The Estimation of Vector Multiplicative Error Model on Contaminated Data and Its Applications in Forecasting Volatilities

DING, Hao

A Thesis Submitted in Partial Fulfilment

of the Requirements for the Degree of

Doctor of Philosophy

in

Systems Engineering and Engineering Management

The Chinese University of Hong Kong

April 2013

Abstract of thesis entitled:

The Estimation of Vector Multiplicative Error Model on Contaminated Data and Its Applications in Forecasting Volatilities

Submitted by DING, Hao

for the degree of Doctor of Philosophy

at The Chinese University of Hong Kong in April 2013

This thesis studies the estimations of vector Multiplicative Error Model (MEM) under different kinds of model mismatches and its application in forecasting. In the first part of the thesis, two estimation methods, Maximum Likelihood (ML) method and Generalized Method of Moments (GMM), which have previously been used on vector MEM, are compared through different situations of data contaminations. From the comparison results it is found that both ML and GMM estimators are suspected to outliers in data. Therefore in the second part of the thesis a novel estimator is proposed: Weighted Empirical Likelihood (WEL) estimator.

It is shown to be more robust than ML and GMM estimators in simulations, and also in forecasting realized volatility and bipower volatility of S&P 500 stock index including the current financial crisis period. The forecast ability of vector MEM is further addressed in the third part of the thesis, where an alternative decomposition of realized volatility is proposed, and vector MEM is used to model and forecast the two components of realized volatility. From the realized volatility forecasts of S&P 500, NASDAQ and Dow Jones, this decomposition together with vector MEM are illustrated to have superior performances over three competing models which have been applied on forecasting realized volatility before.

这篇论文研究了当假设的数据分布与实际不符时估计多维乘积误差模型参数的方法，和该模型在预测领域的应用。论文的第一部分讨论了两种在以前的文献中被用来估计该模型的估计方法：最大似然估计法和广义矩估计法。并在对数据做了不同的干扰后比较了这两种方法。比较结果显示这两种方法都易受偏离值的影响。因此论文的第二部分提出了一种新的估计方法：权重经验似然估计法。在模拟实验和使用包含了当前经济危机间断数据的标准普尔指数的实际实验中，对比最大似然估计法和广义矩估计法，权重经验似然函数显示出了对偏离值有更好的抗性。论文的第三部分进一步研究了多维乘积误差模型在预测中的应用。并且这一部分还提出了实波动性的一种新的分解方式。分解得到的两个新的变量可以被多维乘积误差模型所模拟。通过比较标准普尔指数和纳斯达克指数的预测结果，比起以前用来估计实波动性的三种模型，多维乘积向量模型和新的分解方式显示出了更强的预测能力。

Acknowledgement

First of all I would like to thank Department of System Engineering and Engineering Management, CUHK. The department has provided an encouraging and inspiring research environment for students, both physically and mentally.

I am deeply indebt to my supervisor, Professor Kai-Pui Lam, for his patience, the most generous use of his time and effort during my PhD study. Most of all, I would like to thank him for allowing me to choose the research areas which I am interested. And also I would like to thank him for all the guidance during the five and half years.

I also want to express my gratitude to all my examiners, Professor Wai Lam, Professor Nan Chen and Professor Philip L.H. Yu, for spending time on my research and giving me valuable comments and suggestions during the past years.

During my study I received a lot of helps and supports from my colleagues, friends and families. My special thanks goes to Carlin for all the helps and his kindness of sharing with me the stock index data he recorded, to Ada for reminding me all the deadlines about graduations and helping me speed up the graduation process, to all my friends in room 615 for the good time we spent together, and finally to my parents and my wife for understanding and supporting my pursuit of PhD degree.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Autoregressive Conditional Heteroscedasticity (ARCH) models proposed in Engle [1982] and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models proposed in Bollerslev [1986] are models that widely used in modeling and forecasting financial time series. These models successfully reproduce the stylized facts of financial processes and allow the underlying variables to be both positive and negative. This will not result any problems if the variables under consideration are logarithm of returns or other time series that include both positive and negative values. However, there are a considerable number of financial data which are nonnegative. For example, the durations between trades, volumes of trades and volatilities. Although transforming these nonnegative

time series with logarithm function will turn them into suitable variables for ARCH and GARCH models, this transformation may not work for several reasons. First of all there may be zeros in the series; secondly even if there are no zeros in the data, a very small positive value could has severe impact on parameter estimation after logarithm transformation; thirdly feasible choices of innovation distributions are lessened [Engle, 2002]; and lastly the relations among variables and between each variable and its own past values do not have straightforward explanations. Therefore it is desirable to model the nonnegative time series directly.

To this end, Autoregressive Conditional Duration (ACD) model is proposed in Engle and Russell [1998] to model nonnegative financial time series, the durations between trades. Later ACD model is generalized into Multiplicative Error Model (MEM) by Engle [2002] to account for a wider range of nonnegative time series. Since then, MEM has been applied to modeling realized volatility [Lanne, 2006], volumes [Brownlees et al., 2011a] and so on. Of all the areas MEM has been applied in, modeling and forecasting volatilities is one area that attracts much interest.

As high frequency data are becoming more and more available,

volatility that can incorporate intra-day information is attracting more and more interest. GARCH models are perhaps one of the most common models on daily returns among practitioners. However they do not directly model volatility. Also due to the nonnegativity of volatility, instead of GARCH, MEM becomes a natural candidate to model volatility. One problem with volatility is that it cannot be directly observed, therefore various kinds of measures are developed for volatility. These different measures can be divided into two groups. The first group of volatility measures, for example, squared return and high-low range, utilizes daily information. Volatility measures in the second group are calculated from high-frequency time series. Realized volatility and bipower volatility, for instances, belong to the second group. No matter which group the measure is from, one single volatility measure is often not enough to encapsulate all the information on the underlying volatility. As a result, vector MEM is introduced in Engle and Gallo [2006], where three volatility measures, absolute return, daily high-low range and realized volatility are jointly modeled by vector MEM. In the past decade, more and more researches adopt this vector MEM framework to model and forecast

different measures of volatilities, and also analyze the interactions between them [Cipollini et al., 2006, 2007, Gallo and Velucchi, 2007].

The recent global financial crisis has caused a lot of unusual movements in the market. Comparing to history observations in the volatility series, these unusual movements are suspected as outliers. Traditional estimation methods like Maximum Likelihood method and Generalized Method of Moments can be heavily influenced by outliers. Motivated by these two phenomenon, this thesis focuses on the parameter estimation of MEM and its application in volatility forecasts in the presence of outliers.

Based on the kinds of outliers two situations are considered in this thesis. In the first situation the assumption is that outliers are rare and they do not contain any information on volatility. Therefore the task is to locate the outliers and to reduce their impact on parameter estimations and volatility forecasts. For this situation a novel approach, Weighted Empirical Likelihood method, is proposed to robustly estimate the parameters of MEM.

The second situation differs from the first situation from three aspects. The first difference is that situation two only considers

outliers in realized volatility. The second difference is that in situation two the outliers are not isolated and they are themselves a time series. This outlier series is recognized as jump component in the literature. And the last difference is in situation two the jump component is assumed to contain useful information on realized volatility and it is involved in the forecasting process. The task in situation two is to decompose realized volatility into continuous and jump components and to model the two components separately. For situation two this thesis proposes a multiplicative decomposition on realized volatility and vector MEM as the joint model for the continuous and jump components.

The details of the two situations are discussed in Chapter 4 and 5. The next section provides a full outline of the present thesis.

1.1 Outline of the thesis

The first part of this thesis includes Chapter 2 and Chapter 3. After the background study in Chapter 2, two estimators for MEM are discussed in Chapter 3. In Cipollini et al. [2006], Maximum Likelihood (ML) estimator and estimation function estimator are discussed in details and the latter estimator is further developed

to Generalized Method of Moments (GMM) in Cipollini et al. [2012]. However, their estimation results when data are contaminated have not been studied. Often the assumptions of ML and GMM estimators for data may be violated because of outliers and model mismatches. Therefore a simulation study on comparing the performance of the two estimators under different data contamination scenarios is conducted in Chapter 3.

Based on the results of comparisons between ML and GMM estimators, Weighted Empirical Likelihood (WEL) method is proposed in Chapter 4, which is the second part of the thesis. This chapter is based on the paper [Ding and Lam, 2012], which is accepted by Journal of Forecasting. The proposed method uses the moment conditions for GMM in Cipollini et al. [2012] as constraints for empirical likelihood functions. And weights in WEL are calculated automatically by depth function and k-means clustering. The impact of outliers on empirical likelihood functions is reduced by tilting their weights in empirical score function. Thus WEL is expected to be more robust than ML and GMM in the presence of outliers. Through a simulation study consisting of three different outlier scenarios and an application on forecasting

volatilities of S&P 500, WEL is proved to be more robust than ML and GMM.

Chapter 5 is the third part of the thesis. In this chapter vector MEM is used to forecast realized volatility by modeling two components from multiplicative decomposition. Evidences support the conclusion, that vector MEM together with the multiplicative decomposition of realized volatility has superior forecast ability than other competing models on realized volatility or additive decomposition, are found in both empirical studies on S&P 500 stock index, NASDAQ stock index and Dow Jones industry average index.

And the last chapter is conclusion and future work.

1.2 Conclusion

In conclusion, vector MEM has played an important role in modeling and forecasting volatilities, especially in the past decade. Motivated by the recent financial crisis, this thesis focuses on the estimation aspect and the application aspect of vector MEM when data contain outliers.

The main contributions are twofold. Firstly empirical likeli-

hood, k-means clustering and depth function are combined to estimate parameters of vector MEM. And the robustness increases. The outliers are detected by depth function and k-means clustering. The number of outliers is determined by k-means clustering automatically without requiring an artificial threshold value to differentiate outliers from ordinary observations. This property may be appreciated by practitioners.

Secondly realized volatility is decomposed in a multiplicative way rather than the traditional additive way so that vector MEM can be employed to model the continuous and jump components of realized volatility. Further the interactions between them can be investigated based on their conditional expectations. Therefore a clearer picture of relations between these two components can be seen without the disturbances of noise in them.

□ **End of chapter.**

Chapter 2

Background study

Chapter 2 introduces the backgrounds of several areas related to the present thesis. Firstly the recent developments of multiplicative error model are discussed. Secondly the vector MEM is explained in details because it is the main model used in this thesis. And finally two functions, copula function and depth function, which are used to analyze multivariate data are briefly introduced.

2.1 Multiplicative Error Model

2.1.1 Introduction

Based on the autoregressive conditional duration model [Engle and Russell, 1998], multiplicative error model (MEM) is proposed

in Engle [2002] to model realized volatility. Since then, MEM has been the workhorse for modeling nonnegative time series data. MEM has been used to model duration [Manganelli, 2005, Brownlees and Gallo, 2011], squared returns and absolute returns [Hautsch, 2008, Cipollini et al., 2012, Engle and Gallo, 2006, Hautsch, 2008, Manganelli, 2005, Cipollini et al., 2007, Gallo and Velucchi, 2007], realized volatility [Luca and Gallo, 2010, Lanne, 2006, 2007, Brownlees and Gallo, 2008, 2010, Otranto, 2012, Cipollini et al., 2012, Engle and Gallo, 2006, Brownlees and Gallo, 2010, Barigozzi et al., 2011, Gallo and Velucchi, 2007, Cipollini et al., 2007], realized kernel volatility [Otranto, 2011, Gallo and Otranto, 2012, Brownlees and Gallo, 2010], bipower volatility [Gallo and Velucchi, 2007, Cipollini et al., 2007], ranged volatility [Lam and NG, 2009, Cipollini et al., 2012, Engle and Gallo, 2006, Cipollini and Gallo, 2010], trading volumes of stock [Hardle et al., 2012, Bodnar and Hautsch, 2012, Hautsch, 2008, Manganelli, 2005], equity index ETF [Brownlees et al., 2011a], and implied volatility [Ahoniemi and Lanne, 2009].

MEM assumes the nonnegative time series can be represented by the multiplication of a random variable called innovation and a

scale factor that evolves in a similar way to GARCH model. This scale factor is the conditional expectation of the nonnegative time series. Therefore innovations are restricted to have a unit mean. The basic univariate multiplicative error model of order (p, q) can be written as:

$$x_t = \mu_t \epsilon_t \tag{2.1}$$

$$\mu_t = \omega + \sum_{i=1}^{p} \alpha_i x_{t-i} + \sum_{j=1}^{q} \beta_j \mu_{t-j} \tag{2.2}$$

$$\epsilon_t | \mathbb{F}_{t-1} \sim D(1, \sigma^2) \tag{2.3}$$

In equation (2.1)-(2.3), x_t is a non-negative univariate process, and $\epsilon_t \sim i.i.d.$, are innovations. They have a nonnegative conditional distribution $D(1, \sigma^2)$. μ_t is the conditional expectation of x_t that depends on MEM parameters $\theta = (\omega, \{\alpha_i\}_{i=1}^{p}, \{\beta_j\}_{j=1}^{q})$. Because x_t is non-negative, both ϵ_t and μ_t are assumed to be non-negative. Moreover according to Cipollini et al. [2006], to ensure the nonnegativity of μ and stationary distributions for x_t, the following restrictions should be imposed on θ:

$$0 < \omega, \alpha_i, \beta_j < 1, i = 1, \cdots, p, j = 1, \cdots, q \tag{2.4}$$

$$\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{q} \beta_j < 1 \tag{2.5}$$

The model discussed above is denoted as basic MEM(p,q), as it is simplest version of MEM. Since Engle [2002], MEM has been frequently used in modeling and forecasting volatilities. To capture the stylized facts of volatility, different modifications or generalizations are made to the basic MEM(p,q), resulting many different configurations. And they can be roughly classified into three groups. In the first group, different distributions are assigned to innovations. Therefore group 1 is related to equation (2.3). The second group changes the structures of equation (2.2). And the last group adds other components to the multiplication in equation (2.1). The details of the three groups are discussed in the following section.

2.1.2 Developments of MEM

In the past decade, many different versions of MEM are developed in the literature. In this section the developments of MEM are dis-

cussed from three aspects, which are corresponding to equations (2.1)-(2.3) respectively.

Equation (2.3) defines the distribution of innovation and its expectation is restricted to 1. Because the innovation is non-negative, the distribution D should be a non-negative distribution. Specifying a suitable distribution is very important to MEM, because it decides the model's capacity to capture the fat tail phenomenon of volatility. The most popular non-negative distribution is perhaps Exponential distribution. And indeed it is the first distribution applied on MEM [Engle, 2002]. The advantage of Exponential distribution is its simplicity. Because the expectation of innovation is restricted to be 1, the rate parameter is fixed to be 1. As a result only the MEM parameters are unknown and need to be estimated. In the empirical analysis Engle [2002] finds Exponential distribution may lead to underdispersion in residuals, thus an alternative distribution, Gamma distribution, is proposed. Comparing to Exponential distribution, Gamma distribution has one parameter left after imposing the unit mean restriction. Therefore it is more flexible than Exponential distribution. Engle [2002] shows Gamma distribution is strongly preferred to Exponential

distribution. Similar to the case of Gamma distribution, Weibull distribution has also been used in MEM [Hardle et al., 2012].

More complicated distributions are also used. Lanne [2006] enhances the model's flexibility by employing a mixture of two Gamma distributions on innovations. In Luca and Gallo [2010], this distribution specification is further generalized by allowing the weights of two Gamma distributions to change over time. And realized volatility forecasts on exchange rates in Luca and Gallo [2010] show 50-step ahead predictions are improved when weights of Gamma distribution are allowed to change over time.

The distribution used in Luca and Gallo [2010] is perhaps the most flexible distribution used on MEM to date. Despite the flexibility, the disadvantage is that it is difficult to generalize the mixture Gamma distribution with time varying weights to higher dimensions. When the input variable contains more than one margins, a multivariate non-negative distribution is required. And multivariate Gamma distribution can only handle positive correlation between margins[Cipollini et al., 2006]. On the other hand, Cipollini et al. [2012] introduces a semi-parametric specification on innovations so that the univariate MEM can be easily gener-

alized to multivariate MEM. This semi-parametric assumption is used in Cipollini et al. [2012], Cipollini and Gallo [2010]. And the only assumption for innovations is that their expectation equals 1.

Equation (2.2) defines the conditional expectation of x_t. Usually the conditional expectation attracts more interest than innovation, because it is used to predict future values of volatility. A common generalization of equation (2.2) when modeling volatility is adding the asymmetric effect, for example Engle [2002], Engle and Gallo [2006], Gallo and Otranto [2012] and Brownlees et al. [2011b]. The asymmetric effect is that volatility responds differently to positive return and negative return. From the perspective of modeling, a new variable is introduced to equation (2.2). The new variable equals 0 when lagged return is positive and equals $-x_{t-1}$ when lagged return is negative. MEM with this generalization is called asymmetric MEM (AMEM). Brownlees and Gallo [2010] considers including 5 different volatility measures in equation (2.2) to better forecast volatility and also to evaluate the contribution of these measures to volatility.

Besides adding lagged variables into the equation, the limita-

tion of equation (2.2) that it cannot account for the level changes in volatility is addressed in Gallo and Otranto [2012]. In Gallo and Otranto [2012] the AMEM is generalized into two AMEM with different parameters representing two states in a discrete Markov chain. Therefore both sudden and persistence changes in level of volatility can be captured. Otranto [2011] changes the parameter ω in equation (2.2) to a smooth transition function. So that the level of volatility reproduced by MEM can change with time and is not restricted within a regime like the ones in Gallo and Otranto [2012].

Equation (2.1) stands for the multiplicative decomposition of volatility. One way to generalize this equation is to include more components into the multiplication. For example in Brownlees and Gallo [2010] μ_t is further decomposed into two components capturing long run and short run persistence respectively. Also in Brownlees et al. [2011a] x_t is the multiplication of four components. These components are calculated from different frequencies, so that daily effects and intra-daily effects can be differentiated and modeled separately. Comparing the modifications made to equation (2.2), the ones made to equation (2.1) are usually

more convenient to be extended to vector MEM, as long as the additional components of different margins are independent.

Following these three directions, MEM can be generalized to a powerful tool for volatility analysis. What is more besides these generalization on equation (2.1)-(2.3), univariate MEM has been generalized to multivariate model, which is the main model used in this thesis. And its details are explained in the next section.

2.1.3 Vector MEM

In this section the multivariate version of MEM, vector MEM, is discussed. The MEM that models more than one non-negative processes first appear in Engle and Gallo [2006] and Manganelli [2005]. The former proposes a three dimensional MEM to jointly model squared return, high-low range and realized volatility while the latter applies vector MEM on duration, volume and returns. The common assumptions in these two works are that firstly the innovations of different processes are independent to each other. Secondly the conditional expectation of one process is only related to observed values of other processes but not the conditional expectations of other processes. These assumptions make equation

by equation estimation via maximum likelihood practicable. However, from today's point of view they are over-simplified. Comparing to the model in Manganelli [2005], the one in Engle and Gallo [2006] is more suitable to be called vector MEM. Because all the three variables in the model are assumed to follow MEM in Engle and Gallo [2006] and the multivariate model in Manganelli [2005] is composed of one GARCH model and two univariate MEM.

After Manganelli [2005]Engle and Gallo [2006], in Cipollini et al. [2006] vector MEM is generalized into a full multivariate model in terms that firstly the innovations are assumed to follow a multivariate model. Thus innovations of different margins are dependent to each other. Secondly the conditional expectation of one margin is related to not only the observations of other margins, but also the conditional expectations of other margins. These generalizations are essential in the development of vector MEM. Because this specification can better capture the dynamics of multivariate volatilities and it divides the interactions between different volatilities into the ones between innovations and the ones between conditional expectations. Also this specification makes equation by equation estimation impossible. Thus it stimu-

lates the inventions of new structures of innovation and estimation methods.

Examples of the new structures of vector MEM different from the one in Cipollini et al. [2006] are: Hautsch [2008] adds a common component to each margin's conditional expectation in order to study the common movements with volatility in trade size and number of trades; In Ahoniemi and Lanne [2009], the mixture MEM proposed in Lanne [2006] is generalized to vector MEM and the implied volatility of put and call options are jointly modeled; Also in the semi-parametric vector MEM proposed by Cipollini et al. [2012], the distribution of innovations is unspecified and the MEM parameters are estimated by GMM estimator.

In this thesis the vector MEM is chosen to have the same specification to the ones in Cipollini et al. [2006]. Let $\vec{x}_t \in \mathbb{R}^{K\times 1}$ be a K-dimensional process with non-negative components, the vector MEM of order (p,q) for $\vec{x}_t$ is defined as:

$$\vec{x}_t = \vec{\mu}_t \odot \vec{\epsilon}_t = \text{diag}\,(\tilde{\mu}_\text{t})\tilde{\epsilon}_\text{t} \tag{2.6}$$

$$\vec{\mu}_t = \vec{\omega} + \sum_{i=1}^{p} A_i \vec{x}_{t-i} + \sum_{j=1}^{q} B_j \vec{\mu}_{t-j} \tag{2.7}$$

$$\vec{\epsilon}_t \sim D(\vec{1}, \Sigma) \tag{2.8}$$

where $\odot$ indicates the Hadmard (element by element) product. $\vec{\mu}_t \in \mathbb{R}^{K\times 1}$ is the conditional expectation on the information set $\mathbb{F}_{t-1}$. $\vec{\omega} \in \mathbb{R}^{K\times 1}$, $A_i \in \mathbb{R}^{K\times K}, i = 1, \cdots, p$ and $B_j \in \mathbb{R}^{K\times K}, j = 1, \cdots, q$ are MEM parameters. If parameter matrix A or B are not diagonal, one variable and its conditional expectation will affect other variables' conditional expectations. Just like the univariate MEM, equation (2.7) for conditional expectation can be extended to include asymmetric effect. By specifying A and B to be non-diagonal, the interactions between input variables and between the conditional expectations of input variables can be modeled respectively.

The innovation vector $\vec{\epsilon}_t$ is a conditionally stochastic K-dimensional $i.i.d.$ process. Its density function is defined over a $[0, +\infty)^K$ support, with unit vector as expectation and a general variance-covariance matrix Σ, which is usually unknown.

These conditions on expectations and variance covariance matrix guarantee that

$$\mathbb{E}(\vec{x}_t|\mathbb{F}_{t-1}) = \vec{\mu}_t \tag{2.9}$$

$$\mathbb{V}(\vec{x}_t|\mathbb{F}_{t-1}) = \vec{\mu}_t\vec{\mu}_t^{'} \odot \Sigma = diag(\vec{\mu}_t)\Sigma diag(\vec{\mu}_t) \tag{2.10}$$

If the innovations have a diagonal variance covariance matrix, then the vector model can be considered as several univariate MEM models. Engle and Gallo [2006] use multivariate MEM with parameter matrix A non-diagonal but both B and the variance covariance matrix Σ diagonal. This assumption on MEM parameters means that the past values of one margin have influences on the conditional expectations of all the margins including itself. But the past values of one margin's conditional expectation are not related to those of other margins. And the MEM parameters of each margin can be estimated by equation by equation method.

However the equation by equation method is not feasible when the innovations' variance covariance matrix is not diagonal. Therefore it is necessary to assign a multivariate distribution to them so that Maximum Likelihood method can be used to estimate the parameters. Cipollini et al. [2006] point out the drawback of multivariate normal distribution and multivariate gamma distribution. Firstly, many researches have shown the assumption that financial data follow normal distribution is false. Secondly, these kinds of distributions have constraints on parameters which cannot be easily met. And most of them require computing numerical

inversions in the pdf formula.

The answer to these disadvantage of multivariate distributions is copula introduced by Cipollini et al. [2006]. Any multivariate distribution can be separated into margin distributions and a copula function which describes the dependence structure. Copula function describes the dependence between variables and are highly involved in multivariate financial analysis nowadays [Trivedi and Zimmer, 2007]. Copula is more flexible and more tractable. It allows different univariate distribution for different margins. These are the reasons to employ copula functions. Among the pool of copulas the most commonly used copula are the Gumbel copula for extreme distributions, normal copula for linear correlation and Archimedean copula and the Student's t copula for dependence in tail. In this thesis Student's t copula and Gamma distribution are chosen for innovations, because Student's t copula generalizes normal copula by allowing for non-zero dependence in extreme tails. It is fast growing in usage because the degree of tail dependency can be set by changing the degrees of freedom parameter ν. Large value for τ say 100, approximates a normal distribution. Conversely small value for τ, say 3, increases the

tail dependency, until τ equals 1 it become a Cauchy distribution. And after restricting expectations of innovations to be one, Gamma distribution still has one parameter for each margin to estimate which increases the model's flexibility [Engle, 2002]. Thus the distribution of innovations is:

$$\epsilon_t|\mathbb{F}_{t-1} \sim C_t(R, \nu) \times \prod_{i=1}^{K} \Gamma(\phi_i) \tag{2.11}$$

In this formula, the first part represents Student's copula, where R is the correlation and τ is the degree of freedom. Both R and τ controls tail dependence which is symmetric.

In order to illustrate the structure of vector MEM in more detail, assume K equals 2, the parameter matrices A and B are diagonal and the order of the model is (1,1). Further assume the innovations follow Gamma margin distribution and their dependence structure is captured by Student's t copula, the model can be written as:

$$x_{1,t} = \mu_{1,t}\epsilon_{1,t} \tag{2.12}$$

$$\mu_{1,t} = \omega_1 + a_1 x_{1,t-1} + b_1 \mu_{1,t-1} \tag{2.13}$$

$$x_{2,t} = \mu_{2,t}\epsilon_{2,t} \tag{2.14}$$

$$\mu_{2,t} = \omega_2 + a_2 x_{2,t-1} + b_2 \mu_{2,t-1} \tag{2.15}$$

where $x_{i,t}$ is a non-negative univariate process, $(\omega_1, a_1, b_1, \omega_2, a_2, b_2)$ are mem parameters. And innovation vector $\{\epsilon_t\}$ is a conditional stochastic 2-dimensional i.i.d. process who has a joint conditional distribution

$$\vec{\epsilon_t}|\mathbb{F}_{t-1} \sim C_{\rho,\tau} \times \prod_{i=1}^{2} Gamma(\phi_i) \tag{2.16}$$

where $C_{\rho,\tau}$ is the cdf of Student's t copula with correlation parameter ρ and degree of freedom τ. $Gamma(\phi(i))$ stands for the cdf function of Gamma distribution with shape parameter $\phi(i)$ and scale parameter $1/\phi(i)$. And the density of student's t copula is :

$$c_{\rho,\tau}(u_1,u_2) = |\rho|^{-\frac{1}{2}} \frac{\Gamma(\frac{\tau+N}{2})[\Gamma(\frac{\tau}{2})]^2(1+\frac{1}{\tau}(t_\tau^{-1}(u_1)^2 + t_\tau^{-1}(u_2)^2) - 2\rho t_\tau^{-1}(u_1)t_\tau^{-1}(u_2))^{-\frac{\tau+2}{2}}}{[\Gamma(\frac{\tau+1}{2})]^2\Gamma(\frac{\tau}{2})\prod_{i=1}^{2}(1+\frac{t_\tau^{-1}(u_i)^2}{\tau})^{-\frac{\tau+1}{2}}} \tag{2.17}$$

where $c_{\rho,\tau}(u_1, u_2)$ is the pdf of Student's t copula with correlation

parameter ρ and degree of freedom τ at (u_1, u_2), and t_τ denotes CDF of student's t distribution with freedom τ.

2.2 Two functions for multivariate analysis

In this section two functions,copula function and depth function, used in this thesis are introduced. Copula function is closely related to the developments of vector MEM. It is an essential component in distribution assumption of innovations. Comparing to univariate distribution, choices for multivariate distribution are limited, let alone non-negative multivariate distribution. With the help of copula function, the multivariate distribution of innovations can be constructed based on different non-negative univariate distributions. Depth function is a powerful tool in multivariate analysis. It can rank multi-dimensional data non-parametrically. Therefore it is used for detecting outliers from innovations.

2.2.1 Copula function

Although analysis methods for multivariate distributed normal data are well developed, general approaches for joint nonlinear model or nonnormal data are not. And copula seems to be a

promising way to handle these problems. Besides this, copula can provide a more flexible multivariate model, it allows each variable to have different marginal distributions.

Definition An K-dimensional copula is a function C from the unit K-cube $[0,1]^K$to the unit interval $[0,1]$ which satisfies the following conditions:

(1) $C(1,\cdots,1,u_k,\cdots,1) = u_k$ for all $k \leq K$ and all u_k in $[0,1]$;

(2) $C(u_1,\cdots,u_K) = 0$ if $u_k = 0$ for any $k \leq K$;

(3) C is K-increasing.

Property 1 says that when the realizations of all other variables are known with probability 1, then the joint probability of the K outcomes is the same as the uncertain one.

Property 2 is referred as the grounded property. The joint probability is zero when any of the outcomes are zero.

In property 3, K-increasing means that:

$$\sum_{i_1=1}^{2}\cdots\sum_{i_K=1}^{2}(-1)^{i_1+\cdots+i_K}C(u_{1,i_1},\cdots,u_{K,i_K}) \geq 0 \tag{2.18}$$

for all hyper-rectangle $\prod_{k=1}^{K}[u_{k,1},u_{k,2}] \in [0,1]^K$.

An K-copula is a K-dimensional distribution function with all K univariate margins following uniform distribution in $[0, 1]$. To see the relation between joint distribution and copula, consider K variables $(x_1, \cdots, x_K)$ with a joint cumulative distribution function $F(x_1, \cdots, x_K)$, and margin cdf $F_1(x_1), \cdots, F_K(x_K)$ with $F_1^{-1}, \cdots, F_K^{-1}$ as the inverse function of margin cdf. Let $u_1, \cdots, u_K$ be random variables following uniform distribution in $[0, 1]$. And let $(U_1, \cdots, U_K) \in [0, 1]^K$ satisfy $U_k = F_k(x_k), \forall k = 1, \cdots, K$. Therefore $U_1, \cdots, U_K$ are the cdf of $x_1, \cdots, x_K$. Then

$$\begin{aligned} & F(x_1, \cdots, x_K) \\ = \; & F(F^{-1}(U_1), \cdots, F_K^{-1}(U_K)) \\ = \; & Pr[u_1 \leq U_1, \cdots, u_K \leq U_K] \\ = \; & C(u_1, \cdots, u_K) \end{aligned} \tag{2.19}$$

The above results imply that copulas can be used to express a multivariate distribution in terms of its marginal distributions.

Sklar's Theorem: Let F be an K-dimensional distribution function with continuous margins $F_1, \cdots, F_K$. The F has a unique

copula representation:

$$F(x_1, \cdots, x_k, \cdots, x_K) = C(F_1(x_1), \cdots, F_k(x_k), \cdots, F_K(x_K)) \tag{2.20}$$

After introducing the main concepts of copula function, in the following four examples of copula functions are listed.

Product copula The simplest copula, product copula, has the form

$$C(u_1, \cdots, u_K) = \prod_{k=1}^{K} u_k \tag{2.21}$$

where $u_k \in [0, 1], k = 1, \cdots, K$. The product copula assumes different margins are independent to each other. And the corresponding density is

$$c(u_1, \cdots, u_K) = 1 \tag{2.22}$$

Multivariate Gaussian copula Gaussian copula is the copula associated with bivariate normal distribution, and is the dependence function implicitly assumed whenever the bivariate normal distribution is used. Let ρ be a symmetric, positive definite matrix with diagonal elements equal 1. ρ is the covariance matrix of

margins. Denote $(u_1, \cdots, u_K)$ as the cdf of margins, Φ as the standard normal distribution and Φ_ρ as the standard multivariate normal distribution with correlation matrix ρ. The K-dimensional multivariate Gaussian copula can be defined as follows:

$$\begin{aligned} & C(u_1, \cdots, u_k, \cdots, u_K; \rho) \\ = \ & \Phi_\rho(\Phi^{-1}(u_1), \cdots, \Phi^{-1}(u_k), \cdots, \Phi^{-1}(u_K)) \end{aligned} \tag{2.23}$$

The corresponding density is

$$c(u_1, \cdots, u_k, \cdots, u_K; \rho) = \frac{1}{|\rho|^{\frac{1}{2}}} \exp(-\frac{1}{2}\varsigma^T(\rho^{-1} - I)\varsigma) \tag{2.24}$$

with $\varsigma = (\varsigma_1, \cdots, \varsigma_K)^T$ and $\varsigma_k = \Phi^{-1}(u_k), \forall k = 1, \cdots, K$.

Multivariate Student's copula Student's t copula is the dependence structure associated with the bivariate Student's t distribution. Let ρ be the same as in Gaussian copula, a symmetric and positive definite matrix with diagonal elements equal 1 and $T_{\rho,\tau}$ the standardized multivariate Student's distribution with τ as the degree of freedom and correlation matrix ρ. The multivariate

Student's copula can be defined as follows:

$$\begin{aligned} & C(u_1, \cdots, u_k, \cdots, u_K; \rho, \tau) \\ = \; & T_{\rho,\tau}(t_\tau^{-1}(u_1), \cdots, t_\tau^{-1}(u_k), \cdots, t_\tau^{-1}(u_K)) \end{aligned} \tag{2.25}$$

where t_τ^{-1} the inverse of the univariate Student's distribution with degree of freedom τ. The corresponding density is

$$\begin{aligned} & c(u_1, \cdots, u_k, \cdots, u_K; \rho, \tau) \\ = \; & |\rho|^{-\frac{1}{2}} \frac{\Gamma(\frac{\tau+K}{2})[\Gamma(\frac{\tau}{2})]^K (1 + \frac{1}{\tau}\varsigma^T \rho^{-1} \varsigma)^{\frac{\tau+K}{2}}}{[\Gamma(\frac{\tau+1}{2})]^K \Gamma(\frac{\tau}{2}) \prod_{k=1}^K (1 + \frac{\varsigma_k^2}{\tau})^{-\frac{\tau+1}{2}}} \end{aligned} \tag{2.26}$$

with $\varsigma = (\varsigma_1, \cdots, \varsigma_K)^T$ and $\varsigma_k = t_\tau^{-1}(u_k), k = 1, \cdots, K$.

The Student's t copula generalizes the normal copula by allowing for non-zero dependence in extreme tails. For example, if two variables exhibit lower tail dependence, then it implies a non-zero probability of observing a large increase in one variable with a large increase on the other. On the other hand Gaussian copula assumes that this probability is zero.

