour using general mixture components”, *Biometrics* 53: 243-261.

Schervish, M. J., and B. P. Carlin (1990), “On the Convergence of Successive Substitution Sampling”, Technical Report 492, *Dept. of Statistics, Carnegie Mellon University*.

Schoutens, W. (2003), *Levy Processes in Finance: Pricing Financial Derivatives*, Wiley.

Schwarz, G. (1978), “Estimating the dimension of a model”, *Annals of Statistics* 6: 461-464.

Sentana, E. (1995), “Quadratic Arch models”, *Reviews of economic studies*, 62, 639-661.

Sepp, A. (2004), “Analytical valuation of lookback option in a double exponential jump-diffusion model”, *working paper*.

Seshadri, V. (1997): Halphen's laws. In S. Kotz, C. B. Read and D. L. Banks (eds.): *Encyclopedia of Statistical Sciences*, Update Volume 1, pp. 302 - 306. Wiley, New York.

Sheppard, K. (2002), “Understanding the Dynamics of Equity Covariance,” Manuscript, *UCSD*.

Siegel, A.F. (1997), “International currency relationship information revealed by cross-option prices”, *Journal of Futures Markets*: 17, 369-384.

Skintzi, V. and A. Refenes (2003), “Implied Correlation Index: A New Measure of Diversification,” *working paper, Athens University*.

Sklar, A. (1959), “Functions de répartition à dimensions et leurs marges”, *Publications de l’Institut de Statistique de l;Universite de Paris*, 8, 229-231.

Smith, A. F. M. and Roberts, G. O. (1993), “Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods (with discussion)” *Journal of Royal Statistical Society* B55: 2-23 and 53-102.

Subbotin, M. T. (1923), “On the law of frequency of errors”, *Mathematicheskii Sbornik* 31: pp. 296–301.

Tanner, M. A. (1996), Tools for Statistic Inference: Methods for the Exploration of Posterior distribution and likelihood function. New York: Springer.

Tanner, M. A. and G. C. Wei (1990), “A Monte Carlo implementation of the EM algorithm and the poor man’s data augmentation algorithms”. *JASA*: 85, 699–704.

Tanner, M. A. and W. H. Wong (1987), “The Calculation of Posterior Distributions by Data Augmentation” *Journal of American Statistical Association* 82: 528-550.

Theodossiou, P. (1998), “Financial Data and the Skewed Generalized *t* Distribution.” *Management Science*, 44 (12-1): 1650-1661.

Tierney, L. (1991), “Exploring Posterior Distributions Using Markov Chains”, in E. M. Keramidas, ed., *Computing Science and Statistics: Proceeding of the 23rd Symposium on the Interface*, pp. 563-570. Fairfax, VA: Interface Foundation of North America, Inc.

Tierney, L. (1994), “Markov Chains for Exploring Posterior Distributions” (with discussion and rejoinder) *Annals of Statistics:* 22 1701-1762.

Tierney, L. and A. Mira (1999), “Some adaptive Monte Carlo methods for Bayesian Inference”, *Statistics in Medicine*: 18, 2507–2515.

Titterington, D. M., Smith, A. F. M., and Markov, U. E. (1985), *Statistical Analysis of Finite Mixture Distributions*: New York: Wiley.

Tjelmeland, H. and B. K. Hegstad (2001), “Mode jumping proposal in MCMC” *Scandinavian Journal of Statistics* 28: 205-223.

Torben, A.G. Bollerslev, T. Diebold, F.X. and P. Labys (2001), “Modeling and forecasting realized volatility”, *NBER working paper*.

Tsay, R.S. (2002) *Analysis of Financial time series*: Wiley, New York.

Tse, A. and K.C. Tsui (2000), “A Multivariate GARCH Model with Time-Varying Correlations” *Economics Working Paper Archive EconWPA*.

Tukey, J. W. (1960), “A survey of sampling from contaminated distributions”, *In Contributions to Probability and Statistics*, I, Olkin, S. G. Ghurye, W. Hoeffding, W. G. Madow, and H. B. Mann (Eds.). Stanford: Stanford University Press, pp. 448-485.

Venkartaraman, S. (1997), “Value at Risk for a mixture of normal distributions: The use of quasi Bayesian estimation techniques”, *Economic Perspectives*.

Verdinelli, I., And Wasserman, L (1991), “Bayesian analysis of outlier problems using the Gibbs sampler” *Statistics and Computing*: 1, 105-117.

Vlaar, P. J. G and Palm, F. C. (1993), “The message in weekly exchange rates in the European Monetary System: mean reversion, conditional heteroskedasticity and jumps”, *Journal of Business and Economic Statistics* 11: 351–360.

Von Neumann, J. (1951), “Various Techniques used in connection with Random Digits: Monte Carlo Methods”, *U.S. National Bureau of Standards Applied Mathematics series* 12: 36-38.

Wald, A. (1949), “Note on the consistency of the maximum likelihood estimate”, *Ann. Math. Stat*. 20: 5955–601.

Walter, C. and J. Lopez (2000), “Is implied correlation worth calculating? Evidence from foreign exchange options”, *Journal of Derivatives*, pp 65-81.

Wang. K., S.K. Ng, and G.J. McLachlan (2009), “Multivariate skew *t* mixture models: Application to fluorescence-activated cell sorting data”, *Statistics papers*.

Whaley, R.E. (1993), “Derivatives on market volatility: hedging tools long overdue”, *Journal of Derivatives*, 1: 71–84.

Wilks, S. S. (1962), *Mathematical Statistics*, 2nd ed. New York: Wiley.

Wolfe, J. H. (1970), “Pattern clustering by multivariate mixture analysis” *Multivariate Behavioural Research* 5: 329-350.

Wong, M. A. (1985), “A bootstrap testing procedure for investigating the number of subpopulations” *Journal of Statistical Computation and Simulation*: 22, 99-112.

Wong, C. S. and Li, W. K. (2001), “On a Mixture Autoregressive Conditional Heteroskedastic Model” *Journal of American Statistical Association*: 96, 982-995.

Yu, B. and Mykland, P. (1997), “Looking at Markov Samplers through CUMSUM path plots: A simple diagnostic idea”, *Statistics and Computing* 8: 275-286.

Yu, J. and Meyer, R. (2006), “Multivariate Stochastic Volatility models: Bayesian estimation and model comparison”, *Econometric Reviews* 25(2-3), 361-384.

Zellner, A. Min, C-K. (1995), “Gibbs Sampler Convergence Criteria”, *Journal of American Statistical Association* 90: 921-927.

Zakoian, J.M (1994), “Threshold Heteroskedastic models”, *Journal of economic dynamics and control*, 18, 931-955.