Archimedean copulas Archimedean copulas are popular because they are easily derived and are capable of capturing wide ranges of dependence. However, the Archimedean copula family is very

difficult to generalized to higher dimensions. Genest and MacKay [1986] define Archimedean copulas as the following:

$$\begin{aligned} & C(u_1, \cdots, u_k, \cdots, u_K) \\ = \; & \varphi^{-1}(\varphi(u_1) + \cdots + \varphi(u_k) + \cdots + \varphi(u_K)) \end{aligned} \tag{2.27}$$

Equation (2.27) holds only if $\sum_{n=1}^{N} \varphi(u_n) \leq \varphi(0)$. Otherwise the Archimedean copula is 0. $\varphi(u)$ is a C^2 function with $\varphi(1) = 0, \varphi^{'} < 0$ and $\varphi^{''} > 0$ for all $0 < u < 1$. $\varphi(u)$ is called the generator of Archimedean copula.

Product copula belongs to the Archimedean copula family. Besides product copula, two bivariate Archimedean copulas examples are provided below:

Copula	$\varphi(u)$	$C(u_1, u_2)$
Gumbel	$(-\ln u)^{\alpha}$	$\exp(-(\tilde{u}_1^{\alpha} + \tilde{u}_2^{\alpha})^{\frac{1}{\alpha}})$
Clayton	$u^{-\alpha} - 1$	$(u_1^{\alpha} + u_2^{\alpha} - 1)^{-\frac{1}{\alpha}}$

where $\tilde{u} = -\ln u$

2.2.2 Depth function

The vector MEM used in this thesis assumes that the innovations follow a non-negative multivariate distribution. Under the circumstance that the distribution is not specified, the only information on innovations is that their expectation is equal to a vector of 1. To study the properties and structure of innovations, nonparametric approaches in multivariate analysis are required.

However, classical multivariate statistics are extended from the univariate ones and the data is assumed to be multivariate normally distributed or "almost" normally distributed. In the case of innovations, this assumption clearly cannot hold. Because the distribution of innovations are non-negative. In this section an alternative non-parametric analysis tool for multivariate data, depth function, is introduced.

General speaking, depth function measures the "outlyingness" of an observation in a data cloud or how "deep" the observation is inside a data cloud. The closer it is to the center of the data cloud, the higher the depth value. Distance to center is the only concern and directions are not differentiated. Depth function can also be considered as a rank tool for multivariate data. The obser-

vation with the highest depth value is the center of the data. The observations with lower depth values have lower ranks. The observations with the lowest depth values have the highest probability to be outliers.

In Zuo and Serfling [2000], four desirable properties for depth function are proposed. They are affine invariance, maximality at center, monotonicity relative to deepest point and vanishing at infinity. Suppose a random variable $\varsigma \in \mathbb{R}^K$ has a distribution F_ς. And the set of distributions on the Borel sets of $\mathbb{R}^K$ is $\mathbb{F}$. Depth function can be defined based on the four properties:

Definition of depth function Let the function $D(\cdot;\cdot) : \mathbb{R}^K \times \mathbb{F} \to \mathbb{R}$ be bounded, non-negative and satisfies the following

(i)$D(Ax + b, F_{AX+b}) = D(x, F_X), \forall$ random variable $X \in \mathbb{R}^K, \forall A \in \mathbb{R}^{K\times K}, detA \neq 0, b \in \mathbb{R}^K$;

(ii) $D(c; F) = \sup_{x\mathbb{R}^K} D(x; F)$ holds for any distribution F having a center c;

(iii) if F has a deepest point c, then $D(x; F) \leq D(ax + (1 - a)c); F), \forall a \in [0, 1], \forall x \in \mathbb{R}^K$;

(iv) $D(x; F) \to 0$ as $\| x \| \to 0, \forall F \in \mathbb{F}, \forall x \in \mathbb{R}^K$.

Then D is a depth function.

In statistics literatures, many different kinds of depth functions are proposed. Here three of them are provided as examples. They are Tukey depth [Tukey, 1975], simplicial depth[Liu, 1990] and Oja depth [Oja, 1983].

Tukey depth(HD) Tukey depth is also called half space depth. It is calculated by the following equation.

$$HD(x;F) = \inf_{H}\{P(H) : H\ is\ a\ closed\ half\ space\ in\ \mathbb{R}^K\ and\ x \in H\} \tag{2.28}$$

The calculation of Tukey depth is illustrated through the following two dimensional example. Suppose we are considering the Tukey depth of an observation, whose coordinate is (a_1, a_2), in a data set A. Randomly draw a line that pass through (a_1, a_2), for example, $a_1y - a_2x = 0$. This line cut the plane into two parts: $a_1y - a_2x > 0$ and $a_1y - a_2x < 0$. By counting the observations within each half plane we get two values. The smaller one is the current depth. If we rotate the line around the point (a_1, a_2) for 360 degree and record the depth values at each degree, after rotating the line to the starting position, i.e. finishing a full circle, the smallest value among all the depth values recorded at different

degrees is the Tukey depth of observation at (a_1, a_2) based on the data set A.

simplicial depth(SD) The SD of x based on data set $(X_1, X_2, \cdots, X_T), X_t \in \mathbb{R}^K, \forall t = 1, \cdots, T$ is:

$$SD(x; F_n) = \binom{T}{K+1}^{-1} \sum_{1 \le i_1 \le \cdots \le i_{K+1} \le T} I(x \in S[X_{i_1}, \cdots, X_{i_{K+1}}]) \tag{2.29}$$

In this equation $I(\cdot)$ is the indicator function and F_n is the empirical distribution function of $(X_1, X_2, \cdots, X_T)$. $S[\cdot]$ is the close simplex with vertices $(X_1, X_2, \cdots, X_T)$, which means S is the set of all the convex combinations of $(X_1, X_2, \cdots, X_T)$.

The calculation of simplicial depth is illustrated through the example in Figure 2.1. The data set contains three observations $A, B, C \in \mathbb{R}^2$. According to the definition, depth value of $O1$ is 1 because there is only one simplex that has three vertices and contains $O1$. And the depth value of $O2$ is 0. Therefore $O1$ is closer to the center of (A, B, C) comparing to $O2$.

Oja depth (OD) Oja depth is developed in Serfling and Zuo [2000] from a location measure proposed by Oja [1983]. For a

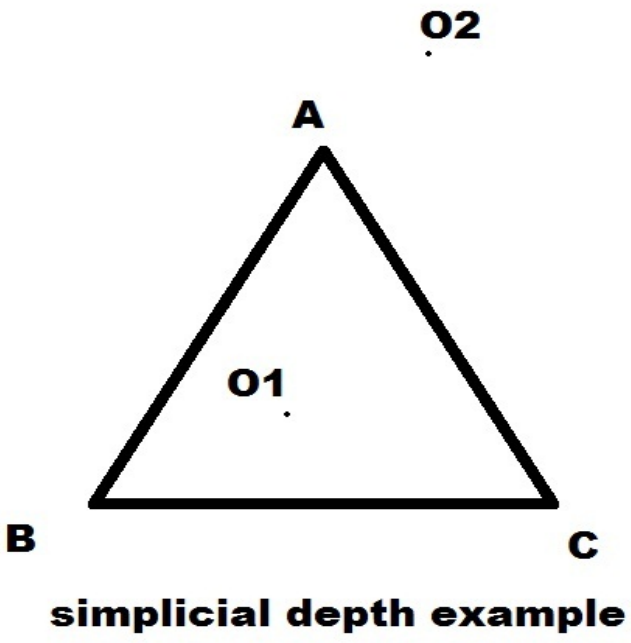

Figure 2.1: simplicial depth

given set of data, $(X_1, X_2, \cdots, X_T), X_t \in \mathbb{R}^K, \forall t = 1, \cdots, T$, the depth value at $x \in \mathbb{R}^K$ is

$$OD(x; F_n) = \binom{T}{K}^{-1} (1 + \sum_{1 \le t_1 \le \cdots \le t_K \le T} v(S[x, X_{t_1}, \cdots, X_{t_K}]))^{-1} \tag{2.30}$$

In this equation $S[\cdot]$ is the same as in SD function, a closed simplex composed of vertices $(x, X_{t_1}, \cdots, X_{t_K})$ and $v(S[\cdot])$ is the volume of simplex S. The volume for the $K + 1$ vertices simplex $(x, X_{t_1}, \cdots, X_{t_K})$is calculated as:

$$\frac{1}{K!} det(X_{t_1} - x, \cdots, X_{t_K} - x) \tag{2.31}$$

One difference between Oja depth and simplicial depth is that the simplex in Oja depth uses x as an vertex while all of the vertices of the simplex in simplicial are from the data set. Figure 2.2 contains an example of Oja depth. The data set is again $A, B, C \in \mathbb{R}^2$. And the simplexes that contain vertex $O1$ is triangles $ABO1$,$BCO1$ and $ACO1$. Therefore the summation of volumes of these simplexes is the area of the triangle ABC. Following the same logic, the simplexes that contain the vertex $O2$ are $ABO2$,$BCO2$ and $ACO2$. And the summation of their volumes

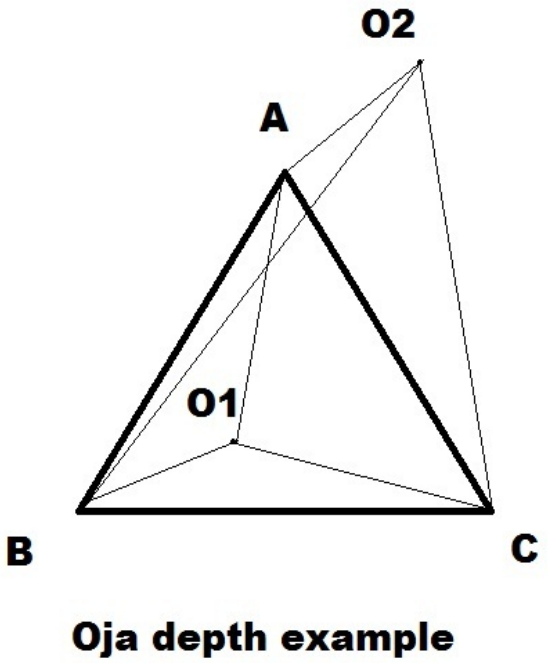

Figure 2.2: Oja depth

is the area of ABC plus two times the area of $ACO2$. Therefore the Oja depth of $O2$ is smaller than the Oja depth of $O1$, which means that $O1$ is closer to the center of data set (A, B, C).

□ **End of chapter.**

Chapter 3

Two Estimators for Vector MEM

This chapter is about the estimation methods that are used on vector MEM previously. As a start two stage ML is discussed. Although it is not a typical method to estimate parameters of vector MEM, its results can be used as starting values for both ML and GMM estimators. Following two stage ML method, ML and GMM estimators for vector MEM are introduced. A comparison based on simulation is conducted at the end of this chapter. The results from simulation show the two estimators can be heavily influenced by outliers, which is the motivation for proposing WEL in Chapter 4.

3.1 Two Stage Maximum Likelihood

Two stage ML estimator is similar to ML estimator. They employ the same distribution assumption on innovations. The difference is that the MEM parameters and copula parameters are estimated separately by two stage ML estimator rather than jointly by ML estimator. In the parameter estimation processes of both ML and GMM estimators, starting values play an important role. Good starting values can quickly lead the optimization to convergence. Therefore two stage ML estimator is used for providing starting values for both ML and GMM estimators. Although GMM estimator only needs the results from stage one of two stage ML estimation. In fact, if stage two of two stage ML is omitted, stage one can be considered as a quasi ML estimator for the copula part. Because when the score function only contain log-likelihood functions of margins, the dependence between innovations are assumed to follow independent copula. According to Chapter 2, the density of independent copula is 1, therefore its logarithm value is 0. If the margins are assumed to follow univariate MEM where Exponential distribution is assigned to innovations, the stage one of two stage ML estimator is a quasi ML estimator for vector

MEM.

Since two stage ML is closely related to ML estimator, and it can provide reasonably good starting values for ML and GMM estimators, it is introduced before the two estimators.

3.1.1 Introduction

Two stage ML estimator is also called inference functions for margin (IFM). Both Patton [2006] and Hu [2006] use this method to estimate models containing copula and margins. IFM is suitable for the situation when multivariate models under consideration have an unknown parameter vector that may be partitioned into parameters only related to the margin distributions and parameters only related to the copula. If such a partition is not possible, full maximum likelihood estimator is the natural estimator to employ. IFM method has the benefit of being computational tractable, at the cost of a loss of full efficiency. Normally results of full ML are not the same with those of IFM. But Patton [2006] shows that IFM is a good alternative to full maximum likelihood method (FML). For it can achieve great computational savings and in addition as the IFM is able to explore all available infor-

mation on both variables, for example when some margins have more data than the other margins. Therefore in some situations IFM actually outperforms FML.

Copula representation splits the parameters into parameters for margin distributions and common parameters of the dependence structure (or the parameters of the copula). The log-likelihood could then be written as

$$\begin{aligned} l(\vec{\theta}) &= \sum_{t=1}^{T} \ln c(F_1(x_{1,t};\vec{\theta_1}),\cdots,F_k(x_{k,t};\vec{\theta_k}),\cdots,F_K(x_{K,t};\vec{\theta_K});\vec{\alpha}) \\ &\quad + \sum_{t=1}^{T}\sum_{k=1}^{K} \ln f_k(x_{k,t};\vec{\theta_k}) \end{aligned} \tag{3.1}$$

where $\vec{\theta_k}$ are the parameters for the kth margins and $\vec{\alpha}$ are the parameters for copula. c is the density of copula function, and F_k and f_k are the cdf and pdf of the distribution corresponding to kth margin.

In vector MEM, if the parameter matrices A and B are both diagonal, the first stage of two stage ML estimator can be optimized via equation by equation method, i.e. the parameters of margins can be estimated separately. For example, the estimator

of the kth margin is

$$\hat{\vec{\theta}}_k = \arg_{\vec{\theta_k}} \max \sum_{t=1}^{T} \ln f_k(x_{k,t}; \vec{\theta}_k) \tag{3.2}$$

And the estimator for $\vec{\alpha}$ given the previous estimation results which are denoted as $(\hat{\vec{\theta}}_1, \cdots, \hat{\vec{\theta}}_K)$ is

$$\hat{\vec{\alpha}} = \arg_{\vec{\alpha}} \max \sum_{t=1}^{T} \ln c(F_1(x_{1,t}; \hat{\vec{\theta}}_1), \cdots, F_K(x_{K,t}; \hat{\vec{\theta}}_K); \vec{\alpha}) \tag{3.3}$$

Very similar to IFM method, in two stage ML the parameter vector α could be estimated without specifying the distribution of margins. The method consists of transforming the data $(x_{1,t}, \cdots, x_{K,t})$ into uniform variables $(\hat{u}_{1,t}, \cdots, \hat{u}_{K,t})$ by empirical distribution, and the estimator can be written as:

$$\hat{\vec{\alpha}} = \arg\max_{\vec{\alpha}} \sum_{i=1}^{T} \ln c(\hat{u}_{1,t}, \cdots, \hat{u}_{K,t}; \vec{\alpha}) \tag{3.4}$$

In this case, $\hat{\alpha}$ could be viewed as the ML estimator given the observed margins(without assumptions on the parametric form of the marginal distributions). Because it is based on the empirical distribuions, we call it the canonical maximum likelihood method (CML).

If parameter matrices A or B is not diagonal, then the first stage of two stage ML method cannot be estimated via equation by equation method. All the log-likelihood functions of margins are added together and during the first stage estimation the copula is assumed to be independent copula. The estimator for first stage in this case is:

$$(\hat{\vec{\theta}}_1, \cdots, \hat{\vec{\theta}}_K) = \arg_{(\vec{\theta_1}, \cdots, \vec{\theta_K})} \max \sum_{t=1}^{T} \sum_{k=1}^{K} \ln f_k(x_{k,t}; \vec{\theta}_k) \qquad (3.5)$$

And the second stage estimation is same as the one when A and B are diagonal. The two stage maximum likelihood estimator is proved to be consistent and asymptotic normal under some regularity conditions [Newey and McFadden, 1994].

3.1.2 Simulation of two stage ML

In this section a simulation study is conducted to test the performance of two stage ML estimator. It is expected that it can provide reasonable results so that ML and GMM estimators can use the results to further estimate the parameters. For each of the two cases considered, there are 100 repetitions, and in each

repetition the data length is 1000.

Diagonal case

In the situation where A and B of the MEM parameters are diagonal, the model used to generate data is:

$$\begin{aligned} x_1 &= \mu_{1,t}\epsilon_{1,t} \\ x_2 &= \mu_{2,t}\epsilon_{2,t} \\ \vec{\epsilon_t} &\sim c(\vec{\epsilon_t}) \cdot \prod Gamma \\ \mu_{1,t} &= \omega_1 + \alpha_1 x_{1,t-1} + \beta_1 \mu_{1,t-1} \\ \mu_{2,t} &= \omega_2 + \alpha_2 x_{2,t-1} + \beta_2 \mu_{2,t-1} \end{aligned} \tag{3.6}$$

Two sets of data are generated from model (3.6). Each data set contain 100 repetitions, the estimation results are tabulated in Table 3.2 and 3.3. In the first set the copula function is Gaussian copula and in the second the copula is assumed to be Student's t copula. The parameters of Gamma distributions for both sets are same, as well are MEM parameters $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2)$. The results in Table 3.2 and 3.3 show that estimated parameters are quite close to their true values. Also this simulation test results

indicate that two stage ML are supposed to perform fairly well on data from Gaussian or Student's t copula.

Non-diagonal case

When MEM parameter matrix A and B are not diagonal, which is often the case in real world application, estimating each margin separately in the first stage of two step ML method is impractical. Because even without the dependence structure, copula function, each margin still affects other margins through off-diagonal elements of A and B. As stated in last section, in this case the parameters in the first stage of two step ML are estimated by assuming the copula function is product copula. This assumption means in the first stage the interactions between innovations are not considered. The model used to generate data in this case is a 2-dimensional vector MEM model:

$$
\begin{aligned}
x_1 &= \mu_{1,t}\epsilon_{1,t} \\
x_2 &= \mu_{2,t}\epsilon_{2,t} \\
\vec{\epsilon}_t &\sim c_{\rho,\tau}(\vec{\epsilon}_t) \cdot \prod Gamma \qquad (3.7) \\
\mu_{1,t} &= \omega_1 + \alpha_{1,1}x_{1,t-1} + \alpha_{1,2}x_{2,t-1} + \beta_{1,1}\mu_{1,t-1} + \beta_{1,2}\mu_{2,t-1} \\
\mu_{2,t} &= \omega_2 + \alpha_{2,1}x_{1,t-1} + \alpha_{2,2}x_{2,t-1} + \beta_{2,1}\mu_{1,t-1} + \beta_{,22}\mu_{2,t-1}
\end{aligned}
$$

In the model (3.7), $c_{\rho,\tau}$ is assumed to be Student's t copula with correlation parameter ρ and degree of freedom τ. In total 100 data sets are generated, and each data set has a size 2 by 1000. The estimation results together with the true values of each parameter are tabulated in Table 3.4. From the results it can be seen that the estimated values are fairly close to their true values. However, comparing to the results in Table 3.3, the bias in Table 3.4 is larger. This is due to the fact that there are more parameters need to be estimated in the non-diagonal case.

From the results of both diagonal case and non-diagonal case, two stage ML method is proved to be able to provide good starting values for other methods which are sensitive to starting values.

3.2 Maximum Likelihood estimator

Following Cipollini et al. [2007] in this paper the ML estimator assumes that the joint distribution of innovations are Gamma distributions linked by Student's t copula. Student's t copula is fast growing in usage because it generalizes normal copula by allowing non-zero dependence in extreme tails through different degrees of freedom, denoted as τ. Student's t copula with large τ, say 100, approximates a normal copula. Conversely small values for τ lead to high tail dependency. After setting mean to one, Gamma distribution has one parameter left to estimate, which increases the model's flexibility comparing to exponential distribution. The distribution of innovations are assumed to be:

$$\vec{\epsilon}_t|\mathbb{F}_{t-1} \sim C_t(R,\tau) \times \prod_{k=1}^{K} F(\phi_k, \epsilon_{k,t}) \tag{3.8}$$

where $C_t(R,\tau)$ is the cdf of student's t copula with correlation matrix R and degree of freedom τ. $F(\phi,\epsilon)$ is the cdf of Gamma distribution at ϵ with shape parameter ϕ and scale parameter $\frac{1}{\phi}$. The density for this joint distribution at time t is:

$$c_{t,R,\tau}(F_1(x_{1,t}/\mu_{1,t}),\ldots,F_K(x_{K,t}/\mu_{K,t}))\times\prod_{k=1}^{K}f_k(x_{k,t}/\mu_{k,t}) \quad (3.9)$$

In the the first part of the above formula, $c_{t,R,\tau}(\cdot)$, represents the density of Student's t copula, where R and τ are the same parameters as in equation 3.8. Following the discussions in Chapter 2, Student's t copula's density at $(\epsilon_{1,t},\cdots,\epsilon_{K,t})$ is:

$$\begin{aligned} & c_{t,R,\tau}(F_1(x_{1,t}/\mu_{1,t}),\ldots,F_K(x_{K,t}/\mu_{K,t})) \\ = \ & |R|^{-\frac{1}{2}}\frac{\Gamma(\frac{\tau+K}{2})[\Gamma(\frac{\tau}{2})]^K(1+\frac{1}{\tau}Q^TR^{-1}Q)^{\frac{\tau+K}{2}}}{[\Gamma(\frac{\tau+1}{2})]^K\Gamma(\frac{\tau}{2})\prod_{k=1}^{K}(1+\frac{q_k^2}{\tau})^{-\frac{\tau+1}{2}}} \end{aligned} \quad (3.10)$$

where $Q=(q_1,\cdots,q_K)^T$ and $q_k=t_\tau^{-1}(F_k(x_{k,t}/\mu_{k,t}))$ and t_τ is the cdf of student's t distribution with degree of freedom τ.

And the second part of equation 3.9 is the density of margins where F_k and f_k are conditional cdf and pdf of $\epsilon_{k,t}$. Since the margin distribution of innovations is Gamma distribution, F_k and f_k can be expressed as below:

$$f_k(\epsilon_{k,t})=\frac{\phi_k^{\phi_k}}{\Gamma(\phi_k)}\epsilon_{k,t}^{\phi_k-1}\exp(-\phi_k\epsilon_{k,t}) \quad (3.11)$$

$$F_i(\epsilon_{k,t}) = \Gamma(\phi_k; \phi_k \epsilon_{k,t}) \tag{3.12}$$

where $\Gamma(\varsigma, x)$ is the incomplete Gamma function with parameter ς computed at x, or in other words, the cdf of a $Gamma(\varsigma, 1)$ random variable computed at x [Cipollini et al., 2006]. And ϕ_k is the Gamma parameter for kth dimension.

For ML estimator, the parameters are estimated by the following maximization problem(4.5) [Cipollini et al., 2007]:

$$\begin{aligned} & (\hat{\theta}_{ML}, \hat{\rho}, \hat{\tau}) \\ = \ & \arg \max_{\theta \in \Theta, \rho \in [-1,1] 0, \tau > 0} [\sum_{t=1}^{T} (\ln c_{t,R,\tau} + \sum_{k=1}^{K} f_k(\frac{x_{k,t}}{\mu_{k,t}}))] \end{aligned} \tag{3.13}$$

where Θ is the feasible region for MEM parameters Cipollini et al. [2006].

3.2.1 Derivatives of score function

The parameters of vector MEM can be estimated by maximizing the log-likelihood function (score function) using different numerical methods. Often the optimization methods, Newton methods for example, require derivatives of the log-likelihood functions.

Although softwares like R and MATLAB can calculate numer-

ical derivatives, the analytic ones are still helpful. For example, Levy [2003] has show that analytic derivatives perform better than numerical derivatives in short GARCH sequence. However, the calculations of numerical derivatives usually cost less time than analytic ones. This is because they use finite-difference approximations which will only evaluate the objective function at several points in the neighborhood of the current estimations of the parameters. This is contrast to analytic derivatives which are computed recursively using all the data.

The derivatives of GARCH models have been developed by Fiorentini et al. [1996], and Cipollini et al. [2006] states the details of calculating the derivatives of vector MEM. Based on their results and for illustration purpose, the calculations of derivatives of the vector MEM used in section 3.1.2 are discussed. The basic vector MEM is assumed to have two margins which follows Gamma distribution. And the copula is assumed to be Student's t copula.

Most derivatives of the model have analytic forms, only the derivative against ϕ , the parameter of margin distribution and τ , the parameter of Student's t copula need to be calculated

numerically.

Derivatives of Margin Log-likelihood

Because $x_{k,t}$ can be expressed as the multiplication of $\mu_{k,t}$ and $\epsilon_{k,t}$, $\frac{\partial x_{k,t}}{\partial \vec{\omega}}$,$\frac{\partial x_{k,t}}{\partial \vec{\alpha}}$,$\frac{\partial x_{k,t}}{\partial \vec{\beta}}$ and $\frac{\partial x_{k,t}}{\partial \vec{\phi}}$ all equal to zero, where $\vec{\omega} = (\omega_1, \omega_2)$, $\vec{\alpha} = (\alpha_1, \alpha_2)$, $\vec{\beta} = (\beta_1, \beta_2)$ and $\vec{\phi} = (\phi_1, \phi_2)$. Also since $\vec{\phi}$ is a parameter only related to $\vec{\epsilon_t}$, $\frac{\partial \vec{\mu_t}}{\partial \vec{\phi}} = 0$.

Denote l_{mt} as the summation of all margin score functions.

$$\frac{\partial l_{mt}}{\partial \omega_1} = \phi_1 \frac{\partial \mu_{1,t}}{\partial \omega_1}(\frac{x_{1,t} - \mu_{1,t}}{\mu_{1,t}^2}) \tag{3.14}$$

$$\frac{\partial l_{mt}}{\partial \alpha_1} = \phi_1 \frac{\partial \mu_{1,t}}{\partial \alpha_1}(\frac{x_{1,t} - \mu_{1,t}}{\mu_{1,t}^2}) \tag{3.15}$$

$$\frac{\partial l_{mt}}{\partial \beta_1} = \phi_1 \frac{\partial \mu_{1,t}}{\partial \beta_1}(\frac{x_{1,t} - \mu_{1,t}}{\mu_{1,t}^2}) \tag{3.16}$$

$$\frac{\partial l_{mt}}{\partial \phi_1} = \ln \phi_1 + 1 - \frac{\Gamma'(\phi_1)}{\Gamma(\phi_1)} + \ln \frac{x_{1,t}}{\mu_{1,t}} - \frac{x_{1,t}}{\mu_{1,t}} \tag{3.17}$$

$\frac{\partial l_{mt}}{\partial \omega_2}$, $\frac{\partial l_{mt}}{\partial \alpha_2}$, $\frac{\partial l_{mt}}{\partial \beta_2}$, and $\frac{\partial l_{mt}}{\partial \phi_2}$ are the same as above besides the index.

In the above equations, all the values are known except the derivatives of MEM parameters on $\mu_{k,t}$. Therefore the calculation of $\frac{\partial \mu_{1,t}}{\partial \omega_1}$ is discussed below. The calculation is very similar to the derivative of variance term in GARCH model.

According to $\mu_{k,t} = \omega_k + \alpha_k x_{i,t-1} + \beta_k \mu_{k,t-1}$, we have

$$\frac{\partial \mu_{k,t}}{\partial \omega_k} = 1 + \beta_k \frac{\partial \mu_{k,t-1}}{\partial \omega_k} \tag{3.18}$$

And if we assume that $\mu_{k,0}$ is a constant which does not depend on $\omega_k, k = 1, 2$, we have

$$\frac{\partial \mu_{k,0}}{\partial \omega_k} = 0 \tag{3.19}$$

Base on this assumption we can calculate $\frac{\partial \mu_{k,t}}{\partial \omega_k}$ for all t. $\frac{\partial \mu_{k,t}}{\partial \alpha_k}$ and $\frac{\partial \mu_{k,t}}{\partial \beta_k}$ can be calculated in the same way.

$$\begin{aligned}
\frac{\partial \mu_{k,t}}{\partial \alpha_k} &= x_{k,t} + \beta_k \frac{\partial \mu_{k,t-1}}{\partial \alpha_k} && (3.20)\\
\frac{\partial \mu_{k,0}}{\partial \alpha_k} &= 0 && (3.21)\\
\frac{\partial \mu_{k,t}}{\partial \beta_k} &= \mu_{k,t} + \beta_k \frac{\partial \mu_{k,t-1}}{\partial \beta_k} && (3.22)\\
\frac{\partial \mu_{k,0}}{\partial \beta_k} &= 0 && (3.23)
\end{aligned}$$

Derivatives of copula log-likelihood

Let l_{ct} denote the logarithm density of copula function at time t, the derivatives of copula log likelihood are:

$$\frac{\ln l_{ct}}{\partial\rho} = \frac{\rho}{1-\rho^2} - \frac{\tau+2}{2}\frac{[2\rho(q_{1,t}^2+q_{2,t}^2-2\rho q_{1,t}q_{2,t})-2(1-\rho^2)q_{1,t}q_{2,t}]}{(1-\rho^2)^2\tau+(1-\rho^2)(q_{1,t}^2+q_{2,t}^2-2\rho q_{1,t}q_{2,t})} \tag{3.24}$$

$$\begin{array}{rc} \frac{\partial l_{ct}}{\partial\tau} = & \frac{1}{2}\frac{\Gamma'(\frac{\tau+2}{2})}{\Gamma(\frac{\tau+2}{2})}+\frac{1}{2}\frac{\Gamma'(\frac{\tau}{2})}{\Gamma(\frac{\tau}{2})}-\frac{1}{2}\frac{\Gamma'(\frac{\tau+1}{2})}{\Gamma(\frac{\tau+1}{2})} \\ -\frac{1}{2} & \ln(1+\frac{q_{1,t}^2+q_{2,t}^2-2\rho q_{1,t}q_{2,t}}{(1-\rho^2)\tau}) \\ -\frac{\tau+2}{2} & \frac{-(q_{1,t}^2+q_{2,t}^2-2\rho q_{1,t}q_{2,t})+\tau(2q_{1,t}\frac{\partial q_{1,t}}{\partial\tau}+2q_{2,t}\frac{\partial q_{2,t}}{\partial\tau}-2\rho q_{2,t}\frac{\partial q_{1,t}}{\tau}-2\rho q_{1,t}\frac{\partial q_{2,t}}{\partial\tau})}{(1-\rho^2)\tau_2+(q_{1,t}^2+q_{2,t}^2-2\rho q_{1,t}q_{2,t})\tau} \\ +\frac{1}{2} & \ln(1+\frac{q_{1,t}^2}{\tau})+\frac{\tau+1}{2}\frac{2\tau q_{1,t}\frac{\partial q_{1,t}}{\partial\tau}-q_{1,t}^2}{\tau^2+\tau q_{1,t}^2}+\frac{1}{2}\ln(1+\frac{q_{2,t}^2}{\tau}) \\ +\frac{\tau+1}{2} & \frac{2\tau q_{2,t}\frac{\partial q_{2,t}}{\partial\tau}-q_{2,t}^2}{\tau^2+\tau q_{2,t}^2} \end{array} \tag{3.25}$$

$$\frac{\partial l_{ct}}{\partial\omega_1} = -\frac{\tau+1}{2}\frac{2q_{1,t}\frac{\partial q_{1,t}}{\partial\omega_1}-2\rho q_{2,t}\frac{\partial q_{1,t}}{\partial\omega_1}}{(1-\rho^2)\tau+q_{1,t}^2+q_{2,t}^2-2\rho q_{1,t}q_{2,t}}+\frac{\tau+1}{2}\frac{2q_{1,t}\frac{\partial q_{1,t}}{\partial\omega_1}}{\tau+q_{1,t}^2} \tag{3.26}$$

$\frac{\partial l_{ct}}{\partial\alpha_1}$,$\frac{\partial l_{ct}}{\partial\beta_1}$ and $\frac{\partial l_{ct}}{\partial\phi_1}$are similar to the formula above, therefore they are omitted here.

$$\frac{\partial l_{ct}}{\partial \omega_2} = -\frac{\tau+1}{2}\frac{2q_{2,t}\frac{\partial q_{2,t}}{\partial \omega_2} - 2\rho q_{1,t}\frac{\partial q_{2,t}}{\partial \omega_2}}{(1-\rho^2)\tau + q_{1,t}^2 + q_{2,t}^2 - 2\rho q_{1,t}q_{2,t}} + \frac{\tau+1}{2}\frac{2q_{2,t}\frac{\partial q_{2,t}}{\partial \omega_2}}{\tau + q_{2,t}^2} \tag{3.27}$$

In the same way the derivatives of l_{ct} against α_2, β_2 and ϕ_2 can be calculated, therefore they are omitted here.

In above formula the derivatives of $q_{k,t}$ can be further calculated as:

$$\frac{\partial q_{k,t}}{\partial \omega_k} = \frac{f(\frac{x_{k,t}}{\mu_{k,t}})x_{k,t}}{t_\tau(q_{k,t})\mu_{k,t}^2}\frac{\partial \tau}{\partial \omega_k} \tag{3.28}$$

$\frac{\partial q_{k,t}}{\partial \alpha_k}$ and$\frac{\partial q_{k,t}}{\partial \beta_k}$ are in the same form of $\frac{\partial q_{k,t}}{\partial \omega_k}$.

$$\frac{\partial q_{k,t}}{\partial \phi_k} = \frac{1}{t_\tau(q_{k,t})}\frac{\partial F_k(\frac{x_{k,t}}{\mu_{k,t}})}{\partial \phi_k} \tag{3.29}$$

Here$\frac{\partial F_k(\frac{x_{k,t}}{\mu_{k,t}})}{\partial \phi_k}$ and $\frac{\partial q_{k,t}}{\partial \tau}$ can only be calculated numerically.

Score functions

In conclusion, the derivatives of the likelihood, i.e., the score functions are tabulated in table 3.1, where θ_k denotes any one of ω_k,α_k or β_k,$k = 1, 2$, and:

$$A_{\theta_1} = \sum_{t=1}^{T}[\frac{(\tau+1)(q_{1,t}\frac{\partial q_{1,t}}{\partial \theta_1} - \rho q_{2,t}\frac{\partial q_{1,t}}{\partial \theta_1})}{(1-\rho^2)\tau + q_{1,t}^2 + q_{2,t}^2 - 2\rho q_{1,t}q_{2,t}} - \frac{(\tau+1)q_{1,t}\frac{\partial q_{1,t}}{\partial \theta_1}}{\tau + q_{1,t}^2}] \tag{3.30}$$

$$A_{\theta_2} = \sum_{t=1}^{T}[\frac{(\tau+1)(q_{2,t}\frac{\partial q_{2,t}}{\partial \theta_2} - \rho q_{1,t}\frac{\partial q_{2,t}}{\partial \theta_2})}{(1-\rho^2)\tau + q_{1,t}^2 + q_{2,t}^2 - 2\rho q_{1,t}q_{2,t}} - \frac{(\tau+1)q_{2,t}\frac{\partial q_{2,t}}{\partial \theta_2}}{\tau + q_{2,t}^2}] \tag{3.31}$$

$$A_{\phi_1} = \sum_{t=1}^{T}[\frac{(\tau+1)(q_{1,t}\frac{\partial q_{1,t}}{\partial \phi_1} - \rho q_{2,t}\frac{\partial q_{1,t}}{\partial \phi_1})}{(1-\rho^2)\tau + q_{1,t}^2 + q_{2,t}^2 - 2\rho q_{1,t}q_{2,t}} - \frac{(\tau+1)q_{1,t}\frac{\partial q_{1,t}}{\partial \phi_1}}{\tau + q_{1,t}^2}] \tag{3.32}$$

$$A_{\phi_2} = \sum_{t=1}^{T}[\frac{(\tau+1)(q_{2,t}\frac{\partial q_{2,t}}{\partial \phi_2} - \rho q_{1,t}\frac{\partial q_{2,t}}{\partial \phi_2})}{(1-\rho^2)\tau + q_{1,t}^2 + q_{2,t}^2 - 2\rho q_{1,t}q_{2,t}} - \frac{(\tau+1)q_{2,t}\frac{\partial q_{2,t}}{\partial \phi_2}}{\tau + q_{2,t}^2}] \tag{3.33}$$

Comparison against numerical derivatives

Table 3.5 shows the derivatives calculated by analytic derivatives and numerical derivatives separately. The data used is a randomly generated data set following equation (3.6). And the copula function is assumed to be student's t copula. A total of 100 repetitions are generated. And in each repetition the data length is 1000. Both the analytic and numerical derivatives use the true values of parameters. From the aspect of programming, the calculations of numerical derivatives are performed by "grad" function in the package "numDeriv" in R. The differences between numer-

ical and analytic derivatives are very small, especially the ones of ω, α, β and ρ. Noted that all of them have full analytic derivatives. Therefore the results verify equations for analytic derivatives are correct. In the estimation processes of ML and GMM methods, analytic derivatives are used.

3.3 GMM estimator

Cipollini et al. [2012] use GMM to estimate vector MEM semi-parametrically. Their method stems from the estimation function proposed in Cipollini et al. [2006]. This method uses less assumptions than ML method. Therefore it is suitable for a more generalized class of models in the vector MEM family as long as the only interest is on the MEM parameters.

In general GMM is computational efficient comparing to ML method. And the less the assumptions, the less likely the model is misspecified. In detail, GMM assumes only the structure of $\vec{\mu}_t$, and do not have any constraints on dependence structure among innovations, in addition, it does not require assigning a distribution for innovations. Thus the GMM estimator will only estimate MEM parameters θ, i.e. the parameters of $\vec{\mu}_t$. GMM involves

an approximation on correlation matrix before solving the GMM estimator. In Cipollini et al. [2012] a sample covariance matrix is used to approximate covariance matrix of innovations for given θ, which only holds when the first moment condition is satisfied. This fact indicates GMM may be sensitive to first moment condition.

The GMM estimator in Cipollini et al. [2012] is based on the conditional moment restriction: $E(\vec{\epsilon}_t|\mathbb{F}_{t-1} = 1)$. The moment restriction used for construct GMM estimator is $E(\vec{\epsilon}_t - \vec{1}|\mathbb{F}_{t-1}) = 0$, and the corresponding instrument variable is:

$$\vec{G}_t = \nabla_\theta \vec{\mu}_t' diag(\vec{\mu}_t)^{-1} \hat{\Sigma}^{-1} \tag{3.34}$$

In equation (3.34) $\hat{\Sigma} = \frac{1}{T}(\vec{\epsilon} - \vec{1})^T(\vec{\epsilon} - \vec{1})$ is an estimator of $\mathbb{V}(\vec{\epsilon}_t|\mathbb{F}_{t-1})$. Thus the GMM estimator $\hat{\theta}_{GMM}$ of vector MEM should satisfy the following equation:

$$\frac{1}{T}\sum_{t=1}^{T} \nabla_\theta \vec{\mu}_t' diag(\vec{\mu}_t)^{-1} \hat{\Sigma}^{-1}(\vec{\epsilon}_t - \vec{1}) = \vec{0} \tag{3.35}$$

By using $\epsilon_{i,t} = \frac{x_{i,t}}{\mu_{i,t}}, i = 1, \cdots, K$ equation (3.35) can be rewrit-

ten as:

$$\frac{1}{T}\sum_{t=1}^{T}\nabla_\theta\vec{\mu}_t'[diag(\vec{\mu}_t)\hat{\Sigma}diag(\vec{\mu}_t)]^{-1}(\vec{x}_t-\vec{\mu}_t)=\vec{0} \tag{3.36}$$

The optimal weight matrix which produces efficient estimator is approximated as [Hall, 2005]:

$$\begin{aligned}\hat{W}_T &= \mathbb{V}(T^{-\frac{1}{2}}\sum_{t=1}^{T}\nabla_\theta\vec{\mu}_t'diag(\vec{\mu}_t)^{-1}\hat{\Sigma}^{-1}(\vec{\epsilon}_t-\vec{1}))^{-1}\\ &= \frac{1}{T}\sum_{t=1}^{T}\nabla_\theta\vec{\mu}_t'[diag(\vec{\mu}_t)\hat{\Sigma}diag(\vec{\mu}_t)]^{-1}\nabla_\theta\vec{\mu}_t\end{aligned} \tag{3.37}$$

Thus the MEM parameters are estimated by solving the following optimization problem rather than solving a system of equations (5.4):

$$\hat{\theta}_{GMM}=\text{argmin}_{\,\theta\in\Theta}\sum_{t=1}^{T}(\vec{G}_t(\vec{\epsilon}_t-\vec{1}))'\times W_T\times\vec{G}_t(\vec{\epsilon}_t-\vec{1}) \tag{3.38}$$

where Θ is the feasible region for MEM parameters Cipollini et al. [2006]. The optimization process of GMM is found to converge faster than ML method in experiments, and thus less time consuming.

3.4 Comparing ML and GMM through simulations

In this section the performances of ML and GMM estimators are compared through a simulation study. In this simulation study, the data are generated by the two variable baisc MEM(1,1), same as the one in equation (3.6). And the innovations are assumed to follow Gamma distributions connected by student's t copula. The simulation contains two parts. The first one is a test on efficiency of both estimators, and the second part is on their robustness. In the first part the input data are generated by vector MEM, which is denoted as clean data. On the other hand in the second part the input data are artificially contaminated data.

There are two stylized facts not captured by ARCH type models. One is that the standardized residues from ARCH type models still display large kurtosis, leading to fat tail phenonmenon. For example, fat tail distribution of risk asset returns is a salient feature in financial time series. And the other sytlized fact is that large shock will seriously impact estimations of ARCH type models.

Thus the clean data generated by equation (3.6) are contaminated from two aspects. Firstly the structure of copula is changed to a mixture of student's t copula and independent copula. The purpose is to test when the tail dependence decreases and copula or marginal distributions are misspecified, whether the original model can capture the dependence structure correctly. Secondly margins are modified by adding outliers to the data generated from vector MEM. This contamination aims at simulating the situation where large shocks appear in data. And the performances of ML and GMM estimators under this situation are tested. The details of two kinds of contaminations are discussed below.

3.4.1 Generation of clean data

With the predetermined MEM parameters, two dimensional data (u, v) are randomly generated, which are in the range of (0,1), using student's t copula. u and v are the cdf of innovation terms $\epsilon_{1,t}$ and $\epsilon_{2,t}$. Then the cdf is inverted into Gamma variables using the corresponding gamma distribution function. Note that the invert functions are determined by the margin distribution assumptions of the vector MEM. After selecting a starting value for $x_{k,0}$ and

$\mu_{k,0}$, $k = 1, 2$, the series $\vec{x}_t$ and $\vec{\mu}_t$ series can be calculated using the formula for them in equation (3.6). The final output data are $\vec{x}_t$.

3.4.2 Data contamination

1 Change copulas

In order to test whether the model can capture dependence structure correctly when the distribution specified is different to the underlying distribution, ten sets of samples are generated. The copula function in these ten sets is a mixture of bivariate Student's t copula and independent copula (product copula). The cdf of copula can be written as:

$$C = wuv + (1 - w)C_{t;\rho,\tau}(u, v) \tag{3.39}$$

where $C_{t;\rho,\tau}$ is the CDF of Student's t copula with parameters ρ, τ. And w and $1 - w$ are the weights of product copula and Student's t copula respectively. When w equals zero, the copula is just Student's t copula. And when w equals one, the two margins are independent (extreme case). There are in total ten sets of data generated that corresponding to different w, where

$w \in (0.1, 0.2, \cdots, 1.0)$. Each set that is corresponding to a value of w contains 100 repetitions and each repetition includes a time series of dimension 1000×2.

2 Add outliers

Outliers are added to the clean data generated from vector MEM as shocks:

$x^*_{k,t} = x_{k,t} * d$ if $t = t_j$, $1 \leq j \leq l = [hT/100]$, else $x^*_{k,t} = x_{k,t}$

where h is the percentage of contamination, $x_{k,t}$ is the non-contaminated series generated by vMEM models (equation (3.6)). $t_1, \cdots, t_l$ are the time slots when the outliers are observed. The values t_j, $1 \leq j \leq l$, were chosen to be equally spaced.

In this outlier test three values, 3, 6 and 9, are chosen to be the sizes of outliers d. And the numbers of outliers are chosen to be 5% and 10% of the total number of clean data. In all there are 6 combinations of outliers with different numbers and sizes. For each combination 100 repetitions are included and each repetition contains a time series of dimension 1000×2.

3.4.3 Optimization

In the simulation it is found that when both ML and GMM estimators are optimized using a constrained optimization method, the efficiencies are not good in general. This is because when optimization is performed with boundary constraints, as long as the derivatives are not zero (or efficiently small), the "optimal point" tends to move to the boundaries. When we apply the model on real data or randomly generated data, it is very hard to guarantee the derivatives of optimal point to be small enough due to the fact that the sample is usually finite. Therefore the optimization is changed into a unconstrained one. If a variable is constrained in $(0, 1)$, MEM parameters for example, a logit transformation is used. Suppose p is between 0 and 1, the logit transformation of p is $q = \log(p/(1-p))$, and its inverse transformation is $\exp(p)/(1+\exp(q))$. If a variable must be positive, a logarithm transformation is used. In this way the constrained problem is changed into a unconstrained one.

3.4.4 Resutls on clean data

The estimation results of ML and GMM on clean data are tabulated in Table 3.6, only common parameters between the two estimators are reported for ease of comparison.

As it can be seen from Table 3.6, the estimated parameters using both estimators are very close to their true values and the variances of errors using both estimators are very small. This shows the performances of ML and GMM estimator are good when the distribution assumptions and moment conditions assigned to the model are accordant to the true distributions and moment conditions of the data. And the performance of ML method is a little better than that of GMM.

As an illustration of the fitting of estimated model on innovations, one repetition is chosen and innovations calculated by estimated parameters are compared with Gamma distribution with true ϕ. The innovations calculated using parameters estimated by ML estimator are depicted in Figure 3.1 and 3.2, corresponding to first margin and second margin separately. And those by GMM estimator are plotted in Figure 3.3 and 3.4, corresponding to first margin and second margin separately.

From Figures 3.1, 3.2, 3.3 and 3.4, it can be seen that the calculated innovations using both estimator are well fitted by Gamma distribution which indicates that the MEM parameters (ω, α, β) are correctly estimated.

In conclusion both estimators work well when the model is correctly specified.

3.4.5 Results on contaminated data

Misspecified copula

In the robustness test, the weight variable w is varied from zero to one, with zero corresponding to extreme case, i.e. independent copula, and one corresponding to Student's t copula. The means of dependencies between two margins estimated by the two estimators are reported in Table 3.7. The results show that when the dependence structure is misspecified, whether assume the dependence structure to be Student's t copula or do not assume any specific dependence structure, the dependencies between innovations can always be estimated accurately. When the weight w equals zero, the mean of estimated dependencies is -0.0004 for ML and -0.0006 for GMM, which are very close to zero. When

the weight w increases to a larger value thus in the mixture the Student's t copula has a larger weight, the estimated dependencies using both estimators show an increasing tendency to move closer to 0.6, which is the true value of the correaltion parameter in Student's t copula.

The detailed estimation results are reported in Table 3.8 for common parameters of the two estimators. The results show that when the copula is misspecified, not only the dependence can be captured accurately by both methods, but also the estimated parameters of MEM model are very close to true values. Comparing the MSE in Table 3.8 and those in Table 3.6, the results are almost the same for both methods. This result indicates that when the dependence structure is misspecified, estimations of neither method is heavily impacted. By assuming the dependence structure to be Student's t copula, ML estimator gains robustness because of the flexibility of Student's t copula. Also GMM gains robustness through its semi-parametric assumption, i.e. no specific distribution is assigned to the innovations but only the first moment conditions.

Outliers

The estimation results are summarized in Table 3.9. First observation is that the results in Table 3.9 indicates that outliers have significant impact on parameter estimations of both estimators. The values of MSE in Table 3.9 are much larger than the ones in Table 3.6 or Table 3.8.

The other observation is that the results are inconclusive on which estimator has a better performance. Neither of the two estimator has a good performances comparing to themselves in the clean data test or misspecified copula test. And also comparing between the two estimators, no evidences are found to support that one of them outperforms the other. When the percentage of ourliers in margin is 5%, GMM perform better than ML. For example when the size of outlier is 9, the MSE of β_2 is 0.0037 for GMM and 0.0267 for ML, which is almost 8 times larger than the one of GMM. When the percentage of outliers is 10%, ML method seems to have a better performance. For instances when the size of outliers is 6, the MSE of β_2 for GMM is 0.0604 and 0.0167 for ML method.

As a conclusion, both ML method and GMM are found to be

sensitive to outliers in margins. The estimation results of both of them are heavily influenced by outliers.

3.5 conclusion

In this chapter, two stage ML method is first introduced and used to estimate the starting values for ML and GMM. A comparison between ML and GMM is conducted through a simulation study. In the simulation study it is found that both estimators have a good performance when the dependence structure is misspecified. The copula used to generate the simulation data is the mixture of independent copula and Student's t copula. Changing the weight of independent copula in the mixture from zero to one, the estimation results of both estimators are not severely affected. However, when outliers exist in the margin, the Gamma distribution assumption in ML estimator and moment condition assumption in GMM estimator are both violated. Consequently the MSE of MEM parameters are much larger using both estimators comparing to themselves when data are not contaminated or only dependence structure is misspecified.

This phenomenon motives us to propose another method to

robustly estimate the parameters of vector MEM in the presence of outliers: Weighted Empirical Likelihood (WEL) method. This method can automatically detect outliers from innovations and the impact of outliers is reduced by adjusting the weights in the likelihood function. Therefore it is expected to be more robust than ML method and GMM when outliers exist in data. The details of WEL are discussed in Chapter 4.

parameter	Score function
θ_k(marginal)	$s_{\theta_k} = \sum_{t=1}^{T}[\phi_k \frac{\partial \mu_{k,t}}{\partial \theta_k}(\frac{x_{k,t}-\mu_{k,t}}{\mu_{k,t}^2}) - A_{\theta_k}], k = 1, 2$
ϕ_k(marginal)	$s_{\phi_k} = \sum_{t=1}^{T}[\ln \phi_k + 1 - \frac{\Gamma^{'}(\phi_k)}{\Gamma(\phi_k)} + \ln \frac{x_{k,t}}{\mu_{k,t}} - \frac{x_{k,t}}{\mu_{k,t}} - A_{\phi_k}]$
ρ(copula)	$s_\rho = \sum_{t=1}^{T}[\frac{\rho}{1-\rho^2} - \frac{\tau+2}{2}\frac{[2\rho(q_{1,t}^2+q_{2,t}^2-2\rho q_{1,t}q_{2,t})-2(1-\rho^2)q_{1,t}q_{2,t}]}{(1-\rho^2)^2\tau+(1-\rho^2)(q_{1,t}^2+q_{2,t}^2-2\rho q_{2,t}q_{2,t})}]$
τ(copula)	$s_\tau = \sum_{t=1}^{T}[\frac{1}{2}\frac{\Gamma^{'}(\frac{\tau+2}{2})}{\Gamma(\frac{\tau+2}{2})} + \frac{1}{2}\frac{\Gamma^{'}(\frac{\tau}{2})}{\Gamma(\frac{\tau}{2})} - \frac{1}{2}\frac{\Gamma^{'}(\frac{\tau-1}{2})}{\Gamma(\frac{\tau-1}{2})} - \frac{1}{2}\ln(1+\frac{q_{1,t}^2+q_{2,t}^2-2\rho q_{1,t}q_{2,t}}{(1-\rho^2)\tau})$ $-\frac{\tau+2}{2}\frac{-(q_{1,t}^2+q_{2,t}^2-2\rho q_{1,t}q_{2,t})+\tau(2q_{1,t}\frac{\partial q_{1,t}}{\partial \tau}+2q_{2,t}\frac{\partial q_{2,t}}{\partial \tau}-2\rho q_{2,t}\frac{\partial q_{1,t}}{\partial \tau}-2\rho q_{1,t}\frac{\partial q_{2,t}}{\partial \tau})}{(1-\rho^2)\tau^2+(q_{1,t}^2+q_{2,t}^2-2\rho q_{1,t}q_{2,t})\tau}$ $+\frac{1}{2}\ln(1+\frac{q_{1,t}^2}{\tau}) + \frac{\tau+1}{2}\frac{2\tau q_{1,t}\frac{\partial q_{1,t}}{\partial \tau}-q_{1,t}^2}{\tau^2+\tau q_{1,t}^2} + \frac{1}{2}\ln(1+\frac{q_{2,t}^2}{\tau}) + \frac{\tau+1}{2}\frac{2\tau q_{2,t}\frac{\partial q_{2,t}}{\partial \tau}-q_{2,t}^2}{\tau^2+\tau q_{2,t}^2}$

Table 3.1: derivatives of score function

parameter	ω_1	α_1	β_1	ϕ_1	ω_2	α_2	β_2	ϕ_2	ρ
True	0.02	0.11	0.7	1	0.05	0.10	0.78	1.1	0.6
Mean	0.020	0.109	0.702	1.00	0.061	0.120	0.780	1.102	0.612
Var	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.001

Table 3.2: Gaussian(repeat 100 times)

parameter	ω_1	α_1	β_1	ϕ_1	ω_2	α_2	β_2	ϕ_2	ρ	ν
True	0.02	0.11	0.7	1	0.05	0.10	0.78	1.1	0.6	2
Mean	0.020	0.109	0.697	1.000	0.060	0.121	0.781	1.10	0.595	2.17
Var	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009

Table 3.3: Student's t(repeat 100 times)

parameter	ω_1	$\alpha_{1,1}$	$\alpha_{1,2}$	$\beta_{1,1}$	$\beta_{1,2}$	ϕ_1	ω_2	$\alpha_{2,1}$	$\alpha_{2,2}$	$\beta_{2,1}$	$\beta_{2,2}$	ϕ_2	ρ	τ
True	1.5	0.3	0.2	0.55	0.2	1	1	0.002	0.01	0.003	0.1	1	0.6	5
Mean	1.4601	0.2997	0.2144	0.5457	0.2594	1.007	0.8999	0.0019	0.0180	0.0025	0.1802	1.006	0.5982	5.199
Var	0.3526	0.0018	0.0542	0.0026	0.1590	0.0018	0.09529	$5.3e-6$	0.0005	$1.3e-5$	0.0725	0.0020	0.0006	1.864

Table 3.4: Full matrix two step estimation)

	ω_1	α_1	β_1	ω_2	α_2	β_2	ρ
ML	0.0002	0.0002	0.0007	0.0006	0.0005	0.0007	0.0005
GMM	0.0002	0.0003	0.0008	0.0006	0.0006	0.0008	0.0008

Table 3.6: MSE of two methods on clean data, true values $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2, \rho) = (0.05, 0.15, 0.8, 0.1, 0.35, 0.6, 0.6)$

Parameters (true value)	Analytic	Numeric
$\omega_1(0.02)$	-0.5748450516	-0.5748457042
$\alpha_2(0.15)$	-0.0368254517	-0.0368254821
$\beta_1(0.6)$	-0.0465226878	-0.0465227371
$\phi_1(0.5)$	0.0185003594	0.0185003860
$\omega_2(0.15)$	0.0674829037	0.0674830201
$\alpha_2(0.05)$	0.0569829442	0.0569830290
$\beta_20.8$	0.0537566592	0.0537567562
$\phi_2(0.5)$	-0.0027189557	-0.0027189860
$\rho(0.6)$	0.0089000310	0.0089000222
$\tau(2)$	-0.0006436333	-0.0006436308

Table 3.5: derivatives

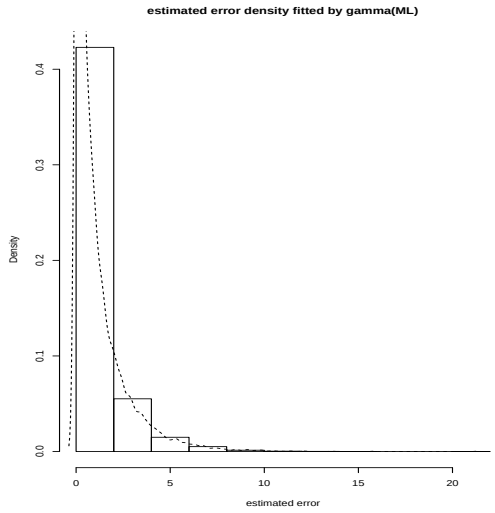

Figure 3.1: Innovations of first margin VS Gamma distribution (ML)

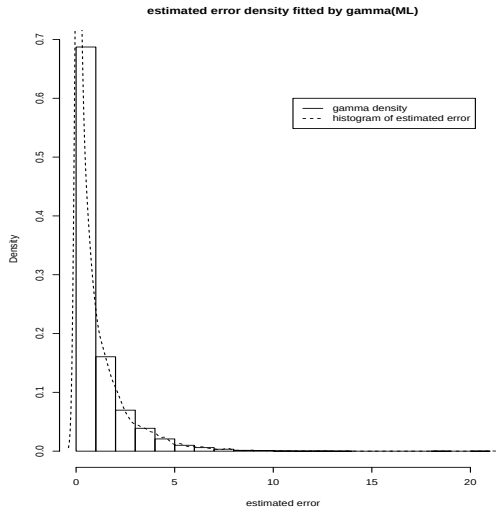

Figure 3.2: Innovations of second margin VS Gamma distribution (ML)

weight	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
ML	−0.0004	0.0132	0.0360	0.1048	0.1852	0.2803	0.3866	0.4858	0.5541	0.5792
GMM	−0.0006	0.0128	0.0398	0.1048	0.1873	0.2799	0.3804	0.4736	0.5400	0.5704

Table 3.7: mean of estimated dependencies

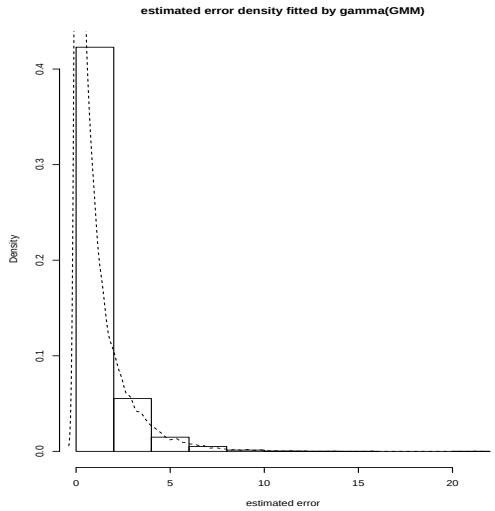

Figure 3.3: Innovations of first margin VS Gamma distribution (GMM)

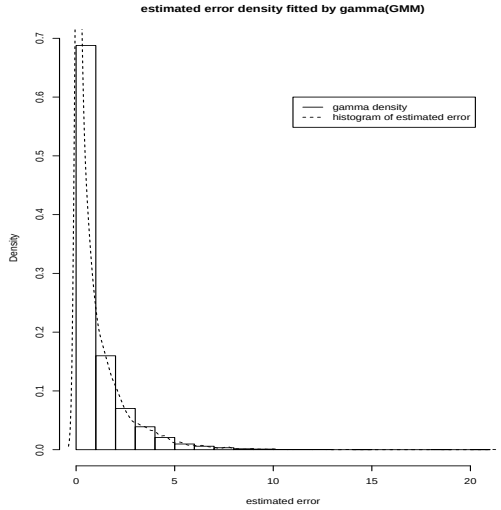

Figure 3.4: Innovations of second margin VS Gamma distribution (GMM)

weight	parameter	ω_1	α_1	β_1	ω_2	α_2	β_2
0	ML	0.0004	0.0004	0.0010	0.0007	0.0008	0.0012
	GMM	0.0004	0.0004	0.0010	0.0007	0.0008	0.0012
0.1	ML	0.0004	0.0005	0.0012	0.0007	0.0009	0.0013
	GMM	0.0004	0.0005	0.0012	0.0007	0.0009	0.0013
0.2	ML	0.0005	0.0003	0.0011	0.0092	0.0007	0.0013
	GMM	0.0005	0.0003	0.0010	0.0074	0.0007	0.0013
0.3	ML	0.0004	0.0004	0.0011	0.0007	0.0006	0.0010
	GMM	0.0004	0.0004	0.0011	0.0007	0.0006	0.0010
0.4	ML	0.0004	0.0005	0.0015	0.0008	0.3935	0.0608
	GMM	0.0004	0.0004	0.0014	0.0008	0.0668	$1.4e-4$
0.5	ML	0.0003	0.0004	0.0010	0.0005	0.0008	0.0011
	GMM	0.0003	0.0004	0.0010	0.0005	0.0008	0.0011
0.6	ML	0.0004	0.0004	0.0011	0.0007	0.0006	0.0011
	GMM	0.0004	0.0004	0.0011	0.0007	0.0006	0.0010
0.7	ML	0.0003	0.0004	0.0008	0.0005	0.0006	0.0008
	GMM	0.0336	0.0005	0.0334	0.0005	0.0007	0.0008
0.8	ML	0.0003	0.0003	0.0010	0.0007	0.0005	0.0009
	GMM	0.0003	0.0003	0.0010	0.0008	0.0006	0.0010
0.9	ML	0.0002	0.0003	0.0006	0.0006	0.0004	0.0007
	GMM	0.0002	0.0003	0.0007	0.0006	0.0005	0.0008

Table 3.8: MSE of ML method and GMM when weight changes. True values $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2, \rho) = (0.05, 0.15, 0.8, 0.1, 0.35, 0.6, 0.6)$

percent of outliers		5			10		
size of outliers		$d=3$	$d=6$	$d=9$	$d=3$	$d=6$	$d=9$
ω_1	ML	0.1057	0.0407	0.0194	0.1013	0.0497	0.0874
	GMM	0.0164	0.0092	0.0078	0.2297	0.1175	0.1489
α_1	ML	0.0007	0.0014	0.0042	0.0008	0.0016	0.0069
	GMM	0.0006	0.0006	0.0013	0.0014	0.0015	0.0041
β_1	ML	0.0782	0.0240	0.0129	0.0627	0.0211	0.0198
	GMM	0.0136	0.0069	0.0064	0.1469	0.0533	0.0548
ω_2	ML	0.1901	0.1990	0.1538	0.4425	0.1301	1.625
	GMM	0.0240	0.0268	0.0221	0.2679	0.5589	0.7038
α_2	ML	0.0015	0.0038	0.0099	0.0021	0.0042	0.0140
	GMM	0.0017	0.0016	0.0008	0.0028	0.0022	0.0025
β_2	ML	0.0339	0.0301	0.0267	0.0669	0.0167	0.0832
	GMM	0.0063	0.0055	0.0037	0.0486	0.0604	0.0573

Table 3.9: MSE of ML and GMM when outliers exist in margins. True values $(\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2, \rho) = (0.05, 0.15, 0.8, 0.1, 0.35, 0.6, 0.6)$

□ End of chapter.

Chapter 4

Weighted Empirical Likelihood Estimator

Vector Multiplicative Error Model (vector MEM) is capable of analyzing and forecasting multidimensional non-negative valued processes. Usually its parameters are estimated by Generalized Method of Moments (GMM) and Maximum Likelihood (ML) methods. However, the estimations could be heavily affected by outliers. To overcome this problem, in this chapter an alternative approach, Weighted Empirical Likelihood (WEL) method, is proposed. This method uses moment conditions as constraints and the outliers are detected automatically by performing a k-means clustering on Oja depth values of innovations. The performance

of WEL is evaluated against those of GMM and ML methods through extensive simulations, in which three different kinds of additive outliers are considered. Besides WEL, Empirical Likelihood method with equal weights is also considered in the simulation study. By comparing the results of WEL and EL, the role of weights calculated by depth function and k-means clustering can be examined. Moreover, the robustness of WEL is demonstrated by comparing the volatility forecasts of the three methods on 10-minute returns of S&P 500 index. The results from both the simulations and the S&P 500 volatility forecasts have shown preferences of using WEL method. Please refer to Ding and Lam [2012] for the journal version of this chapter.

4.1 Introduction

Multiplicative error model (MEM) has attracted much interest in the past decade because it is capable of modeling and forecasting non-negative financial data, for example realized volatility of exchange rates [Engle, 2002], duration and trading volumes [Manganelli, 2005]. One of the areas in which MEM plays an important role is analyzing and forecasting volatilities. Due to the

non-negative nature of volatilities, GARCH models are less suitable for directly analyzing the dynamics of volatilities. Moreover since true volatilities are not observable, different volatility proxies such as squared return, high low range and realized volatility have been used. The problem with proxies is that with only one of them, it may not be enough to encapsulate all the information contained in volatilities. This concern gives rise to extending MEM into vector MEM [Engle and Gallo, 2006, Cipollini et al., 2006]. The study in Cipollini et al. [2006] shows that although vector MEM is a straightforward generalization from univariate MEM, the estimation of vector MEM parameters is not.

Two estimators have been applied on vector MEM: Maximum Likelihood (ML) estimator and Estimating Functions estimator [Cipollini et al., 2006]. The latter estimator is further developed into Generalized Method of Moments (GMM) estimator in Cipollini et al. [2012]. ML estimator is built on the parametric multivariate density function of innovations, which is usually Gamma distributions linked by normal copula [Cipollini et al., 2006, 2007] or Student's t copula [Cipollini et al., 2007]. On the other hand GMM estimator is constructed based on the assump-

tion that the expectations of innovations are equal to a vector of one. Thus GMM estimator avoids choosing a specific multivariate distribution, which is necessary for ML estimator, and the probability of model misspecification is reduced. However, if the data contain outliers, the distribution assumption for ML estimator and the moment assumption for GMM estimator will both be violated. As a result, ML and GMM are susceptible to outliers.

Outliers are not uncommon in financial time series. They can be a result of external influences such as policy changes, news or recording errors. For example, the current financial crisis has caused a considerable number of unusual movements of stock indices in 2008, which are suspected as outliers. The existence of outliers and their impact on ML and GMM estimators motivate us to find a more robust method to estimate the parameters of vector MEM. In this chapter an alternative approach based on Empirical Likelihood method [Owen, 1988] is proposed: Weighted Empirical Likelihood (WEL) method. Empirical likelihood combines the advantages of ML and GMM methods. It allows the parameters to be estimated by likelihood method, without the requirement of a specific distribution assumption. Also it does not require the cal-

culation of the inverse of the weight matrix in GMM. The weighted version of EL method is first studied in Glenn and Zhao [2007], where it is used for non-parametrically estimating the means of a sample from contaminated normal density. There are two characteristics that differentiate the WEL estimator proposed in this chapter from the one in Glenn and Zhao [2007].

The first characteristic is that in this chapter, WEL utilizes moment conditions for vector MEM [Cipollini et al., 2012] as constraints to estimate parameters of a multivariate nonlinear model. Because of these moment conditions WEL estimator is semi-parametric. By contrast, in Glenn and Zhao [2007] WEL is non-parametric. Comparing to ML estimator, this semi-parametric feature is preferred when the underlying distribution of innovations is unknown or contaminated by outliers.

The second and more important characteristic is that different from Glenn and Zhao [2007], in this chapter the weights assigned to empirical likelihood functions are automatically calculated by Oja depth function [Oja, 1983] and k-means clustering. Observations are classified into two groups, outliers and non-outliers, according to the depth values of innovations. Since the number

of outliers is unknown, one main difficulty in outlier detection is to find a suitable threshold value to identify outliers. Rather than choosing this threshold value artificially, k-means clustering is used to automatically separate depth values into two groups corresponding to outliers and non-outliers respectively. And only the weights of outliers are adjusted in empirical likelihood functions. When the data does not contain outliers, the outlier group would only contain a very small number of extreme observations, and the impact of adjusting the weights on these observations is not significant on parameter estimation. Therefore our estimator allows tilting the likelihood function to lessen the influences of outliers, which will increase the robustness of WEL comparing to ML and GMM estimators.

As a result of these two characteristics the performance of the proposed WEL estimator is expected to be better than those of ML and GMM estimators under the situation that outliers exist in data. Vector MEM and the two estimators, ML and GMM, are introduced in Section 4.2, followed by Section 4.3 where the details of WEL estimator are explained. To investigate the performances of the three estimators, a simulation study is conducted in Section

4.4 where different scenarios of additive outliers are considered. In Section 4.7 the three estimators are further compared through an empirical application on 10-minute returns of S&P 500 index, including data collected during the current financial crisis period. The last section is conclusion of this chapter.

4.2 Vector multiplicative error model and two estimation methods

In this section vector MEM and two estimators, ML and GMM, that have previously been discussed in Chapter 2 are briefly reviewed.

Let $\vec{x}_t = (x_{1,t}, \cdots, x_{K,t})^T$ be a K-dimensional nonnegative time series. Vector MEM of order $(1,1)$ can be written as

$$\vec{x}_t = diag(\vec{\mu}_t)\vec{\epsilon}_t \tag{4.1}$$

$$\vec{\mu}_t = \vec{\omega} + A\vec{x}_{t-1} + B\vec{\mu}_{t-1} \tag{4.2}$$

$$\vec{\epsilon}_t|\mathbb{F}_{t-1} \sim D(\vec{1}, \Sigma) \tag{4.3}$$

In equation (4.1), $diag(\vec{\mu}_t)$ is a diagonal matrix whose diagonal elements are $\vec{\mu}_t$. The parameters in equation (4.2), i.e. $\vec{\omega} \in \mathbb{R}^K$, $A \in \mathbb{R}^{(K,K)}$ and $B \in \mathbb{R}^{(K,K)}$, are MEM parameters. These MEM parameters are the main concern of analyzing and forecasting the dynamics of $\vec{\mu}_t$. If both A and B are diagonal, the model is the basic form of vector MEM(1,1). And the 1-dimensional margins of $\vec{x}_t$ influence each other through $\vec{\epsilon}_t$ but not $\vec{\mu}_t$. On the other hand, if at least one of A and B is not diagonal, the interactions between margins of $\vec{x}_t$ are reflected by those between $\vec{\epsilon}_t$ and also between $\vec{\mu}_t$. In this chapter our interest lies in reducing the impact of outliers on parameter estimations but not estimating the interactions between margins of $\vec{x}_t$, therefore only the basic form of vector MEM is considered in the simulation study. In the empirical example both circumstances that A is diagonal and A is not are considered.

In equation (4.3), D is the joint distribution of $\vec{\epsilon}_t$. The expectations of $\vec{\epsilon}_t$ are restricted to $\vec{1}$. And Σ is the variance covariance matrix of $\vec{\epsilon}_t$. The specifications for $\vec{x}_t$ and $\vec{\mu}_t$ are same in ML and GMM estimators. The assumptions for D are where these two estimators differ.

ML estimator requires a specified joint distribution for innovations. In Cipollini et al. [2006, 2007] and Cipollini and Gallo [2010], the joint distribution is decomposed into one dimensional margin distributions and a copula, which is used to describe the dependence structure between margins. In this chapter Gamma distribution and Student's t copula are chosen to construct the ML estimator. As mentioned in Engle [2002], Gamma distribution is more flexible than exponential distribution, which is first used in univariate MEM. And Student's t copula generalizes normal copula by allowing non-zero dependence in extreme tails through different degrees of freedom. The combination of Gamma distribution and Student's t copula is also used in Cipollini et al. [2007] and Cipollini and Gallo [2010]. Following this assumption, denote $c_t(R, \tau)$ as Student's t copula function with correlation matrix R and degree of freedom τ, and $F_\phi(x)$ and $f_\phi(x)$ as the cdf and pdf of Gamma distribution respectively with a shape parameter ϕ and a scale parameter $\frac{1}{\phi}$, the density of innovations at time t can be written as:

$$c_{t,R,\tau}(F_{\phi_1}(x_{1,t}/\mu_{1,t}),\ldots,F_{\phi_k}(x_{K,t}/\mu_{K,t}))\times\prod_{k=1}^{K}f_{\phi_k}(x_{k,t}/\mu_{k,t}) \quad (4.4)$$

Thus when ML estimator is used, the MEM parameters are obtained by solving the following maximization problem [Cipollini et al., 2007]:

$$\hat{\theta}_{ML}=\text{argmax}_{\,\theta\in\Theta}\sum_{t=1}^{T}(\ln c_{t,R,\tau}(F_{\phi_1}(x_{1,t}/\mu_{1,t}),\ldots,F_{\phi_K}(x_{K,t}/\mu_{K,t}))+\sum_{k=1}^{K}\ln f_{\phi_k}(\frac{x_{k,t}}{\mu_{k,t}})) \quad (4.5)$$

where Θ is the feasible region for MEM parameters [Cipollini et al., 2006].

On the other hand, the GMM estimator proposed in Cipollini et al. [2012] is constructed with three components: the moment conditions, instrument variables and optimal weight matrix. The moment conditions are:

$$\mathbb{E}(\vec{\epsilon}_t|\mathbb{F}_{t-1})=\vec{1} \quad (4.6)$$

$$\mathbb{V}(\vec{\epsilon}_t|\mathbb{F}_{t-1})=\Sigma \quad (4.7)$$

where the covariance matrix Σ is unknown.

The instrument variables $\{\vec{G}_t\}_{t=1}^{T}$ and optimal weight matrix $\hat{W}_T$ corresponding to equation (4.6) are:

$$\vec{G}_t = \nabla_\theta \vec{\mu}_t^T diag(\vec{\mu}_t)^{-1} \hat{\Sigma}^{-1} \tag{4.8}$$

$$\begin{aligned} \hat{W}_T &= (\mathbb{V}(T^{-\frac{1}{2}} \sum_{t=1}^{T} \vec{G}_t(\vec{\epsilon}_t - \vec{1})))^{-1} \\ &= \frac{1}{T} \sum_{t=1}^{T} \nabla_\theta \vec{\mu}_t^T [diag(\vec{\mu}_t) \hat{\Sigma} diag(\vec{\mu}_t)]^{-1} \nabla_\theta \vec{\mu}_t \end{aligned} \tag{4.9}$$

where $\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} \vec{\mu}_t \vec{\mu}_t^T$ is the approximation of the unknown covariance matrix of innovations.

The GMM estimations of MEM parameters can be obtained by solving the optimization problem:

$$\hat{\theta} = \text{argmin}_{\,\theta \in \Theta} \sum_{t=1}^{T} (\vec{G}_t(\vec{\epsilon}_t - \vec{1}))^T \times W_T \times \vec{G}_t(\vec{\epsilon}_t - \vec{1}) \tag{4.10}$$

where Θ is the feasible region for MEM parameters [Cipollini et al., 2006].

From equation (4.4), we can see that ML estimator requires choosing a specific margin distribution and copula before estimating MEM parameters. This may cause difficulties because for multivariate model it is often hard to write down a tractable likelihood function when the underlying distribution is unknown. And if the true distribution differs from the one assumed, para-

metric estimator, for example ML estimator, may lead to biased estimations. On the other hand, GMM estimator proposed by Cipollini et al. [2012] only imposes moment restrictions on innovations. Thus GMM is much preferred due to the semi-parametric property. However GMM requires estimating the optimal weight matrix, which is usually calculated via a two step optimization procedure or substituted by an approximation. Further the outliers may contaminate the moment conditions and drive the estimated parameters away from true values. Therefore an alternative semi-parametric estimator, weighted empirical likelihood (WEL) estimator, is proposed in this chapter. WEL is expected to be more robust to outliers. And its details are discussed in the following section.

4.3 Weighted Empirical Likelihood

In this section WEL estimator for vector MEM is introduced. The proposed estimator is a combination of EL method, moment conditions and depth function. WEL method can be considered as a generalization of EL method, in which equal weights are assigned to each empirical likelihood. The rationale for imposing unequal

weights in WEL estimator is that the empirical likelihood functions can be considered as density of innovations. As a result, the values of empirical likelihood functions which are corresponding to outliers should be decreased. The weights in WEL estimator are calculated based on depth function. Depth function serves as a measure of the distance between innovations and their center. And the innovations whose distances to center are larger than others are recognized as outliers. Therefore weights based on depth function can be used to reduce the impact of outliers on empirical likelihood function.

The same moment conditions used in Cipollini et al. [2012] for GMM estimator are employed in this chapter. And the assumptions for innovations are

$$\mathbb{E}(\vec{\epsilon}_t|\mathbb{F}_{t-1}) = \vec{1} \tag{4.11}$$

Following the notations used in Section 4.2, the weighted log empirical likelihood can be written as:

$$f(\vec{p},\theta) = \{\sum_{t=1}^{T} w_t log(p_t) | p_t \geq 0, \sum_{t=1}^{T} p_t = 1, \sum_{t=1}^{T} p_t \vec{G}_t(\vec{\epsilon}_t - \vec{1}) = 0\} \tag{4.12}$$

where $\vec{\epsilon}_t$ is innovation vector, p_t and w_t are the empirical likelihood and weight for $\vec{\epsilon}_t$ respectively, and $\vec{G}_t$ is the instrument variable. As discussed in Owen [2001], maximizing equation (4.12) is equivalent to minimizing the Lagrange dual function:

$$L(\eta, \vec{\lambda}, \vec{p}, \theta) = \sum_{t=1}^{T} w_t \log p_t - \eta(\sum_{t=1}^{T} p_t - 1) - \vec{\lambda}^T \sum_{t=1}^{T} p_t \vec{G}_t(\vec{\epsilon}_t - \vec{1}) \tag{4.13}$$

where $\eta \in \mathbb{R}$ and $\vec{\lambda} \in \mathbb{R}^{K+2K^2}$ are Lagrange multipliers. Using the first order conditions of equation (4.13) and $\sum_{t=1}^{T} w_t = 1$, it can be shown that (the details of the optimization process are discussed in the following section):

$$\eta = 1 \tag{4.14}$$

$$p_t = \frac{w_t}{1 + \vec{\lambda}^T \vec{G}_t(\vec{\epsilon}_t - \vec{1})} \tag{4.15}$$

$$\vec{\lambda} = \text{argmin}_{\vec{\lambda} \in \mathbb{R}^{K+2K^2}} - \sum_{t=1}^{T} w_t \log(1 + \vec{\lambda}^T \vec{G}_t(\vec{\epsilon}_t - \vec{1})) \tag{4.16}$$

Without moment constraints, $\sum_{t=1}^{T} w_t \log(p_t)$ is maximized at $p_t = w_t, t = 1, \cdots, T$. Therefore following Glenn and Zhao [2007], the WEL estimation of parameters θ is obtained by maximizing the weighted empirical likelihood ratio:

$$\begin{aligned} \hat{\theta}_{WEL} &= \text{argmax}_{\theta \in \Theta} \sum_{t=1}^{T} w_t \log(\frac{p_t}{w_t}) \\ &= \text{argmax}_{\theta \in \Theta} \sum_{t=1}^{T} w_t \log(\frac{1}{1 + \vec{\lambda}^T \vec{G}_t(\vec{\epsilon}_t - \vec{1})}) \end{aligned} \quad (4.17)$$

where $\{w_t\}_{t=1}^{T}$ are weights calculated by depth function, $\vec{\lambda}$ are Lagrange multipliers obtained from equation (4.16), $\{\vec{G}_t\}_{t=1}^{T}$ are instrument variables used in GMM estimator (equation (4.8)) and $\vec{\epsilon_t}$ are innovation vectors.

The optimization of empirical likelihood function is discussed in Owen [2001]. Basically the same procedure is used in the optimization of WEL. As can be seen from equation (4.17), there are two sets of parameters in the log empirical likelihood function. The parameters in the first set are MEM parameters θ, which are the main concern of the optimization. The parameters in the second set are Lagrange multipliers $\vec{\lambda}$. Therefore the optimiza-

tion process can be separated into inner optimization (estimate $\vec{\lambda}$ assuming θ is fixed) and outer optimization (estimate θ). The weights are calculated after inner optimization and before outer optimization. Therefore the inner optimization is

$$\hat{\vec{\lambda}} = \text{argmin}_{\vec{\lambda}\in\mathbb{R}^{K+2K^2}} - \sum_{t=1}^{T} w_t \log(1 + \vec{\lambda}^T \vec{G}_t(\vec{\epsilon}_t - \vec{1})) \qquad (4.18)$$

and the outer optimization is

$$\hat{\theta} = \text{argmax}_{\theta\in\Theta} - \sum_{t=1}^{T} w_t \log(1 + \vec{\lambda}^T(\theta) \vec{G}_t(\vec{\epsilon}_t - \vec{1})) \qquad (4.19)$$

where $\vec{\lambda}(\theta)$ are estimated Lagrange multipliers in inner optimization.

Equation (4.19) can be maximized via normal routines like quasi-Newton algorithms or derivative-free algorithms. Equation (4.18) on the other hand is more complicated than equation (4.19) because $p_t = \frac{w_t}{1+\vec{\lambda}^T \vec{G}_t(\vec{\epsilon}_t - \vec{1})}$ needs to be constrained in $(0, 1)$. The details of inner optimization is discussed in the following section.

4.3.1 Inner optimization

The inner optimization intends to estimate $\{p_t\}_{t=1}^T$ when θ is fixed. If T is large then too many parameters are estimated at the same time. To reduce the number of parameters, Lagrange dual of the original problem is considered.

The Lagrange function associated with equation (4.12) is:

$$L(\eta, \vec{\lambda}, \vec{p}, \theta) = \sum_{t=1}^{T} w_t \log p_t - \eta(\sum_{t=1}^{T} p_t - 1) - \vec{\lambda}^T \sum_{t=1}^{T} p_t \vec{G}_t(\vec{\epsilon}_t - \vec{1}) \tag{4.20}$$

where $\eta \in \mathbb{R}^1$ and $\vec{\lambda} \in \mathbb{R}^{K+2K^2}$ are Lagrange multipliers.

Note in inner optimization θ are fixed, thus the first order conditions of lagrange function is:

$$\frac{\partial L(\eta, \vec{\lambda}, \vec{p}, \theta)}{\partial p_t} = \frac{w_t}{p_t} - \eta - \vec{\lambda}^T \vec{G}_t(\vec{\epsilon}_t - \vec{1}) = 0, t = 1, \ldots, T \tag{4.21}$$

$$\frac{\partial L(\eta, \vec{\lambda}, \vec{p}, \theta)}{\partial \eta} = -(\sum_{t=1}^{T} p_t - 1) = 0 \tag{4.22}$$

$$\frac{\partial L(\eta, \vec{\lambda}, \vec{p}, \theta)}{\partial \lambda_i} = -\sum_{t=1}^{T} p_t G_{i,t}(\epsilon_{i,t} - 1) = 0, \quad i = 1, \ldots, K + 2K^2 \tag{4.23}$$

where $G_{i,t}$ and $\epsilon_{i,t}$ are the ith elements of vector $\vec{G}_t$ and $\vec{\epsilon}_t$ respec-

tively. If (4.21) holds, then

$$\sum_{t=1}^{T} p_t \frac{\partial L(\eta, \vec{\lambda}, \vec{p}, \theta)}{\partial p_t} = 0 \tag{4.24}$$

From equation (4.21),(4.22), (4.23) ,(4.24) and $\sum_{t=1}^{T} w_t = 1$ we have

$$1 - \eta = 0$$

Therefore

$$\eta = 1 \tag{4.25}$$

$$p_t = \frac{w_t}{1+\vec{\lambda}^T \vec{G}_t(\vec{\epsilon}_t - \vec{1})} \tag{4.26}$$

$$\sum_{t=1}^{T} p_t G_{i,t}(\epsilon_{i,t} - 1) = 0, i = 1, \ldots, K + 2K^2 \tag{4.27}$$

Observations from equation 4.25-4.27 are that if we substitute p_t in equation 4.27 with equation 4.26, then we have

$$\sum_{t=1}^{T} \frac{w_t G_{i,t}(\epsilon_{i,t} - 1)}{1 + \vec{\lambda}^T \vec{G}_t(\vec{\epsilon}_t - \vec{1})} = 0, i = 1, \ldots, K + 2K^2 \tag{4.28}$$

This is exactly the first order condition of the following uncon-

strained optimization problem:

$$\min_{\vec{\lambda}\in\mathbb{R}^{\frac{K(K+1)}{2}}} -\sum_{t=1}^{T} w_t \log(1+\vec{\lambda}^T\vec{G}_t(\vec{\epsilon}_t-\vec{1})) \tag{4.29}$$

Therefore the first order conditions 4.21, 4.22 and 4.23 for $L(\eta,\vec{\lambda},\vec{p},\theta)$ are solved by:

$$\eta = 1 \tag{4.30}$$

$$p_t = \frac{w_t}{1+\vec{\lambda}^T\vec{G}_t(\vec{\epsilon}_t-\vec{1})} \tag{4.31}$$

$$\vec{\lambda} = \text{argmin}_{\vec{\lambda}\in\mathbb{R}^{K+2K^2}} -\sum_{t=1}^{T} w_t \log(1+\vec{\lambda}^T\vec{G}_t(\vec{\epsilon}_t-\vec{1})) \tag{4.32}$$

One choice of calculating $\vec{\lambda}$ is solving a system of equations (equation (4.28)). Thus it is possible to solve $\vec{\lambda}$ for any given θ. Further we have the same number of variables as the number of equations, therefore the solution is unique. This method is impractical because even in the current step θ may be close to their true values, there is no guarantee the unique solution of $\vec{\lambda}$ can result in a positive joint density p_t. This is especially the case when the starting values are far from the true values. Thus we do not solve the equations directly in inner optimization but rather transfer them into a minimization problem (4.29).

The problem that p_t may be out of feasible region is solved by monitoring its value in each step of inner optimization. Since $p_t = \frac{w_t}{1+\vec{\lambda}^T\vec{G}_t(\vec{\epsilon}_t-\vec{1})}$ and $w_t, p_t \in (0,1)$, we have $1+\vec{\lambda}^T\vec{G}_t(\vec{\epsilon}_t-\vec{1}) \in (0,\infty)^K$, for $t = 1,\ldots,T$. In each step if $1+\vec{\lambda}^T\vec{G}_t(\vec{\epsilon}_t-\vec{1})$ is smaller than a small constant, the step size is divided by two. In this paper this small constant is chosen as $\frac{1}{T}$.

To sum up, the function to be minimized in inner optimization is:

$$-\sum_{t=1}^{T} w_t \log(1+\lambda^T\vec{G}_t(\vec{\epsilon}_t-\vec{1})) \tag{4.33}$$

And its first and second order derivatives against λ are:

$$\begin{aligned} &\frac{d}{d\lambda_i}(-\sum_{t=1}^{T} w_t \log(1+\vec{\lambda}^T\vec{G}_t(\vec{\epsilon}_t-\vec{1}))) \\ = \; &-\sum_{t=1}^{T} \frac{w_t G_{i,t}(\epsilon_{i,t}-1)}{1+\vec{\lambda}^T\vec{G}_t(\vec{\epsilon}_t-\vec{1})}, i = 1,\ldots,K+2K^2 \end{aligned} \tag{4.34}$$

$$\begin{aligned} & \frac{d^2}{d\lambda_i d\lambda_j}(-\sum_{t=1}^{T} w_t \log(1+\vec{\lambda}^T\vec{G}_t(\vec{\epsilon}_t-\vec{1}))) \\ = & \sum_{t=1}^{T} \frac{w_t G_{i,t}(\epsilon_{i,t}-1)G_{j,t}(\epsilon_{j,t}-1)}{1+\vec{\lambda}^T\vec{G}_t(\vec{\epsilon}_t-\vec{1})}, \\ & i,j=1,\ldots,K+2K^2 \end{aligned} \tag{4.35}$$

4.3.2 Calculation of weights

An ideal weight function should have the following properties: firstly it can assign small values to innovations far from the center; secondly it can be calculated non-parametrically.

These preferred properties can be achieved by using depth function. Depth values can rank multivariate data from center to outward non-parametrically. For a multivariate data set, its center will receive the highest depth value. The farther away the observations are from the center, the lower the depth values. In the past decade, depth function has received growing interest in areas such as multi-dimensional exploratory data analysis and inference [Zuo and Cui, 2005, Zuo et al., 2004, Massé, 2009] and outlier detection [Gervini, 2011].

The depth function used in this chapter is Oja depth(OD).

Oja depth is developed in Serfling and Zuo [2000] from a location measure proposed by Oja [1983]. For a given set of innovations, $\{\vec{\epsilon}_t\}_{t=1}^T$, the depth value d at $\vec{\epsilon}_0 \in \mathbb{R}^K$ is

$$d(\vec{\epsilon}_0; \{\vec{\epsilon}_t\}_{t=1}^T) = \binom{T}{K}^{-1} (1 + \sum_{1 \le t_1 \le \cdots \le t_K \le T} v(S[\vec{\epsilon}_0, \vec{\epsilon}_{t_1}, \cdots, \vec{\epsilon}_{t_K}]))^{-1} \tag{4.36}$$

In equation (4.36), $S[\cdot]$ is a closed simplex composed of vertices $(\vec{\epsilon}_0, \vec{\epsilon}_{t_1}, \cdots, \vec{\epsilon}_{t_K})$ and $v(S[\cdot])$ is the volume of simplex S.

After calculating the depth values, a function is needed to transfer them to weights. This is because the summation of depth values is not necessarily one, and also the depth values should be divided into two groups. If the depth value is smaller than a certain value, resulting the detection of an outlier, the corresponding weight is adjusted. The weight function introduced in Zuo et al. [2004] is used in this chapter:

$$WT(d) = (\exp(-J(1 - (\frac{d}{C})^2)^2) - \exp(-J))I(d < C) + I(d \ge C) \tag{4.37}$$

In the weight function (4.37), d represents Oja depth, $I(\cdot)$ is

indicator function and $C \in (0, 1)$, $J \in (0, \infty)$ are range and shape parameters separately. C indicates the range of depth values in which the weight is adjusted. For example, if $C = 0.5$, it means all innovations with depth values smaller than 0.5 are identified as outliers and their corresponding weights are reduced. And if an innovation's depth value is larger than 0.5, its weight is 1. The shape parameter J represents how fast the weight decreases when the depth gets smaller. When J is very large, say more than 100, the weights of all the outliers are very close to 0. This means that the outliers are neglected in the estimation process. On the other hand, when J is small, only a few outliers with the lowest depth values are neglected. The weights of the other outliers are between 0 and the weights of non-outliers. In our simulation study we set J as 2. The reason to choose 2 is that when an observation is classified as outliers, the weight of the corresponding empirical likelihood function is not reduced to zero immediately. The reduction on weight is decided by how small the depth value is comparing to depth values of other outliers. The advantage of reducing the weight gradually is that the impact of misclassification is lessened. When the data does not contain outliers, the

outliers group would contain a small number of extreme observations, their weights are only slightly adjusted because their depth values are not as low as outliers. Therefore WEL can estimate the parameters accurately when data do not contain outliers. Figure 4.1 and Figure 4.2 plot the the weight function for different values of C and J.

The weights used in (4.17) can be written as:

$$\widetilde{w}_t = WT(d(\vec{\epsilon}_t; \{\vec{\epsilon}_t\}_{t=1}^T)) \tag{4.38}$$

$$w_t = \frac{\widetilde{w}_t}{\sum_{t=1}^T \widetilde{w}_t} \tag{4.39}$$

where equation (4.39) guarantees the summation of weights is 1.

From equation (4.38), it can be seen that the weights $\{w_t\}_{t=1}^T$ depend on innovation vectors, and also the range parameter C. In practice C can be considered as a benchmark value, and it is used to determine the bar of depth values under which their corresponding weights will be adjusted. One way to calculate C is setting it as a certain quantile of depth values. However this involves choosing a cut off percentage artificially, and it is difficult to find a suitable percentage when the number of outliers is unknown. In the present chapter the cut off value C is chosen

automatically by using k-means clustering.

Since firstly depth function transfers the outlyingnesses of multi-dimension data into one dimension measure: depth value. And secondly the depth value represents the distance to the center, i.e. smaller values are more likely to be associated with outliers. K-means clustering is used to divide the one dimension depth values into two groups. They are outlier group and non-outlier (ordinary) group respectively. The group with a larger valued mean contains depth values of the ordinary innovations. And the group with the smaller valued mean contains those of outliers. The cut off value C is set as the largest depth of innovations in the outlier group. And smaller weights are assigned to the log empirical likelihood corresponding to the outlier group. Although some extreme innovations may be classified as outliers, through the simulation we find the number of miss-classified outliers is small and the influences on parameter estimation are insignificant.

4.4 Simulation study on outliers

In this section the performances of WEL, ML and GMM estimators are compared through simulations. Firstly the three es-

timators are tested using clean data (data generated by vector MEM without outliers). Secondly they are compared under three circumstances of data contamination. In each circumstance a different kind of outliers is added to the clean data.

The model used to generate the clean data is a $2-$dimensional basic Vector MEM$(1,1)$, where the joint distribution of innovations is assumed to be two $1-$dimensional Gamma distributions linked by Student's t copula. The combination of Gamma distribution and Student's t copula is also used in Cipollini et al. [2007] and Cipollini and Gallo [2010].

$$\begin{aligned} x_1 &= \mu_{1,t}\epsilon_{1,t} \\ x_2 &= \mu_{2,t}\epsilon_{2,t} \\ \vec{\epsilon_t} &\sim c(\vec{\epsilon_t}, 0.6, 5) \cdot \prod Gamma(10) \\ \mu_{1,t} &= \omega_1 + \alpha_1 x_{1,t-1} + \beta_1 \mu_{1,t-1} \\ \mu_{2,t} &= \omega_2 + \alpha_2 x_{2,t-1} + \beta_2 \mu_{2,t-1} \end{aligned} \tag{4.40}$$

Equation (4.40) is the vector MEM that generates the clean data. In equation (4.40) $c(\vec{\epsilon_t}, 0.6, 5)$ stands for the density of Stu-

dent't copula, whose correlation parameter is 0.6 and degree of freedom is 5. And $Gamma(10)$ is 1-dimensional Gamma distribution with a scale parameter 10 and a shape parameter 0.1. The MEM parameters $\theta = (\omega_1, \alpha_1, \beta_1, \omega_2, \alpha_2, \beta_2)$ are set as $(0.05, 0.1, 0.85, 0.01, 0.15, 0.8)$.

The evaluation method used in the simulation study is mean square error (MSE). MSE of the kth MEM parameter $\theta(k) \in \theta$ is defined as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{\theta}_i(k) - \theta_0(k))^2 \tag{4.41}$$

where N is the number of repetitions, $\hat{\theta}_i(k)$ is the estimated MEM parameter for the ith repetition and $\theta_0(k)$ is the true value of $\theta(k)$.

4.4.1 Clean data

The performance of a robust estimator should be good whether the input data contain outliers or not. Therefore before testing the robustness of the three estimators against outliers, their performances are compared using clean data.

The clean data are generated by equation (4.40) and contain 500 repetitions of 2−dimensional basic MEM (1,1) series. And in each repetition the size of the MEM series is 2×1000. The esti-

mations of the three estimators are reported in Table 4.1. We can see that when data does not contain outliers, the performances of all the three estimators are good. The results of GMM and WEL estimators are very close and their differences on each parameter are all smaller than 0.0001. The similar performances between WEL and GMM estimators are expected, because they employ the same moment conditions and very few observations are classified as outliers by WEL estimator. By comparing the results of ML estimator with the results of GMM and WEL estimators, it can be seen that most of the MSE values are same except the ones of ω_1 and β_1 of ML estimator are slightly larger than those of GMM and WEL. The reason is that when both parameter matrices A and B are diagonal, there are 8 MEM parameters to estimate. However for ML estimator there are additional parameters for Gamma distribution and Student's t. Therefore ML estimator has 12 parameters to estimate, which is about 50% more than the number of parameters of GMM and WEL. As a result the optimization result of ML estimator is slightly worse than those of GMM and WEL estimators.

4.4.2 Outliers

According to Engle [2002] the square roots of 1-dimensional MEM input variables, $\{\sqrt{\vec{x}_t}\}$, can be considered as variables following a zero mean GARCH model. Therefore MEM is isomorphic to GARCH model, and adding outliers to the clean data generated by MEM is similar to the situation where the clean data are generated by the GARCH model. Following Mulera and Yohai [2008] and Chalabi and Wuertz [2010], a portion of the clean data used in Section 4.4.1 are replaced by the true expectations multiplied by the size of the outliers.

Since multivariate input data are used in the simulation study, different from univariate data, outliers may only appear in some of the margins. Also the appearances of the outliers can follow a periodic or random pattern. These phenomena motivate us to design three kinds of additive outliers: outliers that simultaneously appear in every margin periodically; outliers that simultaneously appear in every margin, but the time intervals between consecutive outliers are random; and outliers that do not necessarily appear in every margin simultaneously nor do they appear in a periodic pattern. The details of the three scenarios are explained

below.

Scenarios

The first kind of outliers is equidistant-in-time outliers, which means outliers appear in a periodic pattern. The outliers appear in $x_{1,t}$ and $x_{2,t}$ simultaneously following the equation:

$$x^*_{k,t} = \begin{cases} s \cdot \mu_{k,t} & \text{if } t = t_i, 1 \leq i \leq l = [hT/100], k = 1, 2 \\ x_{k,t} & \text{else}, k = 1, 2. \end{cases}$$

where h is the percentage of contamination, $x_{k,t}$ is the clean data used in Section 4.4.1 and $\mu_{k,t}$ is its conditional expectation. $t_1, \ldots, t_l$ are the time slots in which original data are replaced by outliers. The values t_i, $1 \leq i \leq l$, are equally spaced. The size of outliers s is chosen from the set $(3, 6, 9)$. And the total numbers of outliers, l, are 50 and 100. They are corresponding to the percentages h, which are equals to 5 and 10 respectively. When there are 5% outliers out of 1000 pairs of observations, one outlier will appear in every 20 observations. This periodic pattern indicates they are monthly outliers. While in the situation of h equals 10, outliers will appear in every 10 observations. Therefore they are biweekly

outliers.

The second kind of outliers is designed to appear simultaneously in both margins but time intervals between two consequent outliers are random. They are generated according to the following equation:

$$x^*_{k,t} = x_{k,t} * (1 - b_t) + s \cdot \mu_{k,t} \cdot b_t, k = 1, 2$$

where $s \in (3, 6, 9)$ is the size of outliers and b_t is an i.i.d. series following Bernoulli distribution. The numbers of outliers are 5% and 10% out of 1000 observations. Random variables b_t corresponding to 5% and 10% outliers follow $Bernoulli(0.05)$ distribution and $Bernoulli(0.1)$ distribution respectively.

The third kind of outliers is completely random outliers in terms that the time intervals between consecutive outliers in the same margin are random and further outliers in different margins are unrelated. They are generated according to the following equation

$$\begin{aligned} x^*_{1,t} &= s \cdot \mu_{1,t} \cdot b_{1,t} + x_{1,t} * (1 - b_{1,t}) \\ x^*_{2,t} &= s \cdot \mu_{2,t} \cdot b_{2,t} + x_{2,t} * (1 - b_{2,t}) \end{aligned}$$

where $s \in (3, 6, 9)$ is the size of outliers and $b_{k,t}, k = 1, 2$ are i.i.d. Bernoulli distributed variables. Two numbers of outliers are considered. When the percentage of outliers is 5%, $b_{k,t} \sim Bernoulli(0.05), k = 1, 2$. And when the percentage of outliers is 10%, $b_{k,t} \sim Bernoulli(0.1), k = 1, 2$.

After adding outliers to the clean data, the MEM parameters are estimated based on $\{x^*_{1,t}, x^*_{2,t}\}_{t=1}^{T}$. Different combinations of sizes and numbers of outliers used in the simulation are summarized in Table 4.2. And each combination is repeated 500 times.

4.4.3 Simulation results

The results for data containing outliers of scenario 1 are displayed in Table 4.3. From the results we can see that the performance of the WEL estimator is the best among the three estimators. Comparing to ML estimator, the MSE of WEL and GMM estimators are often ten times smaller, which shows the advantage of semi-parametric assumption of WEL and GMM. As to the comparison between WEL and GMM, although they use the same moment conditions, WEL estimator considerably reduces the MSE of MEM parameters ω_1, ω_2, β_1 and β_2. For example, when the size of

outliers is 9 and the percentage of outliers is 10%, the MSE of β_2 corresponding to WEL is 0.0091. While the value corresponding to GMM is 0.0573, which is more than 5 times larger than 0.0091. The MSE of α_1 and α_2 are similar between WEL and GMM. This is because these two parameters are less impacted by outliers and there is little room for WEL estimator to improve.

Table 4.4 reports the results of the three estimators in scenario 2. The results show that the performances of GMM and WEL are similar to each other and are better than ML. Comparing the results of Table 4.3 and 4.4, we can see that the parameters' MSE of all the three estimators are larger in scenario 1. This observation shows that when outliers appear simultaneously in both margins, periodic outliers have more influences on the estimations of MEM parameters than random outliers.

Results of scenario 3 are tabulated in Table 4.5. Among the three estimators WEL is least affected by outliers and achieves the best performance. For instance, when data contain 10% outliers of size 9, MSE of β_2 using WEL is 0.0105 where those using ML and GMM are 0.4642 and 0.1611 respectively. The substantial reductions in MSE show that the impact of outliers is successfully

decreased using WEL estimator. The results in Table 4.5 also suggest that comparing to themselves, the performances of the three estimators are worse when outliers are completely random in scenario 3 than when they appear simultaneously in scenario 1 and 2. This phenomenon indicates the three estimators may not be feasible when the outliers in different margins are independent to each other. However if the input data are different volatility measures of one time series, this problem is not significant because the outliers tend to appear simultaneously.

To conclude, WEL estimator produces the smallest MSE among all the three estimators and is more stable than the other two estimators. When the size of outliers increases, the parameter estimations of WEL are least affected. On the other hand, estimations of ML and GMM differ significantly as the size becomes larger. Another observation from the results is that the estimations are less affected by outliers that appear simultaneously in each margin. For the simultaneously appeared outliers, periodic outliers have more influences on parameter estimation than the outliers that do not have a periodic pattern.

4.5 Computations of high dimension vector MEM

In this section, the computation issues of ML, GMM and WEL methods when the dimension K is large are discussed.

4.5.1 The influences of dimension on ML

When the dimension K increases, the first problem is the increase of number of parameters, which is referred to "the curse of dimension" in literatures. For a K dimension variable, there are $2 * K^2 + K$ MEM parameters. As can be seen the number of MEM parameters increases rapidly as K increases. Beside the MEM parameters, different from GMM and WEL methods, ML method also need to estimate the covariance matrix,the Gamma parameters and the degree of freedom parameter. Therefore the increase of the number of parameters affects ML method most among the three methods.

In ML method, the covariance matrix of innovations are required to be estimated. And for K dimension variables, the number of parameters in covariance matrix is $K(K+1)/2$. When $K = 2$, there are 10 MEM parameters, which is the parameters need to be estimated by all the three methods. And ML method

needs to estimate the Gamma parameters and also the covariance matrix and degree of freedom in Student's t copula. Therefore ML method has 14 parameters to estimate, on the other hand GMM and WEL methods have 10 parameters to estimate. When $K = 3$, there are 18 MEM parameters to estimated. The parameters in the covariance matrix increase to 6. As a result, GMM and WEL methods need to estimate 21 parameters and ML method needs to estimate 31 parameters, which is almost 50% more than the number of parameters in GMM and WEL methods. When $K = 4$, there are 36 MEM parameters and ML method needs to estimate 51 parameters.

In order to reduce the number of parameters, the MEM parameter $\vec{\omega}$ is expressed as $(I - A - B)E(X)$ for all three methods. In this way in all the three methods there are only $2 * K^2$ MEM parameters.

Generally, the larger the number of parameters, the less accurate the estimations are. Thus when K is large, ML method is expected to have the worst estimation results among the three methods.

4.5.2 The influences of dimension on GMM

Although the impact of the increase of parameters is not as severe on GMM as on ML, in the optimization process GMM requires the calculation of derivatives of MEM parameters and the weight matrix which could cause the optimization process slow and the estimation results inaccurate. In some cases it may even make the optimization process crash.

The core function for GMM is the moment condition:

$$\vec{G}_t(\vec{\mu}_t - \vec{1}) = \nabla_\theta \vec{\mu}_t^T (diag(\vec{\mu}_t)\hat{\Sigma} diag(\vec{\mu}_t))^{-1}(\vec{x}_t - \vec{\mu}_t) \quad (4.42)$$

$$\hat{W}_T = (\mathbb{V}(T^{-\frac{1}{2}} \sum_{t=1}^{T} \vec{G}_t(\vec{\epsilon}_t - \vec{1})))^{-1}$$

$$= \frac{1}{T} \sum_{t=1}^{T} \nabla_\theta \vec{\mu}_t^T [diag(\vec{\mu}_t)\hat{\Sigma} diag(\vec{\mu}_t)]^{-1} \nabla_\theta \vec{\mu}_t \quad (4.43)$$

In the above equation, θ refers to the parameter matrix A and B. By using the equation $\vec{\omega} = (I - A - B)E(X)$, ω can be represented by A and B, therefore MEM parameters only contains A and B. $\vec{G}_t$ is the moment condition at time t, and it is $2 * K^2$ by 1 vector. On the right hand side of equation (4.42), $\nabla_\theta \vec{\mu}_t$ is the vector contains the derivatives of MEM parameters on the con-

ditional expectation vector, and it is a $2 * K^2 \times K$ matrix. As K increases, the size of this matrix increases quickly. The rest part of right hand side of equation (4.42), $(diag(\vec{\mu}_t)\hat{\Sigma}diag(\vec{\mu}_t))^{-1}$ is a $K \times K$ matrix, the size of this matrix does not increase as fast as the others therefore the impact is small. In conclusion, for equation (4.42) the dimension is:

$$[2 * K^2, K] \times [K, K] \times [K, 1] \tag{4.44}$$

From equation (4.44) it can be seen the most difficult part in this equation is calculating the derivatives of MEM parameters on conditional expectations.

However, the most difficult part of GMM estimation is the calculation of the inverse of $\hat{W}_T$, which is the weight matrix. Because the object function in the GMM optimization process involves $\hat{W}_T^{-1}$, the calculation of the inverse of weight matrix is essential for GMM method. The dimension of each component in equation (4.43) is:

$$[2 * K^2, K] \times [K, K] \times [K, 2 * K^2] \tag{4.45}$$

As a result of equation (4.43), the size of weight matrix is $2 * K^2 \times 2 * K^2$. Although the number of parameters to estimate is smaller for GMM than that for ML method, when K increases, the calculation difficulty is also increased mainly because of the calculation of inverse weight matrix.

4.5.3 The influences of dimension on WEL

Unlike ML method, the parameters in WEL method are only the MEM parameters. And also different from GMM method, in WEL method the weight matrix is not required, only equation (4.42) is involved. However, WEL is also affected by the increase of dimension K.

For WEL method, the main difficulty lies in the calculation of depth values for innovations. Oja depth for innovations of dimension K and length T is defined be the following equation:

$$d(\vec{\epsilon}_0; \{\vec{\epsilon}_t\}_{t=1}^T) = \binom{T}{K}^{-1} (1 + \sum_{1 \le t_1 \le \cdots \le t_K \le T} v(S[\vec{\epsilon}_0, \vec{\epsilon}_{t_1}, \cdots, \vec{\epsilon}_{t_K}]))^{-1} \tag{4.46}$$

In the above equation, the depth value of $\vec{\epsilon}_0$ is calculated based

on the data set $\{\vec{\epsilon}_t\}_{t=1}^{T}$. In order to calculate the Oja depth of $\vec{\epsilon}_0$, the volumes of all the $K+1$ vertices simplices are calculated. Therefore a total number of $\binom{T}{K}$ simplices are considered. When T or K increases, the calculation burden rapidly increases. When K or T increases, the number of simplices increases following the equation:

$$\binom{T}{K+1} = \binom{T}{K} * \frac{T-K}{K+1} \tag{4.47}$$

$$\binom{T+1}{K} = \binom{T}{K} * \frac{T+1}{T-K+1} \tag{4.48}$$

Usually K is much smaller than T, therefore $\frac{T-K}{K+1}$ is much larger than $\frac{T+1}{T-K+1}$. This means that the increase in K has a much larger impact on calculation burden of Oja depth than the increase in T.

4.5.4 Simulation

When the dimension K increases, the accuracy of all the three methods, ML, GMM and WEL, is expected to decrease comparing to the case when K is small. In this section three values of K are chosen, and they are K=3, K=4 and K=5. And all the MEM

parameter matrices are chosen to be diagonal. Each data set contains 200 samples generated from vector MEM, and in total 100 data sets are generated.

Following this set up, when $K = 3$ there are 16 parameters in ML method and 9 parameters in GMM and WEL methods. The results are tabulated in Table 4.10. The results show that the three methods have very similar performances. And among the MEM parameters, the estimations on α are closest to true values. The estimated ω are farthest to true values.

When $K = 4$, there are 23 parameters in ML methods and 12 parameters in WEL and GMM methods. A similar observation can be seen from results in Table 4.11: all the three methods estimate α better than ω and β. Another observation is that ML method has a worse performance than WEL and GMM, whose results are close to each other. The performance of ML method can be explained by the increasing of parameters. Comparing to $K = 3$, there are only 3 more parameters for GMM and WEL method. On the other hand, the parameters increase from 16 to 23 for ML method.

When $K = 5$, there are 10 parameters for WEL and GMM

methods and 26 parameters for ML method. The simulation results are displayed in Table 4.12. Same to the case where $K = 4$, GMM and WEL methods have similar performances and both of them outperform ML method.

In general for all the three methods the estimated parameters are fairly close to their true values. However comparing to the situation where $K = 2$ in Table 4.1, the MSE is much larger when K increases. Although the sample length is different for the case $K = 2$ and the cases $K = 3, 4, 5$, the differences between the results are very large, which shows the impact of K on parameter estimation. In practice when K is very large, certain dimension reduction techniques may be used to reduced the dimension, for example principle component analysis.

4.6 Compare weighted empirical likelihood and empirical likelihood

In this section, the un-weighted version of empirical likelihood (EL), equal weight empirical likelihood, is compared with WEL method.

The same object function and optimization process of WEL

can be followed in EL. The object function of EL when moment condition in GMM method is used is:

$$f(\vec{p},\theta) = \{\sum_{t=1}^{T} \frac{1}{T} log(p_t) | p_t \geq 0, \sum_{t=1}^{T} p_t = 1, \sum_{t=1}^{T} p_t \vec{G}_t(\vec{\epsilon}_t - \vec{1}) = 0\} \tag{4.49}$$

where $\vec{p}$ are the empirical likelihood, $\vec{G}_t$ are moment conditions used in GMM method (equation (4.42)), and $\vec{\epsilon}_t$ are innovations at time t.

By using the Lagrange multipliers, the optimization problem (4.49) can be transformed to:

$$\eta = 1 \tag{4.50}$$

$$p_t = \frac{1}{T} \frac{1}{1+\vec{\lambda}^T \vec{G}_t(\vec{\epsilon}_t - \vec{1})} \tag{4.51}$$

$$\vec{\lambda} = \text{argmin}_{\vec{\lambda} \in \mathbb{R}^{K+2K^2}} - \frac{1}{T} \sum_{t=1}^{T} \log(1 + \vec{\lambda}^T \vec{G}_t(\vec{\epsilon}_t - \vec{1})) \tag{4.52}$$

where $\eta, \vec{\lambda}$ are Lagrange multipliers.

Without moment constraints, $\sum_{t=1}^{T} \frac{1}{T} \log(p_t)$ is maximized at $p_t = \frac{1}{T}, t = 1, \cdots, T$. Therefore the equal weight EL estimation of parameters θ is obtained by maximizing the empirical likelihood

ratio:

$$\begin{aligned}\hat{\theta}_{EL} &= \text{argmax}_{\theta\in\Theta}\frac{1}{T}\sum_{t=1}^{T}\log(\frac{p_t}{\frac{1}{T}}) \\ &= \text{argmax}_{\theta\in\Theta}\frac{1}{T}\sum_{t=1}^{T}\log(\frac{1}{1+\vec{\lambda}^T\vec{G}_t(\vec{\epsilon}_t-\vec{1})}) \quad (4.53)\end{aligned}$$

The EL method is applied to the same data sets used in Section 4.4. For ease of comparison, the results are tabulated together with WEL, ML and GMM methods in Table 4.1, Table 4.3, Table 4.4 and Table 4.5 corresponding to clean data test, outlier scenario 1, scenario 2, and scenario 3 respectively.

In clean data test, EL has a good performance. And the results are similar to the other three methods. In the outlier test, comparing the results of EL to those of GMM when the sizes and the numbers of outliers are small, EL method has a very similar performance to GMM and WEL methods, especially for the estimation on α. For example, when in scenario 1, the size of outliers is 3 and the number of outliers is 50 out of 1000, the MSE of α_1 and α_2 for EL are 0.0006 and 0.0018, the values for GMM are 0.0006 and 0.0017, and the values for WEL are 0.0006 and

0.0018.

However, when the size and number of outliers increase, EL estimations are heavily affected by outliers. For example, in scenario 3 when the size of outliers is 9 and the number of outliers is 100 out of 1000, the MSE of β_1 and β_1 for EL are 0.1851 and 0.1395, the values for GMM are 0.1564 and 0.1611, and the values for WEL are 0.1159 and 0.0105.

In general EL method have a similar performance to GMM, and a similar performance to WEL when the size and number of outliers are small. When there are more outliers in the data and their sizes are larger, WEL outperforms EL. However, due to the semi-parametric property, same to GMM, EL method outperforms ML method in all three scenarios.

4.7 Empirical example

In this section the forecasts of the three estimators on Standard and Poor's composite stock index (S&P 500) are compared. An interesting observation from Section 4.4 is that the performances of the three estimators are better when the outliers tend to appear simultaneously in $x_{1,t}$ and $x_{2,t}$ (scenario 1 and 2). Therefore

two closely related volatility proxies, realized volatility (RV) and realized bipower volatility (BV), are chosen as input variables for vector MEM. The two volatility series are calculated from 10-minute returns of S&P 500 observed from 15 August 2005 to 16 March 2011, consisting 1386 days. And each day contains 39 intraday returns from 9:30 AM to 16:00 PM. Following Cipollini et al. [2007], RV and BV are defined as:

$$RV_t = \sum_{i=1}^{39} r_{t,i}^2 \tag{4.54}$$

$$BV_t = \frac{\pi}{2} \sum_{i=2}^{39} r_{t,i} r_{t,i-1} \tag{4.55}$$

In equation (4.54) and (4.55) $r_{t,i}$ is the ith return in day t. RV and BV are all scaled to have the same mean 1. The statistics of BV and RV are summarized in Table 5.1. And the two series are plotted in Figure 4.3

One observation from Table 5.1 is that the maximum values of RV and BV are considerably larger than the mean of both series. Also visual observations from Figure 4.3 show unusual fluctuations occur immediately after September 2008, the month in which the

current financial crisis begins. Both observations indicate a high possibility that outliers exist in the two series.

Based on the observations we separate the series into two parts, corresponding to in-sample period and out-of-sample period. The in-sample period contains the first 1200 out of 1386 observations (from 15 August 2005 to 21 June 2010) and the out-of-sample period contains the rest 186 observations (from 22 June 2010 to 16 March 2011). In this way the in-sample period covers all the days on which abnormal values are observed in Figure 4.3. Thus the impact of outliers will be reflected on the parameter estimations using in-sample data. Since the true values of the parameters are unknown, out-of-sample forecasts are used to compare the impact on different estimators.

4.7.1 Model

The filter applied on (RV_t, BV_t) is a two variable MEM(1,1). The general form is defined as:

$$RV_t = \mu_t^{RV} * \epsilon_t^{RV} \tag{4.56}$$

$$BV_t = \mu_t^{BV} * \epsilon_t^{BV} \tag{4.57}$$

$$\begin{pmatrix} \mu_t^{RV} \\ \mu_t^{BV} \end{pmatrix} = \begin{pmatrix} \omega^{RV} \\ \omega^{BV} \end{pmatrix} + A \begin{pmatrix} RV_{t-1} \\ BV_{t-1} \end{pmatrix} + B \begin{pmatrix} \mu_{t-1}^{RV} \\ \mu_{t-1}^{BV} \end{pmatrix} \tag{4.58}$$

where $A = [a_{i,j}]_{2\times 2} \in \mathbb{R}^{2\times 2}$,$B = [b_{i,j}]_{2\times 2} \in \mathbb{R}^{2\times 2}$ and $\epsilon_t^{RV}, \epsilon_t^{BV}$ are i.i.d. innovations whose expectations equal 1.

Two different configurations of the filter are considered in this example. Filter one assumes both parameter matrices A and B are diagonal, while filter two assumes only B is diagonal. Filter one leaves all the interactions between RV and BV into innovations. While in filter two besides the interactions between innovations, interdependence of BV and RV is also described by the linear relationships between BV_{t-1} and μ_t^{RV} and the ones between RV_{t-1} and μ_t^{BV} as well. The assumptions for innovations $(\epsilon_t^{RV}, \epsilon_t^{BV})^T$are the same as the ones in Section 5.4: $(\epsilon_t^{RV}, \epsilon_t^{BV})^T$ are assumed to follow Gamma distributions linked by Student's t copula in ML es-

timator; WEL and GMM assume the expectations of $(e_t^{RV}, e_t^{BV})^T$ are equal to $(1,1)^T$.

4.7.2 Forecast comparison criteria

The forecasts using the three estimators are evaluated by Mincer-Zarnowitz regression. Mincer-Zarnowitz regression is based on the linear model:

$$(\sum_{j=0}^{k-1} V(t+j)) = c+b_{wel}\cdot(\sum_{j=0}^{k-1} \hat{V}_{wel}(t+j))+b_{model}\cdot(\sum_{j=0}^{k-1} \hat{V}_{model}(t+j))+e_t \tag{4.59}$$

Where $V(t)$ is the value of RV or BV series at time t, $\hat{V}_{wel}(t+j)$ is the j-day ahead forecast calculated using WEL parameters and $\hat{V}_{model}(t+j)$ is the j-day ahead forecast calculated using ML or GMM parameters. And the j-day ahead forecast is the summation of all the forecasts from day $i+1$ to day $i+j$ based on the information up to day i. The forecast horizon k is set as two values, 1 and 5. And $\{e_t\}_{t=1}^{T-k+1}$ is an $i.i.d.$ series that follows normal distribution. By restricting b_{model} or b_{wel} to zero, the regression tests the explanation power of the regressor. The

forecasts from different estimators can be evaluated by comparing the coefficients of the regressors in the linear model. If the null hypothesis $b_{wel} = 1$ or $b_{model} = 0$ cannot be rejected, namely b_{wel} is significantly larger than 0 or b_{model} is not significantly different from 0, then volatilities forecasted by WEL contain more information than those forecasted by the other estimator and therefore the WEL forecasts are superior.

4.7.3 Results

The parameter estimations are tabulated in Table 4.7. From the results we can see that ML and GMM estimators tend to underestimate B and overestimate A comparing to WEL.

The regression results are summarized in Table 4.8 (in-sample forecasts) and Table 4.9 (out-of-sample forecasts). The R square corresponding to WEL is always the highest among all the one variable regressions, both in-sample and out-of-sample. This result indicates that the volatility forecasts of WEL have more explanation power than the ones of ML and GMM. What is more in most two variable regressions, b_{wel} is statistically different from zero with 5% level of significance and more closer to 1 than b_{model}.

Also the R square does not improve much when the volatility forecasts using ML or GMM are included in the regression. These observations prove volatility forecasts using WEL estimator are superior to the ones using ML and GMM.

To conclude, through the S&P 500 example we have shown that WEL estimator can improve the parameter estimations and forecasts when outliers exist in data and is more robust than ML and GMM estimators.

4.8 Conclusions

In this chapter our contributions are as follows: firstly the moment conditions are incorporated in empirical likelihood method to estimate the parameters of vector MEM. Thus specifying a particular multivariate distribution is avoided. Secondly a novel automatic routine to detect outliers using Oja depth and k-means clustering is introduced. Finally the robustness is increased by adjusting the weights of empirical likelihood functions corresponding to outliers. As a result the proposed estimator is semi-parametric and can accommodate data containing additive outliers well.

Two estimators that have been applied on vector MEM, ML

and GMM, are compared with WEL through a simulation study containing three different outlier scenarios. The first scenario is outliers simultaneously appear in each margin and the time intervals between them are equal. The second scenario is outliers simultaneously appear in each margin but the time intervals between them are random. And the last scenario is outliers appear in a completely random pattern, which means that outliers in $x_{1,t}$ and $x_{2,t}$ are independent to each other and the intervals between outliers are random. Overall the results show that WEL estimator can largely reduce the MSE of MEM parameters in these three scenarios and therefore is able to estimate MEM parameters more accurately and stably than ML and GMM estimators.

The results of the three outlier scenarios show that the performance of WEL estimator is best when outliers appear simultaneously. This observation suggests that it is more suitable to use WEL when input data are closely related. For example it can be used to analyze different measures of volatility. Therefore we apply the three estimators on RV and BV calculated from 10-minute returns of S&P 500 composite index. The in-sample period covers the current financial crisis, during which a lot of unusual

fluctuations of RV and BV are observed. The parameter estimations show that comparing to WEL, ML and GMM estimators tend to underestimate MEM parameters B and overestimate A. Mincer-Zarnowitz regressions of both in-sample and out-of-sample forecasts show that forecasts using WEL estimator are superior to those using the other two estimators in the presence of outliers. This result confirms the conclusions from the simulation study.

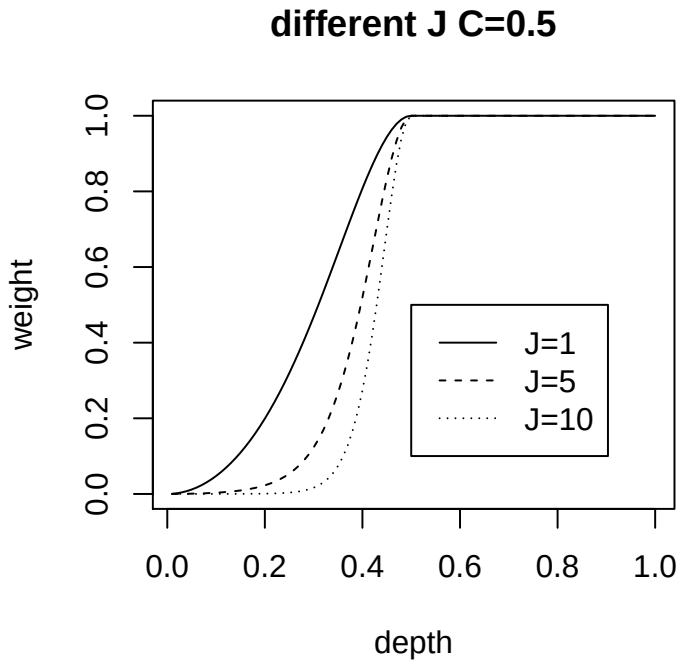

Figure 4.1: different C K=1 i=1

parameter		ω_1	α_1	β_1	ω_2	α_2	β_2
MSE	WEL	0.0005	0.0003	0.0011	0.0022	0.0003	0.0010
	ML	0.0007	0.0003	0.0014	0.0021	0.0003	0.0010
	GMM	0.0005	0.0003	0.0011	0.0022	0.0003	0.0010
	EL	0.0019	0.0003	0.0025	0.0115	0.0003	0.0008

Table 4.1: clean data tests

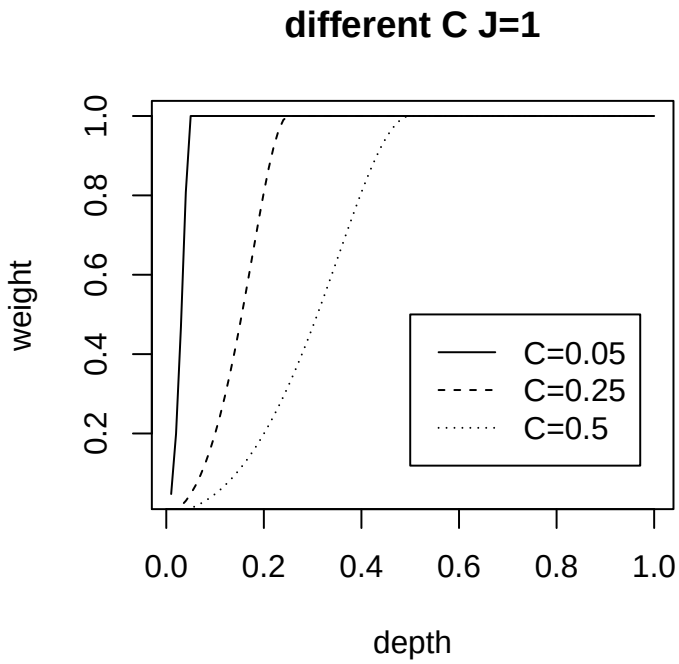

Figure 4.2: different K C=0.5 i=1

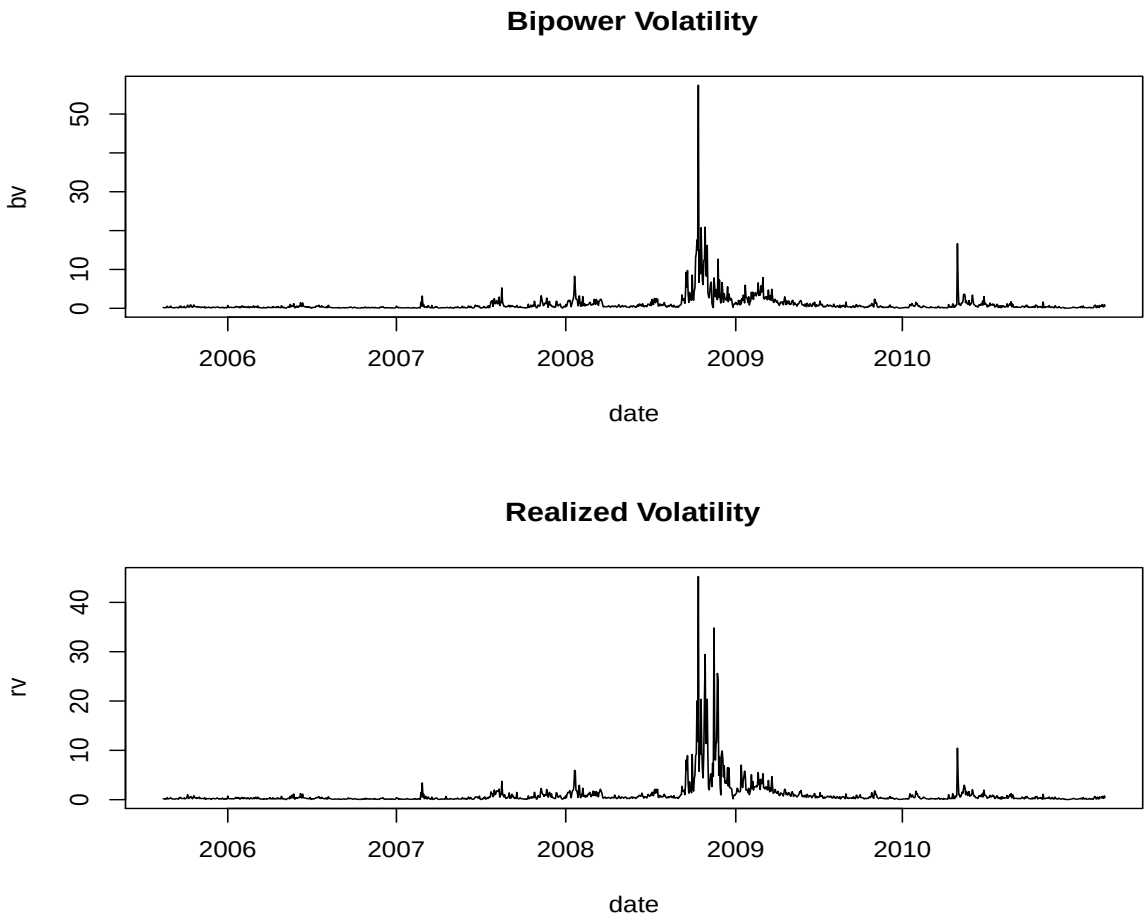

Figure 4.3: RV and BV of S&P 500, Aug/15/2005-Mar/16/2011

scenarios	simultaneously appear in each dimension or simultaneous not	time interval	number of outliers	size of outliers
1		between equidistant outliers	5%	$3, 6, 9$
			10%	$3, 6, 9$
2	simultaneous	random	5%	$3, 6, 9$
			10%	$3, 6, 9$
3	random	random	5%	$3, 6, 9$
			10%	$3, 6, 9$

Table 4.2: scenarios summary

Statistics	realized volatility	Bipower volatility
mean	1.000	1.000
median	0.3611	0.4132
min	0.0262	0.0260
max	45.24	57.43
sd	2.662	2.394
skewness	8.646	12.10
Kurtosis	100.6	235.9

Table 4.6: RV BV statistics(Aug/15/2005-Mar/16/2011)

percent of outliers		5			10		
size of outliers		$s=3$	$s=6$	$s=9$	$s=3$	$s=6$	$s=9$
ω_1	WEL	0.0109	0.0073	0.0056	0.1936	0.0591	0.0269
	ML	0.1057	0.0407	0.0194	0.1013	0.0497	0.0874
	GMM	0.0164	0.0092	0.0078	0.2297	0.1175	0.1489
	EL	0.0268	0.0079	0.0065	0.2408	0.0741	0.0692
α_1	WEL	0.0006	0.0006	0.0013	0.0013	0.0009	0.0022
	ML	0.0007	0.0014	0.0042	0.0008	0.0016	0.0069
	GMM	0.0006	0.0006	0.0013	0.0014	0.0015	0.0041
	EL	0.0006	0.0006	0.0013	0.0014	0.0008	0.0026
β_1	WEL	0.0093	0.0055	0.0050	0.1261	0.0289	0.0128
	ML	0.0782	0.0240	0.0129	0.0627	0.0211	0.0198
	GMM	0.0136	0.0069	0.0064	0.1469	0.0533	0.0548
	EL	0.0208	0.0060	0.0055	0.1517	0.0316	0.0256
ω_2	WEL	0.0132	0.0212	0.0132	0.2079	0.1446	0.0905
	ML	0.1901	0.1990	0.1538	0.4425	0.1301	1.625
	GMM	0.0240	0.0268	0.0221	0.2679	0.5589	0.7038
	EL	0.0265	0.0426	0.0381	0.3950	0.2773	0.3642
α_2	WEL	0.0018	0.0018	0.0011	0.0029	0.0017	0.0010
	ML	0.0015	0.0038	0.0099	0.0021	0.0042	0.0140
	GMM	0.0017	0.0016	0.0008	0.0028	0.0022	0.0025
	EL	0.0018	0.0016	0.0008	0.0030	0.0017	0.0008
β_2	WEL	0.0042	0.0048	0.0026	0.0388	0.0204	0.0091
	ML	0.0339	0.0301	0.0267	0.0669	0.0167	0.0832
	GMM	0.0063	0.0055	0.0037	0.0486	0.0604	0.0573
	EL	0.0041	0.0051	0.0027	0.0582	0.0270	0.0254

Table 4.3: MSE of parameters in scenario 1

percent of outliers		5			10		
size of outliers		$s=3$	$s=6$	$s=9$	$s=3$	$s=6$	$s=9$
ω_1	WEL	0.0018	0.0020	0.0027	0.0022	0.0026	0.0031
	ML	0.1503	0.1711	0.2410	0.2201	0.0827	0.3341
	GMM	0.0093	0.0027	0.0035	0.0034	0.0037	0.0031
	EL	0.0095	0.0030	0.0030	0.0023	0.0028	0.8495
α_1	WEL	0.0003	0.0005	0.0013	0.0003	0.0015	0.0035
	ML	0.0005	0.0007	0.0012	0.0006	0.0011	0.0015
	GMM	0.0003	0.0006	0.0013	0.0003	0.0015	0.0035
	EL	0.0003	0.0006	0.0013	0.0003	0.0015	0.0093
β_1	WEL	0.0021	0.0022	0.0032	0.0021	0.0030	0.0044
	ML	0.1129	0.0945	0.1053	0.1391	0.0330	0.0892
	GMM	0.0083	0.0028	0.0038	0.0030	0.0036	0.0044
	EL	0.0077	0.0029	0.0034	0.0021	0.0031	0.1851
ω_2	WEL	0.0031	0.0050	0.0075	0.0057	0.0074	0.0110
	ML	0.1367	0.2397	0.4747	0.0898	0.1528	0.1860
	GMM	0.0032	0.0089	0.0089	0.0073	0.0143	0.0109
	EL	0.0168	0.0271	0.0271	0.0209	0.0292	3.1648
α_2	WEL	0.0010	0.0010	0.0007	0.0009	0.0005	0.0005
	ML	0.0011	0.0020	0.0028	0.0011	0.0012	0.0015
	GMM	0.0010	0.0009	0.0005	0.0009	0.0005	0.0005
	EL	0.0010	0.0009	0.0005	0.0009	0.0005	0.0180
β_2	WEL	0.0014	0.0014	0.0013	0.0016	0.0009	0.0009
	ML	0.0257	0.0334	0.0500	0.0140	0.0144	0.0131
	GMM	0.0013	0.0020	0.0014	0.0019	0.0017	0.0009
	EL	0.0016	0.0021	0.0012	0.0016	0.0010	0.1395

Table 4.4: MSE of parameters in scenario 2

percent of outliers		5			10		
size of outliers		$s=3$	$s=6$	$s=9$	$s=3$	$s=6$	$s=9$
ω_1	WEL	0.2792	0.4071	0.4792	0.3098	0.4275	0.5415
	ML	0.3164	0.8934	1.547	0.6872	0.8581	2.522
	GMM	0.3199	0.3811	0.4598	0.4691	0.4994	0.7358
	EL	0.3174	0.4901	0.2881	0.4043	0.4395	0.8555
α_1	WEL	0.0024	0.0072	0.0090	0.0036	0.0083	0.0099
	ML	0.0038	0.0086	0.0094	0.0050	0.0083	0.0095
	GMM	0.0023	0.0071	0.0084	0.0037	0.0079	0.0091
	EL	0.0025	0.0074	0.0084	0.0038	0.0079	0.0094
β_1	WEL	0.2058	0.1986	0.1850	0.1783	0.1347	0.1159
	ML	0.2407	0.4904	0.6694	0.4238	0.3328	0.6364
	GMM	0.2369	0.1824	0.1644	0.2738	0.1594	0.1564
	EL	0.2336	0.2382	0.2137	0.2344	0.1384	0.1851
ω_2	WEL	0.4514	1.292	1.523	0.6149	1.839	2.080
	ML	0.5107	1.212	4.090	0.5450	1.785	7.697
	GMM	0.5406	1.224	1.639	0.9892	2.330	3.290
	EL	0.5034	1.688	2.229	1.043	2.473	3.168
α_2	WEL	0.0041	0.0107	0.0142	0.0064	0.0139	0.0159
	ML	0.0061	0.0157	0.0191	0.0082	0.0155	0.0185
	GMM	0.0039	0.0112	0.0162	0.0063	0.0148	0.0181
	EL	0.0040	0.0113	0.0164	0.0066	0.0147	0.0180
β_2	WEL	0.0836	0.1512	0.1348	0.0935	0.1378	0.0105
	ML	0.0932	0.1705	0.4268	0.0877	0.1607	0.4642
	GMM	0.1006	0.1417	0.1303	0.1485	0.1758	0.1611
	EL	0.0850	0.1808	0.1602	0.1454	0.1745	0.1395

Table 4.5: MSE of parameters in scenario 3

parameter		ω_{RV}	ω_{BV}	$a_{1,1}$	$a_{1,2}$	$a_{2,1}$	$a_{2,2}$	$b_{1,1}$	$b_{2,2}$	ϕ_1	ϕ_2	ρ	τ
filter one	WEL	0.0136	0.0166	0.3334	-	-	0.3318	0.6542	0.6529	-	-	0.8919	-
	ML	0.0286	0.0352	0.4762	-	-	0.4530	0.4978	0.5145	2.776	2.774	0.9266	4.666
	GMM	0.0185	0.0232	0.3589	-	-	0.3596	0.6243	0.6190	-	-	0.8894	-
filter two	WEL	0.0118	0.0169	0.2614	0.0780	−0.0280	0.3609	0.6509	0.6519	-	-	0.9011	-
	ML	0.0400	0.0545	0.7045	−0.1340	0.2754	0.3035	0.3913	0.3670	2.458	2.480	0.9325	12.45
	GMM	0.0203	0.0305	0.3740	0.0786	0.0401	0.4206	0.5300	0.5106	-	-	0.9320	-

Table 4.7: Full vector MEM: Estimated parameters of ML, GMM and WEL

parameter		ω_1	α_1	β_1	ω_2	α_2	β_2	ω_3	α_3	β_3
True value		0.6	0.2	0.6	0.3	0.25	0.55	0.4	0.3	0.5
Mean	WEL	0.5127	0.1672	0.6566	0.3008	0.2283	0.5632	0.3455	0.2795	0.5434
	ML	0.4825	0.1623	0.6663	0.2837	0.2106	0.5920	0.3707	0.2587	0.5522
	GMM	0.5351	0.1584	0.6551	0.3258	0.2132	0.5619	0.4097	0.2514	0.5403
MSE	WEL	0.1237	0.0087	0.0315	0.0475	0.0115	0.0425	0.0522	0.0154	0.0391
	ML	0.1251	0.0052	0.0287	0.0334	0.0085	0.0333	0.0420	0.0116	0.0348
	GMM	0.0901	0.0069	0.0277	0.0381	0.0097	0.0368	0.0486	0.0138	0.0375

Table 4.10: 3-dimension estimation results

parameter		ω_1	α_1	β_1	ω_2	α_2	β_2	ω_3	α_3	β_3	ω_4	α_4	β_4
True value		0.6	0.2	0.6	0.3	0.25	0.55	0.4	0.3	0.5	0.5	0.35	0.45
Mean	WEL	0.5109	0.1574	0.6733	0.2686	0.2138	0.6025	0.3568	0.2577	0.5637	0.4245	0.3265	0.4988
	ML	0.4333	0.1667	0.6888	0.2152	0.1970	0.6597	0.3199	0.2460	0.5969	0.3727	0.2953	0.5534
	GMM	0.5205	0.1565	0.6706	0.2868	0.2052	0.5990	0.3718	0.2526	0.5606	0.4685	0.3100	0.4984
MSE	WEL	0.1684	0.0078	0.0341	0.0370	0.0087	0.0313	0.0650	0.0134	0.0395	0.0665	0.0129	0.0329
	ML	0.1339	0.0076	0.0378	0.0272	0.0135	0.0416	0.0638	0.0142	0.0435	0.0580	0.0166	0.0384
	GMM	0.1470	0.0073	0.0323	0.0313	0.0090	0.0314	0.0459	0.0133	0.0361	0.0511	0.0124	0.0312

Table 4.11: 4-dimension estimations results

			filter one				filter two			
	d.v.	regressor	b_0	b_{EL}	b_{model}	Adj.R^2	b_0	b_{EL}	b_{model}	Adj.R^2
1-step	BV	WEL	0.1184	0.8915*	-	0.4672	0.1197	0.8886*	-	0.4642
		ML	0.1536*	-	0.8587*	0.4521	0.2297*	-	0.7876*	0.4236
		GMM	0.1197	-	0.8902*	0.4648	0.1605	-	0.8512*	0.4541
		WEL+ML	0.1128	1.309*	−0.4118*	0.4689	0.1211	0.7634*	0.1240	0.4655
		WEL+GMM	0.1210*	2.306*	−1.416	0.4677	0.1198	0.8863*	0.0022	0.4642
	RV	WEL	0.1069	0.9034*	-	0.5100	0.0705	0.9347*	-	0.5066
		ML	0.1449*	-	0.8686*	0.5029	0.2227*	-	0.7968*	0.4918
		GMM	0.1080	-	0.9024*	0.5096	0.1150	-	0.8945*	0.5013
		WEL+ML	0.1101	0.7672*	0.1333	0.5103	0.1023	0.6360*	0.2699*	0.5113
		WEL+GMM	0.1069	0.9331	−0.0296	0.5101	0.0734	0.8504*	0.0817	0.5067
5-step	BV	WEL	0.8129*	0.8481*	-	0.6062	0.8217*	0.8452*	-	0.6039
		ML	0.8007*	-	0.8503*	0.5935	1.598*	-	0.7033*	0.4858
		GMM	0.7582*	-	0.8581*	0.6045	0.9109*	-	0.8299*	0.5916
		WEL+ML	0.8255*	0.9536*	−0.1078	0.6064	0.7901*	0.7606*	0.0917*	0.6059
		WEL+GMM	0.8567*	1.4263*	−0.5863	0.6065	0.8161	0.6973*	0.1502	0.6045
	RV	WEL	0.8382*	0.8448*	-	0.6208	0.5267*	0.9015*	-	0.6309
		ML	0.9497*	-	0.8247*	0.5978	1.571*	-	0.7119*	0.5469
		GMM	0.8164*	-	0.8488*	0.6167	0.7284*	-	0.8649*	0.6109
		WEL+ML	0.8479*	1.202*	−0.3592*	0.6231	0.5305*	0.8937*	0.0072	0.6309
		WEL+GMM	0.9632*	3.518*	−2.696*	0.5101	0.4665*	1.499*	−0.5866*	0.6347

Note: The * symbol indicates the corresponding coefficient is statistically different from zero with 5% significance

Table 4.8: Mincer-Zarnowitz regression, in sample

			filter one				filter two			
	d.v.	regressor	b_0	b_{EL}	b_{model}	Adj.R^2	b_0	b_{EL}	b_{model}	Adj.R^2
1-step	BV	WEL	0.0805	0.8142*	-	0.3157	0.0943*	0.7985*	-	0.3336
		ML	0.1045*	-	0.7659*	0.3305	0.0890	-	0.8080*	0.2881
		GMM	0.0838	-	0.8087*	0.3492	0.1035*	-	0.7772*	0.3174
		WEL+ML	0.0726	1.716*	−0.8869	0.3630	0.1143	1.120*	−0.3644	0.3381
		WEL+GMM	0.0801	4.016	−3.203	0.3619	0.0966	1.254*	−0.4593	0.3362
	RV	WEL	0.0637	0.8106*	-	0.3790	0.0900*	0.7544*	-	0.3527
		ML	0.0836*	-	0.7600*	0.3560	0.0816*	-	0.7761*	0.3127
		GMM	0.0655*	-	0.8067*	0.3757	0.1009*	-	0.7232*	0.3426
		WEL+ML	0.0584	1.383*	−0.5612	0.3839	0.1003	0.9610*	−0.2345	0.3381
		WEL+GMM	0.1069	3.513	−2.703	0.5101	0.0885	1.003	−0.2440	0.3362
5-step	BV	WEL	0.3155*	0.7692*	-	0.5538	0.3254*	0.7462*	-	0.5503
		ML	0.2228*	-	0.7612*	0.5251	0.0650	-	0.9095*	0.4970
		GMM	0.2467*	-	0.7759*	0.5470	0.2828	-	0.7725*	0.5272
		WEL+ML	0.4635*	1.440*	−0.6882	0.5620	0.3780*	0.8283*	−0.1098	0.5508
		WEL+GMM	0.6540*	3.908*	−3.188*	0.5662	0.3906*	1.203*	−0.4877	0.5542
	RV	WEL	0.2776*	0.7277*	-	0.5980	0.3137*	0.6654*	-	0.5941
		ML	0.2047	-	0.7186*	0.5661	0.1814	-	0.7668*	0.5310
		GMM	0.2236*	-	0.7338*	0.5920	0.2977*	-	0.6635*	0.5714
		WEL+ML	0.3837*	1.296*	−0.5825	0.6058	0.3710*	0.7930*	−0.1618	0.5960
		WEL+GMM	0.5895*	4.336*	−3.658*	0.6116	0.3647	1.261*	−0.6100	0.6008

Note: The * symbol indicates the corresponding coefficient is statistically different from zero with 5% significance

Table 4.9: Mincer-Zarnowitz regression, out of sample

parameter		ω_1	α_1	β_1	ω_2	α_2	β_2	ω_3	α_3	β_3	ω_4	α_4	β_4	ω_4	α_4	β_4
True value		0.6	0.2	0.6	0.3	0.25	0.55	0.4	0.3	0.5	0.5	0.35	0.45	0.2	0.4	0.4
Mean	WEL	0.4006	0.3253	0.6076	0.2667	0.3588	0.5515	0.2880	0.3789	0.5485	0.3532	0.3849	0.5426	0.1413	0.4137	0.5118
	ML	0.4200	0.3332	0.5964	0.2167	0.3620	0.5642	0.2625	0.3685	0.5651	0.3434	0.3797	0.5501	0.1389	0.4128	0.5147
	GMM	0.4061	0.3237	0.6081	0.2698	0.3578	0.5518	0.3156	0.3731	0.5469	0.3811	0.3819	0.5406	0.1585	0.4093	0.5082
MSE	WEL	0.1285	0.0496	0.0238	0.0297	0.0363	0.0172	0.0454	0.0266	0.0176	0.0757	0.0219	0.0286	0.0117	0.0158	0.0307
	ML	0.1718	0.0503	0.0288	0.0300	0.0383	0.0238	0.0657	0.0287	0.0276	0.0956	0.0234	0.0330	0.0102	0.0174	0.0338
	GMM	0.1245	0.0497	0.0238	0.0278	0.0362	0.0168	0.0385	0.0263	0.0171	0.0685	0.0218	0.0283	0.0093	0.0162	0.0298

Table 4.12: 5-dimension estimations results

□ End of chapter.

Chapter 5

Forecast RV by Vector MEM

Forecasting ability on realized volatility (RV) can be increased by decomposing RV into continuous and jump components. Despite the jump component accounts for a considerable amount of total variation, it only plays a minor role in forecasting RV comparing to the continuous component. To better utilize jumps to predict RV, in this chapter the jump component is constructed as RV divided by its continuous component. Thus RV is decomposed into two components multiplicatively. Moreover both components are non-negative so they can be jointly modeled by vector multiplicative error model (vector MEM). A clearer picture of the interactions between them can be seen after filtering them through vector MEM as the impact of noises in each component is reduced. 10-

minute returns of S&P 500, Nasdaq and Dow Jones indices are used to investigate the out-of-sample forecast ability of the vector MEM. Moreover vector MEM is compared with three other models, all of which have been used to forecast RV before. Also vector MEM is compared with the model which utilize logarithm jump. The results show preferences for using vector MEM and multiplicative decompositions to forecast RV over the other models.

5.1 Introduction

Volatility plays an important role in empirical finance and time series analysis. One of the volatilities that has attracted much attention in the past decade is realized volatility (RV). RV integrates high frequency intra-day data into daily data and can be analyzed by discrete models. RV is first decomposed into continuous and jump components in Barndorff-Nielsen and Shephard [2004], where the decomposition is based on the differences between two measures of volatility, RV and bipower volatility (BV). The continuous component is found to be more persistent and predictable than the jump component in Andersen et al. [2007]. Therefore sep-

arating jump component from RV will result in significant gains in volatility forecasting. However, the jump component is suggested by Andersen et al. [2007] and Huang and Tauchen [2005] to account for a non-trivial part of total variation, which should not be overlooked. In addition, the jump component is found to be associated with central bank intervention and macroeconomic news [Beine et al., 2006, Barndorff-Nielsen and Shephard, 2006]. Thus the jump component is beginning to generate more and more research interest.

To this end, the jump component has been employed as a lag regressor in a univariate linear regression model of RV [Andersen et al., 2007] or a lag regressor in a multivariate model of squared return, RV and BV [Gallo and Velucchi, 2007]. The structure of the jump component is analyzed as well. For example, the jump component is modeled by autoregressive model in Bollerslev et al. [2009], two-regime Markov model in Lanne [2007] and a combination of autoregressive conditional hazard model and log-linear model in Andersen et al. [2011].

Although in theory the jump component can be calculated by subtracting bipower or multipower variation measure from RV,

for financial data with given frequency, this simple subtraction can include a lot of noises in the jump component. In some cases the noises will even lead to negative subtraction results. For this reason an artificial threshold value is used to eliminate the non-significant jumps and the negative valued jumps [Andersen et al., 2003, Lanne, 2007, Andersen et al., 2011, Gallo and Velucchi, 2007]. A side effect to use a threshold value is that it will produce a lot of zeros in the jump series. Therefore the difficulty of modeling jumps is increased. Further if the threshold value is too large, small valued jumps may be ignored, and the corresponding continuous components would be overestimated. On the other hand, if the threshold value is too small, some positive noises may be included in the jump series, and the corresponding continuous components would be underestimated.

This phenomenon motivates us to propose a novel approach to decompose RV: multiplicative decomposition, which is the first contribution of this chapter. In this study the jump component is calculated as RV divided by its continuous component, BV for instance. The division results are called multiplicative jump (M-jump) and are similar to the jumps used in Bollerslev et al. [2009],

except that logarithm function is not used on M-jumps. From the computation of M-jump it can be seen that M-jump is a non-negative process and both negative and positive valued noises are included in the series. Therefore multiplicative error model (MEM) is suitable to filter the M-jump series. Lam et al. [2010] first propose using MEM to filter the noises in the jump series which is obtained from additive decomposition of RV. Following Lam et al. [2010] and to further investigate the interactions between M-jump and the continuous component, vector MEM is used in this chapter to model and forecast RV. This is the second contribution of this chapter.

The approach used in this chapter is similar to the one used in Lanne [2007]. Both approaches forecast RV by decomposition. And both approaches use MEM to model the continuous component of RV.

However there are two differences between the model used in this chapter and the one used in Lanne [2007]. The first difference is that vector MEM is used in this chapter rather than univariate mixture MEM in Lanne's paper. The inputs for vector MEM in this chapter are the continuous components of RV and the cor-

responding M-jumps. The second difference is that instead of modeling jumps by two-regime Markov model as an independent process to the continuous component, the interactions between the continuous components and M-jumps are considered in this chapter. Since these interactions could be masked by the noises in the two components, in this chapter the relations between M-jumps and continuous components are examined based on their conditional expectations. In this way the impact of noises in both components is reduced, and the results are expected to be more reliable. In contrast in Lanne [2007] the continuous and jump components are assumed to be independent to each other.

Besides the model proposed in Lanne [2007], vector MEM is compared with two other models which have been used to forecast RV in Andersen et al. [2003]. Both of them belong to ARFIMA family and the forecasts are obtained directly based on past values of RV. These three models together with vector MEM are tested on three data sets. They are 10-minute returns of S&P 500 index, Nasdaq index and Dow Jones index. Vector MEM is expected to have superior performance over the other models.

The remainder of this chapter is organized as follows: Section

5.2 gives a short introduction to jumps and vector MEM. In Section 5.3 vector MEM and three other competing models are tested on three data sets. The last section is conclusion and future work.

5.2 Multiplicative jump and vector MEM

In this section two kinds of multiplicative decompositions of RV based on BV and quadpower volatility (QV) are introduced. After the introduction to multiplicative decomposition, vector MEM and its maximum likelihood estimator are presented.

5.2.1 Multiplicative jump

In theory, the jump series is proved to be the difference between RV and BV [Barndorff-Nielsen and Shephard, 2004]. However it is not always true for discrete financial data of given frequency. First of all, subtracting BV from RV can result in negative values, which contradicts the fact that jumps are non-negative in theory. Secondly, a positive result of the subtraction does not necessarily indicate the existence of jump. There are three methods in the literature to handle these problems.

The first method proposed in Bollerslev et al. [2009] calcu-

lates the jump series as the logarithm differences of RV and BV (which means RV is allowed to be smaller than BV), thus the jump component includes all the negative and positive valued noises produced by the subtraction. The second method adopted by Barndorff-Nielsen and Shephard [2004] and Gallo and Velucchi [2007] uses zero to truncate the jump series and ignores the negative values. Therefore the jump component is meaningful but the process will include small positive valued noises. In the third method [Andersen et al., 2011, Lanne, 2007], a positive threshold value is used to eliminate the negative values and small positive noises. And only the jumps which are statistically significant are considered. Jumps constructed following this procedure have the least noises among all the three methods, at the price of two drawbacks. Firstly this method relies on an artificial value to recognize significant jumps. And secondly by using a threshold, the possibility of mistakenly eliminating positive but relatively small valued jumps is increased.

A common characteristic of the second and the third methods is that a lot of zeros will be included in the jump series. In this case comparing to the models used on RV or its continuous component,

more complicated models are required for the jump component. For instances, in Lanne [2007] jumps are calculated using the third method and their distribution is assumed to switch between two normal distributions following a Markov process. This model, however, allows negative values where jumps can only be positive. In Andersen et al. [2011], after computing jumps using the third method, the series is represented by two new series. One contains the values of jumps and the other is the durations between significant jumps. In this way extra variables are introduced into the model.

To avoid mistakenly eliminating small valued jumps and also to avoid zeros in the jump series, in this chapter a similar approach to the first method [Bollerslev et al., 2009] is used. Both the negative and positive valued noises are included in the jump series. However the difference is that the jump series is calculated as RV divided by its continuous component, which does not involve logarithm function. And this jump component is denoted as multiplicative jump (M-jump). Comparing to Bollerslev et al. [2009], the logarithm is removed so that the interactions between M-jumps and continuous component have more straightforward

explanations. And more importantly, M-jumps constructed in this way are also positive like the continuous component. As a result it is possible to use vector MEM to model the two components jointly.

One important difference between additive jump and M-jump is that in additive jump series, a non-zero value indicates the existence of jumps. On the other hand M-jump is a series blended with jumps and noises. Therefore it is not an indicator of the presence of jumps.

Let $Y_{t,j}$ denote the jth log price of an asset at day t, where total numbers of j, M, depends on the intra-day time interval of log price series. And the jth return $y_{t,j}$ at day t is:

$$y_{t,j} = Y_{t,j} - Y_{t,j-1} \tag{5.1}$$

Then RV and BV can be written as:

$$RV_t = \sum_{j=1}^{M} y_{t,j}^2 \tag{5.2}$$

$$BV_t = u_1^{-2} \sum_{j=1}^{M-1} |y_{t,j+1}||y_{t,j}| \tag{5.3}$$

where $u_1 = 2^{1/2}\Gamma(1)/\Gamma(1/2)$. The M-jump corresponding to BV

is calculated as:

$$J_{BV,t} = \frac{RV_t}{BV_t} \tag{5.4}$$

Market microstructure noises can result in a bias for BV especially when the data frequency is high. One possible solution to overcome this problem is using QV instead of BV. And QV is defined as:

$$QV_t = u_{\frac{1}{2}}^{-4} \sum_{j=1}^{M-3} \sqrt{|y_{t,j} y_{t,j+1} y_{t,j+2} y_{t,j+3}|} \tag{5.5}$$

where $u_{1/2} = 2^{1/4}\Gamma(3/4)/\Gamma(1/2)$.

Thus the jump series based on QV is:

$$J_{QV,t} = \frac{RV_t}{QV_t} \tag{5.6}$$

Both $(BV^{1/2}, J_{BV}^{1/2})$ and $(QV^{1/2}, J_{QV}^{1/2})$ are used as inputs to vector MEM in the empirical analysis.

5.2.2 Vector MEM for jump and continuous components

Vector MEM is used to reduce the noises and possible serial correlations in the continuous component and the corresponding M-jumps, and further to investigate the relations between them based on their conditional expectations. MEM separates a non-negative process into a multiplication of two parts. The first part is the conditional expectation, which is assumed to follow a GARCH model. And the second part is a non-negative random noise vector, called innovations. Due to the multiplicative property, the expectations of innovations are restricted to one. Suppose $\vec{x}_t = (BV_t^{\frac{1}{2}}, J_{BV,t}^{\frac{1}{2}})^T$ or $\vec{x}_t = (QV_t^{\frac{1}{2}}, J_{QV,t}^{\frac{1}{2}})^T$, then vector MEM of order (p,q) is:

$$\vec{x}_t = \vec{\mu}_t \vec{\varepsilon}_t \tag{5.7}$$

$$\vec{\mu}_t = \vec{\omega} + \sum_{i=1}^{p} A_i \vec{x}_{t-i} + \sum_{j=1}^{q} B_j \vec{\mu}_{t-1} \tag{5.8}$$

where $\vec{\mu}_t$ is the conditional expectation vector of $\vec{x}_t$, $\vec{\varepsilon}_t$ is a non-negative random noise process and $\vec{\omega} \in \mathbf{R}^{2\times 1}, \{A_i \in \mathbf{R}^{2\times 2}\}_{i=1}^{p}, \{B_j \in \mathbf{R}^{2\times 2}\}_{j=1}^{q}$ are the MEM parameters.

The MEM parameters $\vec{\theta} = (\vec{\omega}, \{A_i\}, \{B_j\})$ in equation (5.8)

can be estimated by Maximum Likelihood method after assigning a parametric joint distribution to innovations. Following Cipollini et al. [2006] the multivariate non-negative distribution is structured as margin distributions linked up by a copula function. Gamma distribution is used as margin distribution in this chapter for its flexibility comparing to exponential distribution. Because the expectations of innovation are restricted to one, the scale parameter of Gamma distribution is assumed to be the inverse of its shape parameter. And Student's t copula is chosen because comparing to normal copula, Student's t copula considers the tail dependence between margins. The combination of Gamma distribution and Student's t copula is also used in Cipollini et al. [2009] and Cipollini and Gallo [2010]. The cdf F_ϕ and the pdf f_ϕ of Gamma distribution with a shape parameter ϕ and a scale parameter $\frac{1}{\phi}$ can be written as:

$$f_{\phi_k}(\varepsilon_{k,t}) = \frac{\phi_k^{\phi_k}}{\Gamma(\phi_k)} \varepsilon_{k,t}^{\phi_k - 1} \exp(-\phi_k \varepsilon_{k,t}) \tag{5.9}$$

$$F_{\phi_k}(\varepsilon_{k,t}) = \Gamma(\phi_k; \phi_k \varepsilon_{k,t}) \tag{5.10}$$

where $\Gamma(\varsigma, x)$ is the cdf of a random variable following $Gamma(\varsigma, 1)$ distribution computed at x. Denote $c_{\rho,\tau}$ as two-dimensional Student's t copula function with correlation parameter ρ and degree of freedom τ, then $c_{\rho,\tau}$ at (u_1, u_2) can be written as

$$c_{\rho,\tau}(u_1, u_2) \tag{5.11}$$
$$= |\rho|^{-\frac{1}{2}} \frac{\Gamma(\frac{\tau+T}{2})[\Gamma(\frac{\tau}{2})]^2(1+\frac{1}{\tau}(t_\tau^{-1}(u_1)^2+t_\tau^{-1}(u_2)^2)-2\rho t_\tau^{-1}(u_1)t_\tau^{-1}(u_2))^{-\frac{\tau+2}{2}}}{[\Gamma(\frac{\tau+1}{2})]^2\Gamma(\frac{\tau}{2})\prod_{k=1}^{2}(1+\frac{t_\tau^{-1}(u_k)^2}{\tau})^{-\frac{\tau+1}{2}}}$$

where t_τ denotes cdf of student's t distribution with degree of freedom τ.

The log-likelihood function for vector MEM with Gamma distribution and Student's copula assumptions is:

$$\begin{aligned} L(\vec{\theta}, \vec{\phi}, \rho, \tau) = \quad & \sum_{t=1}^{T} \log(f_{\phi_1}(\varepsilon_{1,t})) + \sum_{t=1}^{T} \log(f_{\phi_2}(\varepsilon_{2,t})) \\ & + \sum_{t=1}^{T} \log(c_{\rho,\tau}(F_{\phi_1}(\varepsilon_{1,t}), F_{\phi_2}(\varepsilon_{2,t}))) \end{aligned} \tag{5.12}$$

where T is the total number of observations, ϕ_k is the parameter for Gamma distribution of kth margin, and $\{\varepsilon_{1,t}\}_{t=1}^{T}$ and $\{\varepsilon_{2,t}\}_{t=1}^{T}$ are the innovations of the continuous and jump components re-

spectively.

The MEM parameters $\vec{\theta}$, Gamma distribution parameters $\vec{\phi} = (\phi_1, \phi_2)$, and Student's t copula parameters (ρ, τ) can be estimated by maximizing equation (5.12):

$$(\hat{\vec{\theta}}, \hat{\vec{\phi}}, \hat{\rho}, \hat{\tau}) = \arg \max_{\vec{\theta}\in\Theta, \vec{\phi}\in(0,\infty), \rho\in(-1,1), \tau\in(0,\infty)} L(\vec{\theta}, \vec{\phi}, \rho, \tau) \quad (5.13)$$

where Θ is the feasible region for MEM parameters discussed in Cipollini et al. [2006].

5.3 Empirical analysis

Data used in this empirical analysis are taken from three stock indices. Moreover three other models are chosen as competing models against vector MEM. And the forecasts are evaluated by two criteria, which help to analyze the relative performances of the models from different angles. This empirical analysis services two purposes: first, to investigate whether vector MEM and multiplicative decomposition could improve the forecasts of RV or not; and second, to study the interactions between the continuous component and M-jumps. In this empirical analysis, all the mod-

els are estimated by Maximum Likelihood method. The reason to choose ML method is that the model of Lanne [2006] used in comparison can only be estimated by ML method. To make the comparison fair, all the models are estimated by ML method

5.3.1 Data summary

Three data sets are used to test the forecast abilities of different models in this section. The first data set is 10-minute returns of Standard and Poor's composite stock index (S&P500), spanning from March 13th, 2006 to November 3rd 2009, which includes 858 days. The second data set is 10-minute returns of Nasdaq composite stock index (NASDAQ) from August 16th, 2005 to October 23rd 2009, which includes 1000 days. And the last data set is 10-minute returns of Dow Jones industry average index (DOW JONES). It contains 858 days' data from March 13th 2006 to November 4th 2009. Returns in all data sets are recorded from 9 : 30 to 16 : 00 thus there are 39 intra-day returns for each trading day. In this analysis two different decompositions of $RV^{1/2}$, based on BV and QV respectively, are investigated. Following Andersen et al. [2003], Gallo and Velucchi [2007] and Lanne [2007],

the square root of RV, BV and QV are used in this analysis. $RV^{1/2}$ and its decomposed components calculated from the three data sets are depicted in Figure 5.1, 5.2 and 5.3.

From the figures of the three indices, sudden increases in RV, BV and QV around September 2008 can be observed. September 2009 is the month in which Lehman Brothers Holding Inc. declared bankruptcy, and the current crisis began. Therefore the data of the indices can be naturally divided into two periods, which are corresponding to the before-crisis period and the crisis period. Out-of-sample forecast comparisons are conducted in both periods. The data from the before-crisis period serves as clean data and the data from the crisis period is the noisy data. The jump component is expected to contribute more to the forecasts of RV during the crisis period.

For S&P 500, the before-crisis period is from March 13th, 2006 to August 6th 2008, which includes 558 days. And the crisis period is from August 7th 2008 to November 3rd 2009, which includes 300 days. For NASDAQ, the before-crisis period is from August 16th, 2005 to August 4th 2008, which includes 700 days. And the crisis period is from August 5th 2008 to October 23rd 2009, which

includes 300 days. As for DOW JONES, the before-crisis period contains days from March 13th 2006 to August 14th 2008, which is 558 days in total. And the crisis period is from August 15th 2008 to November 4th 2009, which includes 300 days. The three data sets are separated in this way so that all of them have 300 days in the crisis period.

The statistics of three indices are tabulated in Table 5.1. From the statistics it can be seen that RV, BV and QV series possess strong autocorrelations, their Ljung-Box statistics are all very large comparing to those of the M-jump series. For example, during the before-crisis period the Ljung-Box statistic for $BV^{1/2}$ of S&P 500 is 3165.8, which is more than 10 times larger than the one for $J_{BV}^{1/2}$ (24.16). The Ljung-Box statistics are in line with the visual observations from Figure 5.1, 5.2 and 5.3. $BV^{1/2}$ and $QV^{1/2}$ in the three figures show obvious clustering effects. On the other hand the M-jumps show little clustering effects. This observation indicates that the M-jumps are less related to their past values than $BV^{1/2}$ and $QV^{1/2}$. Therefore the diagonal elements of parameter matrix A and B in vector MEM corresponding to the M-jumps are expected to be smaller than the ones corresponding

to $BV^{1/2}$ and $QV^{1/2}$.

However, by comparing the Ljung-box statistics of M-jumps during the two periods, it can be found that the statistics during the crisis period are significantly larger than those before the crisis. For example before the crisis, the Ljung-box statistics of $J_{BV}^{1/2}$ and $J_{QV}^{1/2}$ for S&P 500 are 24.16 and 16.66 respectively. During the crisis period, the corresponding statistics increase to 137.5 and 241.1. Therefore during the crisis period, the jump components are less random than before. And by applying suitable filters on jump series, they are expected to contribute more to RV forecasts.

The skewnesses of all the five series are similar. However, the kurtoses of two jump series are higher than three volatility measures, which indicates the jump series have heavier tails and their Gamma parameters ϕ are expected to be larger than those of BV and QV in vector MEM.

5.3.2 Models

Vector MEM is used to forecast RV based on two different multiplicative decompositions, $(BV^{1/2}, J_{BV}^{1/2})$ and $(QV^{1/2}, J_{QV}^{1/2})$. In our

experiments the specification of vector MEM is chosen as

$$
\begin{aligned}
x_{1,t} &= \mu_{1,t}\varepsilon_{1,t} && (5.14)\\
\mu_{1,t} &= \omega_1 + a_{1,1}x_{1,t-1} + a_{1,2}x_{2,t-1} + c_1 x_{1,t-2} \\
&\quad + b_{1,1}\mu_{1,t-1} + b_{1,2}\mu_{2,t-1} && (5.15)\\
x_{2,t} &= \mu_{2,t}\varepsilon_{2,t} && (5.16)\\
\mu_{2,t} &= \omega_2 + a_{2,1}x_{1,t-1} + a_{2,2}x_{2,t-1} + +c_2 x_{2,t-2} \\
&\quad + b_{2,1}\mu_{1,t-1} + b_{2,2}\mu_{2,t-1} && (5.17)
\end{aligned}
$$

This vector MEM set the order p and q in equation (5.8) as 2 and 1 respectively. And the first order parameter matrices A_1 and B_1 are assumed to be full matrices whose elements are $\{a_{i,j}\}$ and $\{b_{i,j}\}$ respectively. The second order matrix A_2 is a diagonal matrix whose diagonal element is c_1 and c_2.

The inputs $\{(x_{1,t}, x_{2,t})\}_{t=1}^{T}$ are $\{BV_t^{1/2}, J_{BV,t}^{1/2}\}_{t=1}^{T}$ or $\{QV_t^{1/2}, J_{QV,t}^{1/2}\}_{t=1}^{T}$, depending on which decompositions are used on RV. And $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are assumed to follow Gamma distribution linked by Student's t copula. The scale and shape parameters of the Gamma distribution are denoted as ϕ_k and $1/\phi_k$, $k = 1, 2$. And the correlation parameter and degree of freedom of Student's t copula are denoted

as ρ and τ respectively.

Besides vector MEM, three competing models are used to forecast RV. Two of them forecast RV without decomposition, they are ARFIMA(0,d,0) and ARFIMA(5,d,0). In Andersen et al. [2003] and Lanne [2006], these two models are proved to have a good performance and thus they are used as benchmarks in this analysis. The inputs for both ARFIMA models are logarithm RV. And the forecasts are transformed to $RV^{\frac{1}{2}}$ for comparison purpose.

The third model is the mixture MEM combined with Markov regime switching model proposed in Lanne [2007]. In Lanne's paper, the jump series is calculated as RV subtracted by BV and a threshold is imposed on the results. The final jump series only contains the significant jumps and is modeled by standard Markov regime switching model. The continuous component, on the other hand, is modeled by mixture MEM.

Among the models that forecast RV by additive decomposition, the third competing model is chosen for the following reasons. First of all MEM is used for the continuous component of RV in Lanne [2007], and in this chapter as well. Therefore the differences in forecast abilities are debited to the jump com-

ponent. Secondly, models involve variables other than RV, BV and jumps are not considered in this analysis. Because usually adding new variables to the model will improve the forecasts and thus the comparisons between additive jumps and multiplicative jumps are biased. These two reasons show that the comparisons between the two models can reflect the different characteristics of M-jumps and additive jumps meanwhile the influences from other factors are reduced to minimum.

For the above reasons the model used in Lanne [2007] is chosen as the third comparison model in this analysis. And the same assumptions and procedures described in Lanne [2007] are followed. First the significant jumps are selected at 0.95 confidence level. Then the continuous component is calculated as $RV^{1/2}$ subtracted by the square roots of significant jumps. Finally the continuous component is modeled by mixture MEM(2,1) and the jump series is modeled by standard two regime Markov switching model. The forecast of $RV^{\frac{1}{2}}$ is the summation of the forecasts of the continuous and jump components. This model is denoted as "mixture MEM" in the following discussions and "mix-MEM-d" in the tables for simplicity although it also includes a Markov

regime switching model.

Besides the three models mentioned above, the model used in Bollerslev et al. [2009] is compared with vector MEM. The model in Bollerslev et al. [2009] use a system of three equations to model the return, RV and jump series. The jump series is constructed as the logarithm of RV divided by BV. Because the structure of jump series is very similar to M-jumps, this model is included as a competing model to vector MEM. Two different model specifications are considered. The exact specification used in Bollerslev et al. [2009] and a simplified version are used in comparison. In the simplified model the asymmetric effect in the original model is ignored. The reason for this change is that the vector MEM framework does not involve any asymmetric effect. Therefore it is fair to compare the two models when neither of them include the asymmetric effect.

5.3.3 Forecast comparison criteria

In the empirical analysis section, two criteria are selected to compare the forecasts of different models: mean square error (MSE)

and Mincer-Zarnowitz regression. MSE is defined as:

$$MSE = \frac{1}{T}\sum_{i=1}^{T}((\sum_{j=0}^{k-1} rv(i+j))^{\frac{1}{2}} - (\sum_{j=0}^{k-1} \hat{rv}(i+j))^{\frac{1}{2}})^2 \quad (5.18)$$

where T is the total number of forecasts, $k = 1, 10$ is the forecast horizon and $\hat{rv}(i+j)$ is the j-step ahead out-of-sample forecast of RV at day i.

Following Andersen et al. [2003], the second criterion is Mincer-Zarnowitz regression. The summation of the first k-step ahead forecasts of vector MEM is compared with those of other models by the following linear regression:

$$\begin{aligned}(\sum_{j=0}^{k-1} rv(i+j))^{\frac{1}{2}} &= c + b_1 \cdot (\sum_{j=0}^{k-1} \hat{rv}_{MEM}(i+j))^{\frac{1}{2}} \\ &\quad + b_2 \cdot (\sum_{j=0}^{k-1} \hat{rv}_{model}(i+j))^{\frac{1}{2}} + e_i \quad (5.19)\end{aligned}$$

$$(5.20)$$

where $\hat{rv}_{MEM}(i+j)$ is the j-step ahead forecast at day i using vector MEM and $\hat{rv}_{model}(i+j)$ is the j-step ahead forecast at day i using one of the other competing models. e_i is an $i.i.d.$ series that follows normal distribution. The explanation power of a single regressor is represented by R square if one of b_1 and b_2

is restricted to 0. When there are two regressors, the forecasts of vector MEM are superior over those of the competing model if b_1 is significant and close to 1.

5.3.4 Before-crisis period

Following Andersen et al. [2003] and Lanne [2007], 1-step ahead and 10-step ahead out-of-sample forecasts of different models are compared in this analysis. The forecasts of these models are computed using rolling window method. And the estimation sample size in the before-crisis period is fixed to 400. Take k-step out-of-sample forecast for example, the first sample contains the data from day 1 to day 400 and is used to estimate the parameters. Then out-of-sample forecasts for the next k days (from day 401 to day $400+k$) are calculated. The second sample contains the data from day 2 to day 401, and the parameters are re-estimated for this sample, after which the forecasts for the next k days (from day 402 to day $401+k$) are made using the re-estimated parameters. This procedure goes on until the forecast period includes the 558th day for S&P and DOW JONES, and the 700th day for NASDAQ. Therefore for 1-step ahead, there are 158, 300 and 158

forecasts respectively when data used are S&P 500, NASDAQ and DOW JONES. And for 10-step ahead there are 149, 291 and 149 forecasts.

Table 5.2 displays the parameter estimation results of vector MEM specified in equation (5.15)-(5.17). The results, as expected, show that the Gamma parameters ϕ of M-jumps are higher than those of BV and QV. $a_{2,2}$ and $b_{2,2}$ are smaller than $a_{1,1}$ and $b_{1,1}$, which confirms the supposition that jump series are less related to their past values than BV and QV. From $b_{1,2}$ of the three indices it can be seen that the conditional expectations of jumps have significant negative influences on those of BV and QV, which is not found in the previous literatures using additive decomposition. For S&P 500 the value of $b_{1,2}$ is -1.541 and -5.577 respectively when the inputs are $(BV_t^{1/2}, J_{BV,t}^{1/2})$ and $(QV_t^{1/2}, J_{QV,t}^{1/2})$. The corresponding values for NASDAQ are -0.5547 and -0.4684. And for DOW JONES the values are -0.2358 and -0.1958. This negative impact of M-jumps is confirmed by the negative correlation parameter ρ of Student's t copula function. The values of ρ based on BV and QV are -0.2663 and -0.3800 respectively for S&P 500. And the corresponding values are -0.3083 and -0.4360 for

NASDAQ, and -0.2390 and -0.3328 for DOW JONES.

From the estimation results in Table 5.2 it can be found that there are some insignificant parameters included in the model. The values of $a_{2,1}$ and $b_{2,1}$ for the three indices are relatively small comparing to the other parameters. These two parameters represent the influences of the continuous components on the conditional expectations of M-jumps. The insignificance of these two parameters shows that the interactions between M-jumps and continuous components are one direction only, i.e., the M-jumps have influences on the continuous components but not the other way around.

Therefore the two parameters, $a_{2,1}$ and $b_{2,1}$, of the model presented in equation (5.15)-(5.17) are restricted to zero. For concision the restricted vector MEM on $\{BV^{1/2}, J_{BV}^{1/2}\}$ and $\{QV^{1/2}, J_{QV}^{1/2}\}$ are denoted as vMEM-BV-r and vMEM-QV-r respectively. And by letting $a_{2,1}$ and $b_{2,1}$ be zero, the rest parameters are re-estimated. The estimation results are tabulated in Table 5.3. Comparing to the unrestricted model, the remaining parameters are hardly affected by the restriction.

Using the estimated parameters of different models, the residu-

als of RV are calculated. The residual at day t is calculated as the true RV at day t minus the one step ahead prediction of different models. And the Ljung box statistics of residuals are tabulated in Table 5.8. The Ljung box statistics of RV for all three data sets are larger than 1000. And the Ljung box statistic of RV for NASDAQ before the crisis is as large as 3058.7. The Ljung box statistics of residuals, however, are generally very small. The Ljung box statistic of residuals for NASDAQ before the crisis, for example, is only 21.11 if the model is v-MEM-BV. These results show that the residuals are less self correlated and thus it is reasonable to apply these models on RV.

The MSE of 1-step ahead and 10-step ahead out-of-sample forecasts before the crisis are tabulated in Table 5.4. For both 1-step ahead and 10-step ahead forecasts, vMEM-BV-r and vMEM-QV-r achieve the lowest MSE values, with the exception of DOW JONES. However the differences between MSE values of mixture MEM and vector MEM on DOW JONES are very small. For 1-step ahead forecast the MSE values of models that utilize decomposition are slightly lower than the ones of ARFIMA models. And for 10-step ahead forecast the advantages of vector MEM and mix-

ture MEM over ARFIMA are more obvious. Consider the case of vMEM-QV-r, in 1-step ahead forecasts of S&P 500, vMEM-QV-r reduces the MSE by 6.3%, 4.5% and 0.14% respectively, comparing to ARFIMA(5,d,0), ARFIMA(0,d,0) and mixture MEM, and 11.2%, 10.6% and 1.8% in 10-step ahead forecasts. When the data used is NASDAQ, vMEM-QV-r reduces the MSE of the 1-step ahead forecasts by 8.6%, 6.7% and 1.1% comparing to ARFIMA(5,d,0), ARFIMA(0,d,0) and mixture MEM respectively, and 18.2% and 13.9% in the 10-step ahead forecasts comparing to ARFIMA(5,d,0) and ARFIMA(0,d,0). For DOW JONES data, vMEM-QV-r reduces the MSE by 5.0% and 5.4% comparing to ARFIMA(5,d,0) and ARFIMA(0,d,0) in 1-step ahead forecasts, and 21.7% and 24.8% in 10-step ahead forecasts.

The Mincer-Zarnowitz regression results in Table 5.5 confirm the conclusion obtained from MSE. It shows that vector MEM has the best performance for the following two reasons.

First when it comes to the values of R square in single variable regressions, vMEM-BV-r and vMEM-QV-r has achieved the highest values for S&P 500 and DOW JONES. Take vMEM-QV-r for example, in the 1-step ahead regressions of S&P 500, the R square

value of vMEM-QV-r is 0.4393, which is 13.1%, 7.3%, and 4.4% higher than mixture MEM, ARFIMA(0,d,0) and ARFIMA(5,d,0) respectively. And for 10-step ahead forecasts the improvements of vMEM-QV-r on R square are 24.4%, 23.8%, and 28.7%. The corresponding improvements of vMEM-QV-r on DOW JONES are 7.1%, 16.6% and 18.3% for 1-step ahead forecasts, and 11.3%, 73.4% and 63.2% for 10-step ahead forecasts.

Secondly besides R square, the coefficients of the regression also indicate vector MEM has superior forecast ability, especially for 10-step ahead forecasts. In most of the the two variable regressions, the coefficients of of vMEM-BV-r and vMEM-QV-r are significant and closer to 1.

The results from regressions also show that the differences between MSE values of vector MEM and mixture MEM for DOW JONES are insignificant. Consider the case of vMEM-QV-r, its R square values of 1-step ahead and 10-step ahead forecasts are 7.1% and 11.3% larger than the values of mixture MEM. Moreover in the two variable regressions, the coefficients of vMEM-QV-r are 0.4707 and 0.5023 for 1-step ahead and 10-step ahead forecasts, both of which are significant. On the contrary, the coefficients of

mixture MEM are 0.3196 and 0.0574, which are insignificant and much smaller than the ones of vMEM-QV-r .

The regression results confirm the conclusions from MSE, that vector MEM achieves the best performance among all the competing models, and also models that utilize decomposition of RV are better than two ARFIMA models for all the data sets.

5.3.5 Crisis period

Similar to the before-crisis period experiment, in this section the 1-step ahead and 10-step ahead out-of-sample forecasts of different models are compared using rolling window method. The in-sample period is set as all the available days in the before-crisis period. For S&P 500, the in-sample period is from March 13th 2006 to August 6th 2008, containing 558 days. For NASDAQ index the in-sample period is from August 16th 2005 to August 4th 2008, containing 700 days. And for DOW JONES index the in-sample period is from March 13th 2006 to August 6th 2008, and same as S&P index, it also containing 558 days. The out-of sample periods of all the indices contain 300 days, spanning from August 6th 2008 to November 3rd 2009 for S&P 500, from August 5th 2008

to October 23rd 2009 for NASDAQ and from August 15th 2008 to November 4th 2009 for DOW JONES respectively. The out-of-sample period is noisier than the before-crisis period, therefore it is expected that the differences in forecast ability will be larger between vector MEM and the three competing models.

Following the configuration of in-sample and out-of-sample periods, the estimation sample sizes of S&P 500 and DOW JONES are fixed to 558 and the one for NASDAQ is fixed to 700. Because the out-of-sample period contains 300 days, for 1-step ahead there are 300 forecasts and for 10-step ahead there are 291 forecasts.

The estimated parameters during the crisis period are provided in the lower part of Table 5.3. As expected, the autocorrelations of M-jumps during the crisis are higher than those before the crisis. Because $a_{2,2}$ are generally larger during the crisis period comparing to the ones before the crisis. For example, $a_{2,2}$ of vMEM-QV-r on S&P 500 during the crisis is 0.1911. And the corresponding value before the crisis is 0.0439. The negative impact of M-jumps on the continuous components still exists and can be observed from the negative values of $b_{1,2}$ and ρ. The rest parameters do not change much before and during the crisis.

And the Ljung box statistics of residuals during the crisis period are tabulated in Table 5.8. Similar to the results before the crisis, the Ljung box statistics of residuals are significantly smaller than those of RV for all three data sets. These results again show that the residuals are less self correlated and thus it is reasonable to apply these models on RV.

The MSE of forecasts are summarized in Table 5.4 under the crisis columns. Vector MEM has the lowest MSE value for all the data sets, with the exception of DOW JONES. In the 1-step ahead forecasts the MSE values of the competing models are only slightly larger than vector MEM. In 10-step ahead forecasts the differences in MSE are larger. Moreover comparing to the MSE of 10-step ahead forecasts before the crisis, the advantages of vector MEM over the other three models in 10-step ahead forecasts during the crisis period are even more obvious. Take vMEM-BV-r for example, for S&P 500 the MSE reductions of vMEM-BV-r are 31.4%, 28.5% and 14.9% comparing to ARFIMA(0,d,0), ARFIMA(5,d,0) and mix-MEM-d. While the largest reduction percentage of 10-step ahead forecasts before the crisis is only 11.2%. When the data is NASDAQ, vMEM-BV-r improves the MSE values of 10-

step ahead forecast by 32.3%, 27.9% and 10.0% comparing to ARFIMA(0,d,0), ARFIMA(5,d,0) and mix-MEM-d. As far as DOW JONES is concerned, the MSE value of vMEM-BV-r is almost the same as the one of mixture MEM in 10-step ahead forecasts. And the reductions comparing to ARFIMA(0,d,0) and ARFIMA(5,d,0) are 36.2% and 32.4% respectively. On the other hand the reductions before the crisis are only 24.8% and 21.7%.

The Mincer-Zarnowitz regression results are summarized in Table 5.6. And the results show that for 10-step ahead forecasts, vMEM-BV-r and vMEM-QV-r dominate the other three models. All the coefficients of vMEM-BV-r and vMEM-QV-r in the two variable regressions are not only significant but also closer to 1 comparing to the other models' coefficients. Further results from single variable regressions show that the R-square values of vMEM-BV-r and vMEM-QV-r are highest in the 10-step ahead forecast regressions. Again take vMEM-BV-r for example and consider 10-step ahead forecasts only, the R-square of single variable regression on S&P 500 is 0.6457, which is 13.9%, 13.7% and 11.8% larger than the ones of mixture MEM, ARFIMA(0,d,0) and ARFIMA(5,d,0) respectively. In the case of NASDAQ index,

the R square of vMEM-BV-r is larger by 6.8%, 15.3% and 10.0%. And the R-square improvements are 0.35%, 15.1% and 18.8% correspondingly for the DOW JONES case. The forecast ability of vector MEM is superior to that of mixture MEM for S&P 500 and NASDAQ indices. For DOW JONES index, the results show that vector MEM and mixture MEM are comparable, and the performance of vector MEM is slightly better than mixture MEM by having larger coefficients in regressions. However neither of the coefficients are significant.

5.3.6 Comparing M-jump and log M-jump

In this section the model used in Bollerslev et al. [2009] is compared with M-jump.

Model in Bollerslev et al. [2009]

The model used in Bollerslev et al. [2009] are related to the work in this chapter because both works use ratio as a proxy for jumps. M-jump defined in this work is RV/BV, while the jump defined in Bollerslev et al. [2009] is $\log(RV/BV)$. However, the models that analyze jumps are different. The model used in Bollerslev et al. [2009] is

$$
\begin{aligned}
r_t &= \gamma_0 + \sum_{j=1}^{d} \gamma_j r_{t-j} + \sqrt{RV_t}\epsilon_t && (5.21)\\
\log(BV_t) &= \alpha_0 + \alpha_d \log(BV_{t-1}) + \alpha_w (\log(BV))_{t-1:t-5} \\
&\quad + \alpha_m (\log(BV))_{t-22:t-1} + \theta_1 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} + \theta_2 I[r_{t-1} > 0] \\
&\quad + \theta_3 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} I[r_{t-1} < 0] + \sqrt{h_t}(u_t + g(\epsilon_t)) && (5.22)\\
h_t &= \omega + \sum_{j=1}^{q} \alpha_j (\sqrt{h_{t-j}}(u_{t-j} + g(\epsilon_{t-j})))^2 \\
&\quad + \sum_{j=1}^{p} \beta_j h_{t-j} && (5.23)\\
\log(\frac{RV_t}{BV_t}) &= \delta_0 + \sum_{j=1}^{n} \delta_j \log(\frac{RV_{t-j}}{BV_{t-j}}) + \phi_1 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} \\
&\quad + \phi_2 I[r_{t-1} > 0] + \phi_3 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} I[r_{t-1} > 0] \\
&\quad + (v_t + m(u_t) + k(\epsilon_t)) && (5.24)\\
& && (5.25)
\end{aligned}
$$

In equation (5.22), r_t, is the daily return, and it is modeled by an AR model. The error term ϵ_t is assumed to follow a standard Gaussian distribution, i.e. $\epsilon_t \sim N(0, 1)$.

Equation (5.23) and (5.24) are used to model the dynamics of $\log(BV_t)$. The mean of $log(BV_t)$ is assumed to follow HAR-RV

model. HAR-RV model utilize the past several days information as a whole. The symbol $(\log(BV))_{t+1-k:t}$ is the average of past k days values of $log(BV)$. The definition is:

$$(\log(BV))_{t+1-k:t} = \frac{1}{k}\sum_{j=1}^{k}\log(BV)_{t-j} \tag{5.26}$$

The error term in equation (5.23), u_t, is assumed to follow a mixture distribution of two Gaussian distributions, in detail: $u_t \sim N(0,1)$ with probability $(1-p_u)$ and $u_t \sim N(\mu_u, \sigma_u^2)$ with probability p_u. The function $g(\cdot)$ in equation (5.23) is assumed to be $g(x) = g_1x + g_2x^2$. This function is used to depict the nonlinear relationships between error terms in equation (5.22) and (5.23).

The asymmetric relation between BV and return is described by three terms: $\frac{|r_{t-1}|}{\sqrt{RV_{t-1}}}$, $I[r_{t-1} > 0]$ and $\frac{|r_{t-1}|}{\sqrt{RV_{t-1}}}I[r_{t-1} < 0]$.

The volatility of $log(BV_t)$, h_t, is assumed to follow a GARCH type model. Originally, the equation also contains the term $\lambda_j BV_{t-j}$ in Bollerslev et al. [2009]. However, this term does not appear in any tables of results in Bollerslev et al. [2009], therefore I assume the author did not include the terms in the model.

The last equation, equation (5.25) is used to capture the self correlations existed in jump series. The same terms used in equa-

tion (5.23) to depict asymmetric effect are used in equation (5.25). The functions $m(\cdot)$ and $k(\cdot)$ are used to describe the nonlinear relationships between the error term of equation (5.25) and those in equation (5.22) and equation (5.23). Same to $g(\cdot)$, they are defined as $m(x) = m_1 x + m_2 x^2$ and $k(x) = k_1 x + k_2 x^2$.

The error term in equation (5.25), v_t, is assumed to follow a mixture distribution of Normal Inverse Gaussian (NIG) distribution and Inverse Gaussian (IG) distribution, i.e. $v_t \sim NIG(\alpha_{NIG}, \beta_{NIG}, \delta_{NIG})$ with probability $(1 - p_v)$ and $v_t \sim IG(\lambda_{IG}, \mu_{IG})$ with probability $(1 - p_v)$.

After applying the model to S&P future data, in Bollerslev et al. [2009] the model is restricted to the following model:

$$
\begin{aligned}
r_t &= \gamma_0 + \gamma_2 r_{t-2} + \gamma_3 r_{t-3} + \sqrt{RV_t}\epsilon_t && (5.27)\\
\log(BV_t) &= \alpha_0 + \alpha_d \log(BV_{t-1}) + \alpha_w (\log(BV))_{t-1:t-5} \\
&\quad + \alpha_m (\log(BV))_{t-22:t-1} + \theta_1 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} \\
&\quad + \theta_3 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} I[r_{t-1} < 0] \\
&\quad + \sqrt{h_t}(u_t + g(\epsilon_t)) && (5.28)\\
h_t &= \omega + \alpha_1 (\sqrt{h_{t-1}}(u_{t-1} + g(\epsilon_{t-1})))^2 \\
&\quad + \beta_1 h_{t-1} && (5.29)\\
\log(\frac{RV_t}{BV_t}) &= \delta_0 + \delta_5 \log(\frac{RV_{t-5}}{BV_{t-5}}) + \phi_1 \frac{|r_{t-1}|}{\sqrt{RV_{t-1}}} \\
&\quad + (v_t + m(u_t) + k(\epsilon_t)) && (5.30)\\
g(\epsilon_t) &= g_1 \epsilon_t + g_2 \epsilon_t^2 && (5.31)\\
k(\epsilon_t) &= k_1 \epsilon_t && (5.32)\\
m(u_t) &= m_1 u_t + m_2 u_t^2 && (5.33)\\
& && (5.34)
\end{aligned}
$$

Model used in comparison with vMEM and M-jump

Because the framework of vector MEM in this chapter does not involve the asymmetric effects, and also does not include the

return variable, another version of the model is also used in this comparison. It is the modified version of the model used in Bollerslev et al. [2009]. As a result, two models are compared with vector MEM in this chapter. One is the model described by equation (5.28) - (5.34). And this model is denoted as "Bollerslev09" in the tables and the discussions below for concision. The other one is described by the following equations:

$$\begin{aligned}
\log(BV_t) &= \alpha_0 + \alpha_d \log(BV_{t-1}) + \alpha_w(\log(BV))_{t-1:t-5} \\
&\quad +\alpha_m(\log(BV))_{t-22:t-1} + \sqrt{h_t}(u_t) && (5.35)\\
h_t &= \omega + \alpha_1(\sqrt{h_{t-1}}u_{t-1})^2 + \beta_1 h_{t-1} && (5.36)\\
\log(\frac{RV_t}{BV_t}) &= \delta_0 + \delta_5 \log(\frac{RV_{t-5}}{BV_{t-5}}) + (v_t + m(u_t)) && (5.37)\\
m(u_t) &= m_1 u_t + m_2 u_t^2 && (5.38)\\
& && (5.39)
\end{aligned}$$

Because this model does not involve the return variable and also it does not consider asymmetric effect. It is denoted as simplified version of the original model in Bollerslev et al. [2009]. The notation "Bollerslev09s" is used to refer to this model.

Compare two models on S&P 500 data

10 minute returns of S&P 500 index data are used to compare the forecast abilities of vector MEM and Bollerslev09 models. The data range from March 13th, 2006 to August 6th 2008, which includes 558 days. And same to the experiments in the previous chapters, the in-sample period is the first 400 days and the out-of sample period is the rest 158 days. Rolling window method is used for both models and their forecast results are compared based on MSE and Mincer-Zarnowitz regression. The parameter estimation results are summarized in Table 5.7.

For 1-step ahead forecasts, the MSE of Bollerslev09s and Bollerslev09 are 1.171 and 1.486 respectively. And the MSE of 10-step ahead for the two models are 23.88 and 29.58. The 1-step ahead MSE of both Bollerslev models are slightly larger than those of vector MEM (0.8590 and 0.8673). And the 10-step ahead MSE of both models are very large comparing to the values of vector MEM (8.403 and 8.218).

The Mincer-Zarnowitz regression results are tabulated in Table 5.9. The regression results of both 1-step ahead forecast and 10-step ahead forecasts show that vector MEM outperforms Bollerslev09.

The performance of Bollerslev09 model is show to be worse than that of vector MEM. There are two possible reasons. The first reason is that the number of parameters of Bollerslev09 is too large. This increases the optimization difficulty. Although in the experiments all the optimizations converge, the accuracy is decreased when the number of parameters is too large. Bollerslev09s has 21 parameters and Bollerslev09 has 30 parameters. In comparison the vector MEM has only 12 parameters. The second reason is that the data used in this study and the data used in Bollerslev et al. [2009] are different. The data used in this thesis is S&P 500 stock index data while the data used in Bollerslev et al. [2009] is S&P future index. Although Bollerslev09 model can provide a good fit on S&P future index data, it is possible that the model does not accommodate S&P 500 stock index data well.

5.3.7 Conclusion on empirical analysis

In this empirical analysis three stock indices, S&P 500, NASDAQ and DOW JONES are used to test the performances of various models. Five observations from the results of parameter estimation and different criteria are in order.

First vector MEM and mixture MEM show their advantages of forecasting RV by decomposition, either additively or multiplicatively, over ARFIMA models. This result is in line with those in Lanne [2007], where mixture MEM is shown to outperform the models that do not decompose RV.

The second observation is that the differences in forecast power between ARFIMA models and models that rely on decomposition become larger as the forecast horizon becomes longer. This observation indicates that single jump, additive jump or M-jump, is hard to forecast and contains little information. Therefore the improvements of using jumps in 1-step ahead forecasts are not obvious. However when jumps of the past few days are considered together, the improvements on forecasts are significant.

The third observation is that the differences in forecast power between vector MEM and the competing models become larger as the data become noisier. The data collected during the crisis period are more volatile than the data in the before-crisis period. The experiments show that the advantages of vector MEM and multiplicative decomposition are expanded during the crisis period. Also the jump component exhibit higher series correlations

during the crisis period. This observation indicates that the jump component plays a more important role when realized volatility contains more noises.

From the parameter estimations on two data sets, it can be seen that M-jumps have significant negative impacts on BV and QV series. This phenomenon is the fourth observation. In previous researches where RV is decomposed into two additive components, the continuous component is assumed to be independent to jumps. But by decomposing RV into two components multiplicatively and filtering them through vector MEM to reduce the impact of noises, the conditional expectations of the continuous component are found to be highly related to those of M-jumps.

Last but not least, for long horizon forecasting, the predictive power of using the multiplicative decomposition and vector MEM is shown to be superior than modeling RV as a whole like ARFIMA models or using additive decompositions like mixture MEM.

5.4 Conclusion

In this chapter we propose decomposing RV multiplivatively and modeling the two components jointly by vector MEM. M-jump

series is constructed as RV divided by the continuous component, therefore it is a non-negative process. And M-jump series is able to be modeled together with the continuous component using vector MEM. Moreover in this way the problems of choosing a specific threshold to rule out the noises in jumps and producing zeros in jumps are avoided. In the empirical analysis section, vector MEM based on $(BV^{1/2}, J_{BV}^{1/2})$ and $(QV^{1/2}, J_{QV}^{1/2})$ are compared with ARFIMA(0,d,0), ARFIMA(5,d,0) and mixture MEM on three data sets: S&P 500, NASDAQ and DOW JONES. And vector MEM is found to have superior forecast ability, especially when the forecast horizon is long.

From the parameter estimation results of the three indices we can see that although M-jumps are less related to past values of themselves comparing to the continuous component, the impact of M-jumps on continuous component of RV is significant and negative, which is not found in previous research where RV is decomposed additionally.

Two evaluation criteria, MSE and Mincer-Zarnowitz regression, are used to compare the forecasts of different models. The results show that first of all the forecast ability is increased by

forecasting RV through its decompositions. Both models which utilize the decomposition, vector MEM and mixture MEM, have better performances than ARFIMA models. Secondly the improvements of decomposition in 10-step ahead forecasts are more obvious than those in 1-step ahead forecasts. This phenomenon indicates that to better utilize the jump component in forecasting RV, consecutive jumps rather than single jumps should be considered. Thirdly improvements of vector MEM over the other three models are larger during the crisis period, which indicates that M-jumps contribute more to RV forecasts when data are noisy. Lastly for long horizon forecasting, little evidence is found in the MSE and regression results against the conclusion that vector MEM on $(BV^{1/2}, J_{BV}^{1/2})$ and $(QV^{1/2}, J_{QV}^{1/2})$ are superior to the other models for all the data sets.

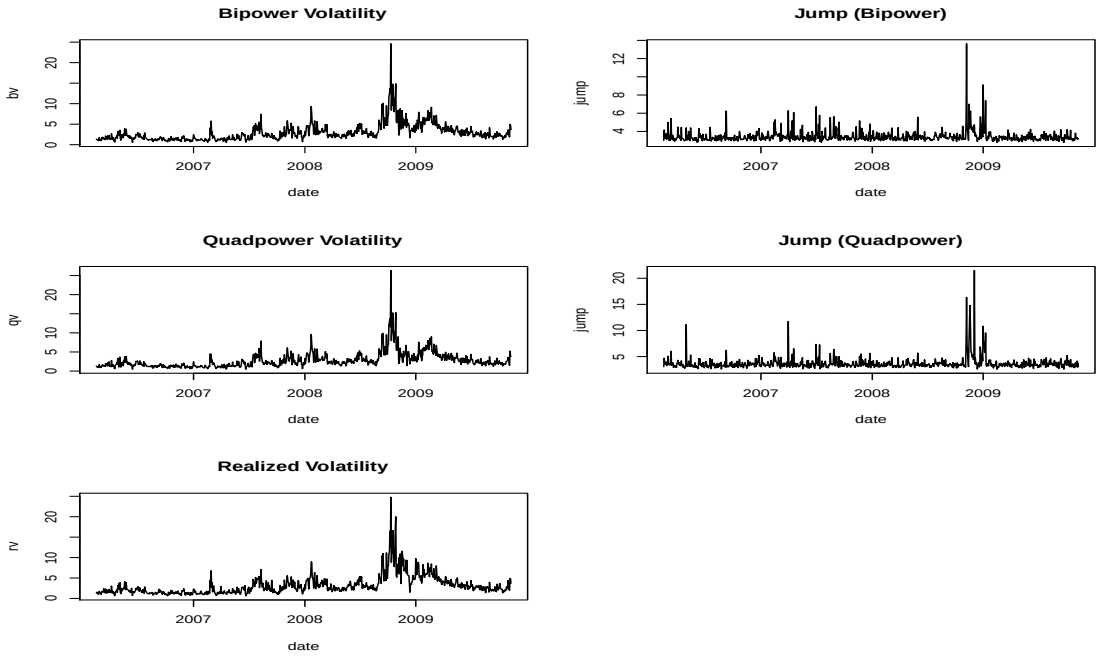

Figure 5.1: S&P 500 (Mar/13/2006-Nov/03/2009)

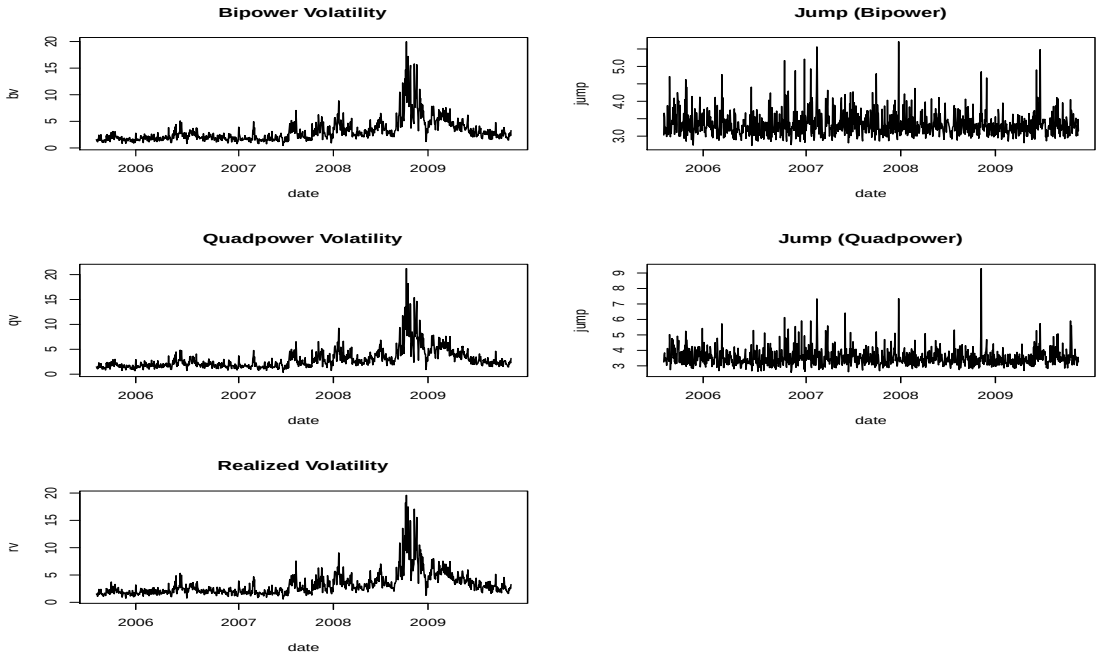

Figure 5.2: Nasdaq (Aug/16/2005-Oct/23/2009)

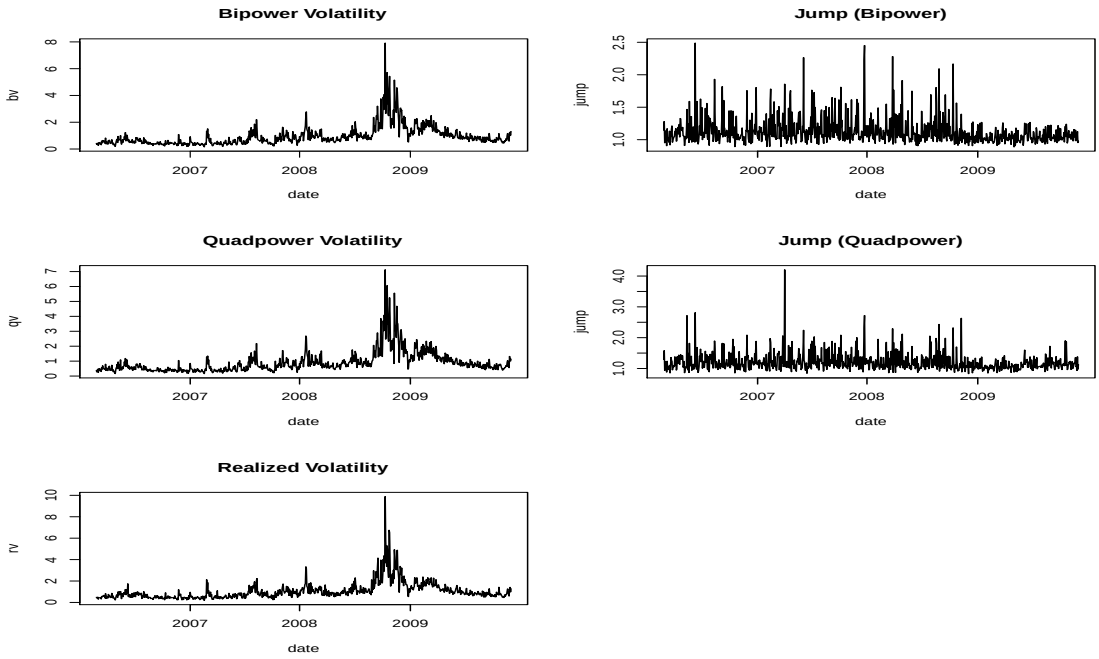

Figure 5.3: DOW JONES (Mar/13/2006-Nov/04/2009)

	Before-crisis period					Crisis period				
Statistics	$RV^{1/2}$	$BV^{1/2}$	$J_{BV}^{1/2}$	$QV^{1/2}$	$J_{QV}^{1/2}$	$RV^{1/2}$	$BV^{1/2}$	$J_{BV}^{1/2}$	$QV^{1/2}$	$J_{QV}^{1/2}$
S&P 500										
mean	2.201	2.069	1.076	2.000	1.131	4.769	4.316	1.108	4.046	1.235
median	1.867	1.775	1.037	1.689	1.079	3.734	3.543	1.043	3.303	1.099
min	0.5962	0.5235	0.8667	0.4699	0.8215	1.323	1.265	0.8790	1.226	0.8120
max	8.993	9.336	2.124	9.613	3.714	24.79	24.62	4.312	26.34	6.800
sd	1.127	1.083	0.1534	1.092	0.2344	3.109	2.708	0.2814	2.735	0.5765
skewness	1.717	1.826	2.755	1.944	4.435	2.340	2.660	6.694	3.125	5.748
Kurtosis	4.488	5.123	10.73	6.031	37.10	8.388	12.07	62.81	16.59	41.61
Ljung-box(20)	3264.8	3165.8	24.16	3022.8	16.66	1874.2	1695.1	137.5	1455.1	241.1
NASDAQ										
mean	2.478	2.361	1.062	2.272	1.120	4.857	4.669	1.049	4.486	1.113
median	2.191	2.054	1.035	1.972	1.082	4.108	3.868	1.033	3.636	1.077
min	0.5773	0.4292	0.8623	0.3534	0.8144	1.261	1.179	0.8876	0.9009	0.8507
max	9.059	8.863	1.806	9.253	2.324	19.61	19.97	1.735	21.24	2.940
sd	1.126	1.106	0.1207	1.112	0.1870	2.937	2.843	0.1031	2.847	0.1935
skewness	1.550	1.521	1.866	1.590	1.977	2.152	2.117	2.293	2.204	3.753
Kurtosis	3.197	3.056	5.836	3.629	6.871	5.894	5.892	9.470	6.856	26.92
Ljung-box(20)	3058.7	2898.3	15.80	2772.0	16.87	1765.4	1723.0	28.610	1627.8	23.33
DOW JONES										
mean	2.404	2.108	1.169	1.999	1.249	4.917	4.582	1.071	4.392	1.141
median	2.088	1.846	1.087	1.720	1.152	3.909	3.685	1.034	3.488	1.080
min	0.5699	0.4751	0.8910	0.4876	0.8431	1.400	1.391	0.8874	1.225	0.8290
max	10.53	8.783	2.489	8.500	4.204	31.27	25.04	2.167	22.52	2.632
sd	1.249	1.102	0.3948	1.070	0.4769	3.572	3.115	0.1494	3.067	0.2280
skewness	1.753	1.641	13.87	1.751	11.26	2.868	2.389	3.476	2.277	2.824
Kurtosis	5.853	4.492	262.6	5.123	189.6	12.54	8.342	18.43	7.142	11.98
Ljung-box(20)	2055.4	2319.9	8.793	2163.5	8.588	1896.7	2027.3	88.84	2043.6	46.68

Table 5.1: Data statistics

	S&P 500				Nasdaq				DOW JONES			
	BV		QV		BV		QV		BV		QV	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
ω_1	1.579	0.6520	2.928	1.253	0.6208	0.1923	0.5753	0.1719	0.5327	0.1723	0.4787	0.1593
ω_2	0.9958	0.0890	0.5098	0.0550	1.034	0.0731	1.113	0.0954	1.238	0.0307	1.344	0.0512
$a_{1,1}$	0.5304	0.0526	0.5337	0.0683	0.3815	0.0778	0.3807	0.0781	0.3223	0.2116	0.3320	0.2222
$a_{1,2}$	0.1991	0.0952	0.8223	0.1879	0.0472	0.0407	0.0240	0.0210	0.0150	0.0143	0.0116	0.0111
$a_{2,1}$	−0.0021	0.0122	−0.0029	0.0043	−0.0472	0.0193	−0.0725	0.0399	−0.0764	0.0733	−0.0705	0.0692
$a_{2,2}$	0.0102	0.0221	0.0101	0.0194	0.0149	0.0159	0.1072	0.0143	0.0475	0.0499	0.0484	0.0492
$b_{1,1}$	0.5437	0.2326	0.6165	0.2181	0.6326	0.2044	0.6714	0.1769	0.8137	0.0554	0.8077	0.0695
$b_{1,2}$	−1.541	0.6591	−5.577	2.262	−0.5547	0.2168	−0.4684	0.1722	−0.2358	0.2466	−0.1958	0.2063
$b_{2,1}$	0.0032	0.0212	0.0015	0.0067	−0.0107	0.0590	−0.0127	0.0817	0.0511	0.0560	0.0357	0.0539
$b_{2,2}$	0.0695	0.0934	0.0952	0.1108	0.0690	0.0657	0.0910	0.0837	0.4169	0.4131	0.4375	0.3970
ϕ_1	10.09	0.3687	9.209	0.3360	11.20	0.9555	10.02	0.7734	6.625	3.711	6.438	3.542
ϕ_2	50.59	2.073	28.92	2.551	79.05	3.543	37.13	1.198	11.41	8.259	8.666	5.665
ρ	−0.2663	0.0213	−0.3800	0.0196	−0.3083	0.0213	−0.4360	0.0174	−0.2390	0.0109	−0.3328	0.0182
τ	248.5	111.0	277.98	103.1	128.1	130.4	195.5	128.3	41.88	42.77	160.0	154.3
c_1	−0.1355	0.2072	−0.2085	0.1977	−0.1362	0.1737	−0.1708	0.1645	−0.1870	0.1833	−0.1919	0.1919
c_2	−0.0012	0.0344	0.0553	0.0419	−0.0210	0.0283	−0.0442	0.0179	−0.0007	0.0048	0.0037	0.0094

Table 5.2: Estimated parameters of vector MEM before the crisis

	S&P 500				NASDAQ				DOW JONES			
	BV		QV		BV		QV		BV		QV	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
Before-crisis period												
ω_1	1.702	0.5971	5.506	1.999	0.7843	0.1574	0.7183	0.1358	0.4054	0.1734	0.3775	0.1367
ω_2	0.9509	0.0757	0.9508	0.1221	1.015	0.0547	1.085	0.0776	0.7323	0.1331	0.7511	0.0945
$a_{1,1}$	0.5250	0.0357	0.5124	0.0638	0.3540	0.0592	0.3434	0.0558	0.4152	0.1037	0.4600	0.0769
$a_{1,2}$	0.2208	0.0347	0.7402	0.1974	0.0463	0.0315	0.0231	0.0157	0.0350	0.0073	0.0268	0.0056
$a_{2,1}$	–	–	–	–	–	–	–	–	–	–	–	–
$a_{2,2}$	0.0435	0.0087	0.0439	0.0206	0.0054	0.0077	0.0012	0.0045	0.0443	0.0168	0.0409	0.0108
$b_{1,1}$	0.7209	0.0955	0.6446	0.1557	0.6180	0.1831	0.6631	0.1533	0.6669	0.0495	0.6887	0.0318
$b_{1,2}$	−1.734	0.5372	−5.525	1.832	−0.5639	0.1981	−0.4848	0.1536	−0.3547	0.1451	−0.3070	0.1094
$b_{2,1}$	–	–	–	–	–	–	–	–	–	–	–	–
$b_{2,2}$	0.0807	0.0695	0.1010	0.1038	0.0479	0.0582	0.0641	0.0690	0.3747	0.1147	0.3878	0.0703
ϕ_1	10.29	0.1866	9.286	0.3554	11.24	0.9007	10.02	0.7456	9.548	0.6832	9.298	0.5882
ϕ_2	50.57	1.241	28.77	2.214	77.81	3.173	36.39	0.9897	18.04	1.308	13.22	0.7996
ρ	−0.2660	0.0097	−0.3810	0.0154	−0.3128	0.0156	−0.4408	0.0127	−0.2025	0.0682	−0.3338	0.0374
τ	116.5	82.01	260.3	111.7	37.68	18.92	96.20	81.62	13.53	5.087	14.61	4.632
c_1	−0.2783	0.1021	−0.2019	0.1520	−0.0852	0.1406	−0.1065	0.1158	−0.1251	0.0756	−0.1871	0.0678
c_2	−0.0027	0.0146	0.0155	0.0194	−0.0085	0.0304	−0.0316	0.0156	−0.0380	0.0215	−0.0223	0.0166
Crisis period												
ω_1	1.246	0.5422	0.9401	0.5212	1.905	0.4649	1.799	0.4000	0.2412	0.1529	0.2549	0.1546
ω_2	2.846	0.4037	2.573	0.5064	3.204	0.2401	3.312	0.1970	0.4882	0.1475	0.4970	0.1375
$a_{1,1}$	0.5465	0.0792	0.5668	0.0959	0.4256	0.0441	0.4160	0.0359	0.2225	0.0957	0.2335	0.1069
$a_{1,2}$	0.1306	0.0314	0.1086	0.0287	0.1202	0.0298	0.0901	0.0250	0.0275	0.0348	0.0232	0.0220
$a_{2,1}$	–	–	–	–	–	–	–	–	–	–	–	–
$a_{2,2}$	0.0107	0.0601	0.1911	0.1034	0.0195	0.0172	0.0400	0.0216	0.0797	0.0165	0.0766	0.0167
$b_{1,1}$	0.7530	0.0353	0.7694	0.0713	0.7774	0.0255	0.7872	0.0156	0.7278	0.0513	0.7417	0.0411
$b_{1,2}$	−0.4773	0.1591	−0.3525	0.1451	−0.6773	0.1556	−0.5896	0.1269	−0.2291	0.1238	−0.2048	0.1188
$b_{2,1}$	–	–	–	–	–	–	–	–	–	–	–	–
$b_{2,2}$	0.0413	0.1232	0.0676	0.1430	0.0292	0.0793	0.0446	0.0479	0.5428	0.1453	0.5719	0.1259
ϕ_1	10.91	0.9235	9.579	0.7644	11.16	0.6545	9.903	0.3919	9.196	1.333	8.741	1.091
ϕ_2	40.36	9.604	20.81	6.095	88.93	5.397	40.55	0.8159	24.62	7.711	16.40	3.795
ρ	−0.2552	0.0793	−0.3730	0.0922	−0.2756	0.0622	−0.4218	0.0083	−0.1035	0.1372	−0.2583	0.1201
τ	270.8	93.90	272.2	112.56	80.33	44.44	227.2	113.5	8.314	2.985	8.141	2.874
c_1	−0.3152	0.0803	−0.3523	0.1280	−0.2164	0.0454	−0.2167	0.0392	0.0307	0.0611	0.0061	0.0809
c_2	0.0302	0.0348	0.0443	0.0395	−0.0054	0.0278	−0.0240	0.0155	−0.0768	0.0462	−0.0610	0.0419

Table 5.3: Estimated parameters of restricted vector MEM

	S&P 500				NASDAQ				DOW JONES			
	Before-crisis		Crisis		Before-crisis		crisis		Before-crisis		crisis	
Forecast Horizon	1	10	1	10	1	10	1	10	1	10	1	10
vMEM-BV-r	0.8590	8.403	3.601	30.53	0.8881	7.556	3.523	27.77	1.197	8.638	5.180	43.48
vMEM-QV-r	0.8673	8.218	3.659	32.21	0.8972	7.838	3.588	28.17	1.219	8.272	5.322	44.19
mix-MEM-d	0.8685	8.366	4.336	35.87	0.9070	7.472	3.522	30.86	1.207	7.840	4.866	43.48
ARFIMA(0,d,0)	0.9080	9.194	3.790	44.51	0.9615	9.100	3.709	41.04	1.289	11.00	5.448	68.17
ARFIMA(5,d,0)	0.9255	9.732	3.580	42.69	0.9817	9.580	3.650	38.53	1.283	10.57	5.408	64.29

Table 5.4: MSE of forecasts

	1-step				10-step			
regressor	b_0	b_{vMEM}	b_{model}	R^2	b_0	b_{vMEM}	b_{model}	R^2
S&P 500								
vMEM-BV-r	0.4240	0.8781*	-	0.4344	4.5394*	0.6338*	-	0.2364
vMEM-QV-r	0.5217*	0.8290*	-	0.4393	4.766*	0.5966*	-	0.2436
mix-MEM-d	0.5124	-	0.8412*	0.3887	5.274*	-	0.5472*	0.1958
ARFIMA(0,d,0)	−0.0423	-	1.077*	0.4195	2.897*	-	0.8452*	0.1968
ARFIMA(5,d,0)	0.1775	-	1.007*	0.4068	3.317*	-	0.8110*	0.1877
vMEM-BV-r+mix-MEM-d	0.3257	0.6682*	0.2381	0.4372	4.441*	0.5213*	0.1241	0.2339
vMEM-BV-r+ARFIMA(0,d,0)	0.3733	0.8028*	0.0960	0.4308	7.927*	1.477*	−1.257*	0.2518
vMEM-BV-r+ARFIMA(5,d,0)	0.4601	0.9512*	−0.0890	0.4309	8.010*	1.595*	−1.411*	0.2615
vMEM-QV-r+mix-MEM-d	0.4001	0.6428*	0.2260	0.4357	4.601*	0.5074*	0.1069	0.2405
vMEM-QV+ARFIMA(0,d,0)	0.6344*	0.2705	0.4380	0.4273	6.832*	0.9988*	−0.6591	0.2485
vMEM-QV-r+ARFIMA(5,d,0)	0.4449	0.7275*	0.1348	0.4364	6.747*	1.013*	−0.6747	0.2513
NASDAQ								
vMEM-BV-r	0.1614	0.9876*	-	0.4948	1.887*	0.9423*	-	0.4524
vMEM-QV-r	0.1636	0.9715*	-	0.4948	1.769*	0.9449*	-	0.4535
mix-MEM-d	0.2562	-	0.9931*	0.4868	2.813*	-	0.8548*	0.4632
ARFIMA(0,d,0)	−0.00052	-	1.102*	0.4953	1.010	-	1.0924*	0.4537
ARFIMA(5,d,0)	0.0182	-	1.078*	0.4833	1.331*	-	1.063*	0.4311
vMEM-BV-r+mix-MEM-d	0.1660	0.8799*	0.1103	0.4932	2.509*	0.2430	0.6421*	0.4627
vMEM-BV-r+ARFIMA(0,d,0)	0.0466	0.4780	0.5770	0.4971	1.309*	0.4529	0.5821	0.4573
vMEM-BV-r+ARFIMA(5,d,0)	0.1380	0.8641*	0.1385	0.4933	1.896*	0.9517*	−0.1106	0.4505
vMEM-QV-r+mix-MEM-d	0.1680	0.8631*	0.1129	0.4933	2.447*	0.2680	0.6205	0.4630
vMEM-QV-r+ARFIMA(0,d,0)	0.0395	0.4794	0.5701	0.4987	1.260*	0.4761	0.5579	0.4584
vMEM-QV-r+ARFIMA(5,d,0)	0.1139	0.7429*	0.2624	0.4944	1.763*	0.9391*	0.0069	0.4516
DOW JONES								
vMEM-BV-r	0.6801*	0.7974*	-	0.2868	5.346*	0.5726*	-	0.2030
vMEM-QV-r	0.8876*	0.7191*	-	0.3049	5.411*	0.5535*	-	0.2290
mix-MEM-d	0.7250*	-	0.7809*	0.2848	5.229∗	-	0.5519*	0.2057
ARFIMA(0,d,0)	0.1485	-	1.045*	0.2615	3.488*	-	0.8336*	0.1321
ARFIMA(5,d,0)	0.2830	-	0.9938*	0.2578	3.838*	-	0.7856*	0.1403
vMEM-BV-r+mix-MEM-d	0.5391	0.4348*	0.4035	0.2994	5.099*	0.2677	0.3116	0.2058
vMEM-BV-r+ARFIMA(0,d,0)	0.5994	0.7235*	0.1071	0.2825	9.760*	1.180*	−1.155*	0.2281
vMEM-BV-r+ARFIMA(5,d,0)	0.6871	0.8051*	−0.0106	0.2822	8.521*	1.809*	0.9256*	0.2199
vMEM-QV-r+mix-MEM-d	0.6651*	0.4707*	0.3196	0.3121	5.328*	0.5023*	0.0574	0.2240
vMEM-QV-r+ARFIMA(0,d,0)	1.004*	0.7882*	−0.1151	0.3008	11.41*	1.232*	−1.426*	0.2757
vMEM-QV-r+ARFIMA(5,d,0)	1.083*	0.8506*	−0.2092	0.3017	10.03*	1.179*	−1.199*	0.2658

Note: The * symbol indicates the corresponding coefficient is statistically different from zero with 5% significance

Table 5.5: Mincer-Zarnowitz regression, out-of-sample, before the crisis

	1-step				10-step			
regressor	b_0	b_{vMEM}	b_{model}	R^2	b_0	b_{vMEM}	b_{model}	R^2
S&P 500								
vMEM-BV-r	−0.0391	1.0345*	-	0.6271	0.2756	1.073*	-	0.6457
vMEM-QV-r	0.0312	1.007*	-	0.5836	0.7502	1.0385*	-	0.6218
mix-MEM-d	0.0483	-	0.9883*	0.5484	−0.1112	-	1.068*	0.5668
ARFIMA(0,d,0)	−0.4123	-	1.173*	0.6319	−2.839*	-	1.397*	0.5680
ARFIMA(5,d,0)	−0.2435	-	1.127*	0.6462	−1.797*	-	1.318*	0.5777
vMEM-BV-r+mix-MEM-d	0.0048	1.120*	−0.0924	0.6263	0.7317	1.292*	−0.2429	0.6470
vMEM-BV-r+ARFIMA(0,d,0)	−0.3764	0.4960*	0.6430*	0.6459	−0.0969	1.005*	0.1018	0.6449
vMEM-BV-r+ARFIMA(5,d,0)	−0.2906	0.3854*	0.7350*	0.6538	−0.0251	0.9985*	0.1034	0.6449
vMEM-QV-r+mix-MEM-d	0.0179	0.9849*	0.0246	0.6181	0.5671	0.9637*	0.0851	0.6208
vMEM-QV-r+ARFIMA(0,d,0)	−0.3898	0.4474	0.6914*	0.6465	−1.349	0.7313*	0.4908*	0.6364
vMEM-QV-r+ARFIMA(5,d,0)	−0.3002	0.3651*	0.7538*	0.6555	−1.001	0.7074*	0.4904*	0.6375
NASDAQ								
vMEM-BV-r	0.2646	0.9505*	-	0.5901	1.427	0.9747*	-	0.6125
vMEM-QV-r	0.3464	0.9207*	-	0.5855	1.758*	0.9393*	-	0.6030
mix-MEM-d	0.1076	-	0.9998*	0.5899	1.265	-	0.9985*	0.5737
ARFIMA(0,d,0)	−0.4807	-	1.184*	0.5953	−3.176*	-	1.406*	0.5312
ARFIMA(5,d,0)	−0.4767	-	1.179*	0.6007	−2.866*	-	1.372*	0.5569
vMEM-BV-r+mix-MEM-d	0.1435	0.4848*	0.4991	0.5941	1.358	0.9197	0.0604	0.6113
vMEM-BV-r+ARFIMA(0,d,0)	−0.2604	0.3208	0.7911*	0.5957	5.838*	1.622*	−1.032*	0.6278
vMEM-BV-r+ARFIMA(5,d,0)	−0.3864	0.1364	1.014*	0.5996	4.683*	1.537*	−0.8475*	0.6202
vMEM-QV-r+mix-MEM-d	0.1556	0.3916	0.5858*	0.5933	1.476	0.7622*	0.2009	0.6034
vMEM-QV-r+ARFIMA(0,d,0)	−0.3121	0.2217	0.9055*	0.5950	5.832*	1.504*	−0.9261*	0.6142
vMEM-QV-r+ARFIMA(5,d,0)	−0.4236	0.0722	1.0894	0.5994	4.257*	1.344*	−0.6273*	0.6064
DOW JONES								
vMEM-BV-r	0.1506	0.9151*	-	0.6267	1.315	0.9275*	-	0.6007
vMEM-QV-r	0.3045	0.8668*	-	0.6184	1.759*	08788*	-	0.6021
mix-MEM-d	−0.4063	-	1.045*	0.6169	−0.2385	-	1.042*	0.5986
ARFIMA(0,d,0)	−0.6314*	-	1.247*	0.6115	−3.482*	-	1.496*	0.5216
ARFIMA(5,d,0)	−0.3260	-	1.167*	0.5997	−1.024*	-	1.292*	0.5075
vMEM-BV-r+mix-MEM-d	0.1557	0.9227*	−0.0087	0.6255	0.7394	0.6095	0.3599	0.6002
vMEM-BV-r+ARFIMA(0,d,0)	0.1433	0.9078*	0.0102	0.6255	6.566*	1.614*	−1.235*	0.6204
vMEM-BV-r+ARFIMA(5,d,0)	0.1229	0.8799*	0.0470	0.6255	5.329*	1.772*	−1.316*	0.6289
vMEM-QV-r+mix-MEM-d	−0.1103	0.3922	0.5771	0.6208	1.1276	0.6228	0.3066	0.6015
vMEM-QV-r+ARFIMA(0,d,0)	−0.0683	0.5657*	0.4439	0.6200	8.904*	1.719*	−1.591*	0.6310
vMEM-QV-r+ARFIMA(5,d,0)	0.0704	0.6390*	0.3220	0.6201	5.940*	1.627*	−1.235*	0.6283

Note: The * symbol indicates the corresponding coefficient is statistically different from zero with 5% significance

Table 5.6: Mincer-Zarnowitz regression, out-of-sample, the crisis period

parameter	mean	sd	parameter	mean	sd	parameter	mean	sd
Full model								
α_0	-0.2376	$1.5e-7$	δ_0	-0.1048	0.0002	γ_0	0.0013	$6.6e-8$
α_d	0.2898	$3.6e-5$	δ_1	0.02386	0.0001	γ_2	-0.0522	0.0001
α_w	0.4592	$7.6e-5$	δ_5	0.0161	$7.7e-5$	γ_3	-0.0211	$4.9e-5$
α_m	0.2456	$5.4e-5$	ϕ_1	-0.0107	0.0001			
θ_1	0.0416	$6.0e-6$	m_1	-0.0642	0.0002			
θ_3	0.1619	$3.8e-6$	m_2	0.0005	0.0004			
g_1	-0.2049	0.0003	k_1	0.0063	$1.3e-5$			
g_2	0.0090	0.0015	p_{v2}	0.0202	$2.2e-8$			
ω	0.9400	0.0010	α_{NIG}	48.38	$7.6e-7$			
α_1	0.0418	$1.4e-9$	β_{NIG}	25.61	$5.3e-7$			
β_1	0.7606	$2.5e-5$	δ_{NIG}	0.2615	$9.9e-5$			
p_{u2}	0.1609	$5.5e-7$	λ_{IG}	0.3009	$1.6e-8$			
μ_{u2}	0.6115	$1.4e-6$	μ_{IG}	0.3362	$1.4e-5$			
σ_{u2}	3.727	0.0001						
Simplified model								
α_0	-0.2138	0.1807	δ_0	-0.2175	0.0010	γ_0	–	–
α_d	0.7527	0.0050	δ_1	0.0245	0.0003	γ_2	–	–
α_w	0.1025	0.0350	δ_5	0.0154	0.0003	γ_3	–	–
α_m	0.0835	0.0620	ϕ_1	–	–			
θ_1	–	–	m_1	-0.6894	0.1954			
θ_3	–	–	m_2	2.402	1.125			
g_1	–	–	k_1	–	–			
g_2	–	–	p_{v2}	0.4149	0.0176			
ω	51.49	12555.7	α_{NIG}	46.95	1.1775			
α_1	0.0731	0.0080	β_{NIG}	23.62	4.692			
β_1	0.8722	0.0162	δ_{NIG}	0.4159	0.0048			
p_{u2}	0.1291	0.1058	λ_{IG}	1.497	0.5256			
μ_{u2}	1.254	3.225	μ_{IG}	0.5644	0.1651			
σ_{u2}	18179.5	$7.8e+9$						

Table 5.7: Estimated parameters of Bollerslev09 model

	S&P 500		NASDAQ		DOW JONES	
	Before-crisis	Crisis	Before-crisis	crisis	Before-crisis	crisis
vMEM-BV-r	20.38	66.64	21.11	39.33	23.47	82.49
vMEM-QV-r	25.24	86.67	19.97	39.80	25.50	84.41
mix-MEM-d	19.07	128.2	20.38	33.21	25.34	86.90
ARFIMA(0,d,0)	21.21	113.0	37.85	93.54	22.64	178.2
ARFIMA(5,d,0)	21.62	76.10	35.58	77.17	21.92	131.8

Table 5.8: Ljung box statistics of residues of realized volatility by different models

	1-step				10-step			
regressor	b_0	b_{vMEM}	b_{model}	R^2	b_0	b_{vMEM}	b_{model}	R^2
S&P 500								
Bollerslev09s	2.249*	0.9985*	-	0.3440	8.950*	0.1859*	-	0.0650
vMEM-BV-r+Bollerslev09s	0.4363	0.7766*	0.1077	0.4300	4.225*	0.6145*	0.0640	0.2566
vMEM-QV-r+Bollerslev09s	0.5205	0.7418*	0.1013	0.4405	4.446*	0.5759*	0.0663	0.2591
Bollerslev09	2.230*	1.145*	-	0.1282	12.08*	−0.1639*	-	0.0395
vMEM-BV-r+Bollerslev09	0.2643	0.8509*	0.0833	0.4277	5.538*	0.6073*	−0.0790	0.2479
vMEM-QV-r+Bollerslev09	0.3180	0.7994*	0.1042	0.438	5.757*	0.5678*	−0.0758	0.2478

Note: The * symbol indicates the corresponding coefficient is statistically different from zero with 5% significance

Table 5.9: Mincer-Zarnowitz regression, out-of-sample, before the crisis period, Bollerslev09 vs vMEM

□ End of chapter.

Chapter 6

Conclusion and future Work

This thesis discusses two aspects of vector MEM, the estimation and forecast. And it can be divided into three parts.

In the first part (Chapter 3), different estimation methods used on vector MEM are discussed. As a start, two stage ML method is introduced. Although two stage ML method may not be as efficient as ML and GMM, it can provide reasonability good starting values for the other two methods. After introducing the ML estimator and GMM estimator, a comparison between them is conducted through three experiments. In the first experiment the input data are generated by vector MEM, and both ML and GMM estimator have good performances. In the second experiment the dependence structure of input data is a linear mixture of product

copula and Student's t copula, which is different from the one assumed in the first experiment. The results show that both ML and GMM estimators can handle this kind of model mismatch well. Finally in the third experiment, outliers are added to clean data used in the first experiment. From the results it is shown that the outliers have a severe impact on estimation results of both estimators. This result motivates us to propose an alternative method, WEL method, to robustly estimate the parameters of vector MEM in the presence of outliers. This WEL estimator is discussed in the second part of the thesis.

In the second part of the thesis (Chapter 4), WEL method for vector MEM is proposed. WEL method is a generalization of empirical likelihood method in which weights for empirical likelihood function are equal to each other. The moment conditions used in GMM estimator is utilized as constraints in WEL method. Moreover k-means clustering are used to automatically determine the amount of outliers based on data depth. This automatic routine avoids choosing a threshold value to differentiate outliers from ordinary observations, which might be appreciated by practitioners. Through simulations which includes three different outlier

scenarios and an empirical study on forecasting RV and BV of S&P 500 stock index, WEL method is shown to have a superior performance over ML and GMM.

The third part (Chapter 5), different from the previous two, focuses on the applications of vector MEM on forecasting. In this part of the thesis the issue of forecasting RV by decomposition is discussed. And a novel decomposition of RV, multiplicative decomposition, is proposed. To our knowledge no researches have been done on this kind of decomposition. The jump component is calculated as RV divided by the continuously component, BV or QV for example. Therefore the so-constructed jump component is nonnegative and can be modeled and forecasted together with continuous component by vector MEM. The empirical analysis includes two data sets, S&P 500 stock index and NASDAQ stock index. In the empirical analysis, it is found that the jump component has a significant and negative impact on the continuous components, which is not reported in previous literatures where RV is decomposed additively. Also, three competing models, two of which do not require decomposition of RV and the rest one forecasts RV by additive decomposition, are considered.

Forecasts of different models are evaluated by MSE and Mincer-Zarnowitz regressions. The results show that vector MEM based on QV multiplicative decomposition has a superior performance over the other models.

Finally we wish to point out that the WEL estimator proposed in part two can be modified to be used on other kinds of multivariate filters besides vector MEM, for example multivariate GARCH models. It is expected to be a competitive estimator if not better than the quasi maximum likelihood or other robust estimators. Moreover in part three only two kinds of multiplicative decompositions of RV are tested. Other decompositions which are robust to jumps and noises such as the one based on staggered BV may further increase the forecast ability. Also, the model proposed in part three is in its simplest form. More complicated generalization may further improve the forecast ability. For example the vector MEM can be extended to include the asymmetric effect. In the empirical analysis section of part three, three models that are most closely related to vector MEM are chosen to compete with vector MEM. The comparison between vector MEM and other models would be interesting. In addition based on the results of

part three, modeling consecutive jumps as a whole rather than modeling single jumps is expected to further improve the forecast ability. These problems mentioned above are beyond the scope of this thesis and are left for future work.

□ **End of chapter.**

Bibliography

Katja Ahoniemi and Markku Lanne. Joint modeling of call and put implied volatility. *International Journal of Forecasting*, 25: 239–258, 2009.

Torben G. Andersen, Tim Bollerslev, Francis X. Diebold, and Paul Labys. Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625, 2003.

Torben G. Andersen, Tim Bollerslev, and Francis X. Diebold. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *REV. ECON. STAT.*, 89(4):701–720, 2007.

Torben G. Andersen, Tim Bollerslev, and Xin Huang. A reduced form framwork for modeling volatility of speculative prices based on realized variation measures. *J. ECONOMETRICS*, 160:176–189, 2011.

Matteo Barigozzi, Christian T. Brownlees, Giampiero M. Gallo, and David Veredas. Disentangling systematic and idiosyncratic risk for large panels of assets. Working Paper, 2011.

Ole E. Barndorff-Nielsen and Neil Shephard. Power and bipower variation with stochastic volatility and jumps (with discussion). *Journal of Financial Econometrics*, 2(1):1–48, 2004.

Ole E. Barndorff-Nielsen and Neil Shephard. Econometrics of testing for jumps in financial econometrics using bipower variation. *J. FINANC. ECONOMET.*, 4(1):1–30, 2006.

Michel Beine, Jérôme Lahaye, Sébastien Laurent, Christopher J. Neely, and Franz C. Palm. Centra bank intervention and exchange rate volatility, its continuous and jump components. Research Division, Federal Research Bank of St. Louis, Working Paper 2006-031C, 2006.

Taras Bodnar and Nikolaus Hautsch. Copula-based dynamic conditional correlation multiplicative error process. Working Paper, 2012.

Tim Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

Tim Bollerslev, Uta Kretschmer, Pigorsch Christian, and George Tauchen. A discrete-time model for daily s & p500 returns and realized variations: Jumps and leverage effects. *J. ECONOMETRICS*, 150:151–166, 2009.

Christian T. Brownlees and Giampiero M. Gallo. On variable selection for volatility forecasting: The role of focused selection criteria. *Journal of Financial Econometrics*, 6(4):513–539, 2008.

Christian T. Brownlees and Giampiero M. Gallo. Comparison of volatility measures: A risk management perspective. *Journal of Financial Econometrics*, 8(1):29–56, 2010.

Christian T. Brownlees and Giampiero M. Gallo. Shrinkage estimation of semiparametric multiplicative error models. *International Journal of Forecasting*, 27:365–378, 2011.

Christian T. Brownlees, Fabrizio Cipollini, and Giampiero M. Gallo. Intra-daily volume modeling and prediction for algorithm trading. *Journal of Financial Econometrics*, 9(3):489–518, 2011a.

Christian T. Brownlees, Fabrizio Cipollini, and Giampiero M.

Gallo. Multiplicative error models. Prepared for Volatility Models and Their Applications, 2011b.

Yohan Chalabi and Diethelm Wuertz. Weighted trimmed likelihood estimator for garch models. MPRA Paper No. 26536, Working Paper, November 2010.

Fabrizio Cipollini and Giampiero M. Gallo. Automated variable selection in vector multiplicative error model. *Comutational Statistics and Data Analysis*, 54(11):2470–2486, June 2010.

Fabrizio Cipollini, Robert F. Engle, and Giampiero M. Gallo. Vector multiplicative error models: Representation and inference. NBER Working Paper No. 12690, Working Paper, November 2006.

Fabrizio Cipollini, Robert F. Engle, and Giampiero M. Gallo. A model for multivariate nonnegative valued processes in financial econometrics. Econometrics Working Papers Archive $wp2007_16$, 2007.

Fabrizio Cipollini, Robert F. Engle, and Giampiero M. Gallo. A model for multivariate nonnegative valued processes in financial

econometrics. Social Science Research Network, working paper, 2009.

Fabrizio Cipollini, Robert F. Engle, and Giampiero M. Gallo. Semiparametric vector mem. *Journal of Applied Econometrics*, 2012. Early View.

Hao Ding and KaiPui Lam. Weighted empirical likelihood estimator for vector multiplicative error model. *Journal of Forecasting*, 2012. Accepted.

Robert F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

Robert F. Engle. New frontiers for arch models. *Journal of Econometrics*, 17(5):425–446, September–October 2002.

Robert F. Engle and Giampiero M. Gallo. A multiple indicators model for volatility using intra-daily data. *Journal of Econometrics*, 131(1–2):3–27, March–April 2006.

Robert F. Engle and Jeffrey R. Russell. Autoregressive conditional duration: A new model for irregularly. *Econometrica*, 66(5): 1127–1162, September 1998.

Gabriele Fiorentini, Giorgio Calzolari, and Lorenzo Panattoni. Analytic derivatives and the computation of garch estimates. *Journal of Applied Econometrics*, 11(4):399–417, July 1996.

Giampiero M. Gallo and Edoardo Otranto. Realized volatility and changeof regimes. Dipartmento di Statistica, Universita degli Studi di Firenze Working Papers 2012/02, 2012.

Giampiero M. Gallo and Margherita Velucchi. On the interaction between ultra-high frequency measures of volatility. Dipartmento di Statistica, Universita degli Studi di Firenze Working Papers 2007/01, 2007.

Christian Genest and Jock MacKay. The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician*, 40(4):280–283, 1986.

Daniel Gervini. Outlier detection and trimmed estimation for general functional data. Department of Mathematical Science, University of Wisconsin-Milwaukee, Working Paper, July 2011.

N L. Glenn and Yichuan Zhao. Weighted empirical likelihood estimates and their robustness properties. *Computation Statistics and Data Analysis*, 51(10):5130–5141, June 2007.

Alastair R. Hall. *Generalized Method of Moments*. Oxford University Press, 2005.

Wolfgang Karl Hardle, Nikolaus Hautsch, and Andrija Mihoci. Local adaptive multiplicative error models for high-frequency forecasts. SFB 649 Discussion Paper 2012-031, 2012.

Nikolaus Hautsch. Capture common components in high-frequency financial time series: A multivariate stochastic multiplicative error model. *Journal of Economic Dynamics and Control*, 32:3978–4015, 2008.

Ling Hu. Dependence patterns across financial markets: A mixed copula approach. *Applied Financial Economics*, 16(10):717–729, 2006.

Xin Huang and George Tauchen. The relative contribution of jumps to total price variance. *Journal of Financial Econometrics.*, 3(4):456–499, 2005.

KaiPui Lam and H. S. NG. Intra-daily information of range-based volatility for mem-garch. *Mathematics and Computers in Simulation*, 79:2625–2632, 2009.

KaiPui Lam, Hao Ding, and Terrance Mak. Conditional jump patterns mining using bidirectional hebbian clustering and sorting. Computational and Financial Econometrics, Senate House, University of London, UK, 2010.

Markku Lanne. A mixture multiplicative error model for realized volatility. *J. FINANC. ECONOMET.*, 4(4):594–616, 2006.

Markku Lanne. Forecasting realized exchange rate volatility by decomposition. *INT. J. FORECASTING*, 23:307–320, 2007.

George F. Levy. Analytic derivatives of asymmetric garch models. *Journal of Computational Finance*, 6(3), Spring 2003.

Regina Y. Liu. On a notion of data depth based on random simplices. *The Annals of Statistics*, 18(1):405–414, 1990.

Giovanni De Luca and Giampiero M. Gallo. A time-varing mixing multiplicative error model for realized volatility. Dipartmento di Statistica, Universita degli Studi di Firenze Working Papers 2010/03, 2010.

Simone Manganelli. Duration, volume and volatility impacts of trades. *Journal of Financial Markets*, 8(4):377–399, November 2005.

Jean-Claude Massé. Multivariate trimmed means based on the tukey depth. *Journal of Statistical Planning and Inference*, 139 (2):366–384, February 2009.

Nora Mulera and Victor J. Yohai. Robust estimates for garch models. *Journal of Statistical Planning and Inference*, 138(10): 2918–2940, October 2008.

Whitney K. Newey and Daniel McFadden. Large sample estimation and hypothesis testing. In James Heckman and Edward Leamer, editors, *Handbook of Econometrics*, volume 4, pages 2111–2245. Elsevier, 1994.

Hannu Oja. Descriptive statistics for multivariate distributions. *Statistics and Probability Letters*, 1(6):327–332, October 1983.

Edoardo Otranto. Classification of volatility in presence of changes in model parameters. CRENOS working paper 2011/13, 2011.

Edoardo Otranto. Spillerover effects in the volatility of financial markets. CRENOS working paper 2012/17, 2012.

Art B. Owen. Empirical likelihood ratio confidence intervals for a single function. *Biometrika*, 75(2):237–249, 1988.

Art B. Owen. *Empirical Likelihood*. Chapman and Hall, 2001.

Andrew J. Patton. Estimation of multivariate models for financial times series of possible different lengths. *Journal of Applied Econometrics*, 21(2):147–173, March 2006.

Robert Serfling and Yijun Zuo. General notions of statistical depth funtion. *The Annals of Statistics*, 28(2):461–482, 2000.

Pravin K. Trivedi and David M. Zimmer. Copula modeling: An introduction for practitioners. *Foundations and Trends in Econometrics*, 1(1):1–111, 2007.

John W. Tukey. Matematics and picturing data. In *Proceedings of the 1975 International Congress of Mathematics*, volume 2, pages 523–531, 1975.

Yijun Zuo and Hengjian Cui. Depth weighted scatter estimators. *The Annals of Statistics*, 33(1):381–413, February 2005.

Yijun Zuo and Robert Serfling. General notions of statistical depth function. *The Annals of Statistics*, 28(2):461–482, 2000.

Yijun Zuo, Hengjian Cui, , and Xuming He. On the stahel-donoho

estimator and depth-weighted mean of multivariate data. *The Annals of Statistics*, 32(1):167–188, 2004.