ESSAYS ON EXPONENTIAL SERIES ESTIMATION AND APPLICATIONS OF COPULAS IN FINANCIAL ECONOMETRICS

A Dissertation

by

CHIN MAN CHUI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2008

Major Subject: Agricultural Economics

ESSAYS ON EXPONENTIAL SERIES ESTIMATION AND

APPLICATIONS OF COPULAS IN FINANCIAL ECONOMETRICS

A Dissertation

by

CHIN MAN CHUI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Ximing Wu
Committee Members,	David A. Bessler
	James W. Richardson
	Suojin Wang
Head of Department,	John P. Nichols

August 2008

Major Subject: Agricultural Economics

ABSTRACT

Essays on Exponential Series Estimation and Applications of Copulas in Financial Econometrics. (August 2008)

Chin Man Chui, B.A., City University of Hong Kong;

M.S., University of Hong Kong;

M.S., Chinese University of Hong Kong

Chair of Advisory Committee: Dr. Ximing Wu

This dissertation contains three essays. They are related to the exponential series estimation of copulas and the application of parametric copulas in financial econometrics. Chapter II proposes a multivariate exponential series estimator (ESE) to estimate copula density nonparametrically. The ESE attains the optimal rate of convergence for nonparametric density. More importantly, it overcomes the boundary bias of copula estimation. Extensive Monte Carlo studies show the proposed estimator outperforms kernel and log-spline estimators in copula estimation. Discussion is provided regarding application of the ESE copula to Asian stock returns during the Asian financial crisis. The ESE copula complements the existing nonparametric copula studies by providing an alternative dedicated to the tail dependence measure.

Chapter III proposes a likelihood ratio statistic using a nonparametric exponential series approach. The order of the series is selected by Bayesian Information Criterion (BIC). I propose three further modifications on my test statistic: 1) instead of putting equal weight on the individual term of the exponential series, I consider geometric and

exponential BIC average weights; 2) rather than using a nested sequence, I consider all subsets to select the optimal terms in the exponential series; 3) I estimate the likelihood ratio statistic using the likelihood cross-validation. The extensive Monte Carlo simulations show that the proposed tests enjoy good finite sample performances compared to the traditional methods such as the Anderson-Darling test. In addition, this data-driven method improves upon Neyman's score test. I conclude that the exponential series likelihood ratio test can complement the Neyman's score test.

Chapter IV models and forecasts S&P500 index returns using the Copula-VAR approach. I compare the forecast performance of the Copula-VAR model with a classical VAR model and a univariate time series model. I use this approach to forecast S&P500 index returns. I apply a modified Diebold-Mariano test to test the equality of mean squared forecast errors and utilize a forecast encompassing test to evaluate forecasts. The findings suggest that allowing a more flexible specification in the error terms using copula tends improve the forecast accuracy. I also demonstrate combined forecasts improved forecasts accuracy over individual models.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my committee chair, Dr. Ximing Wu, for his support, patience, and encouragement throughout my graduate studies. His technical advice was essential to the completion of this dissertation and he has taught me innumerable lessons and insights on the workings of academic research in general. Without his guideline, this dissertation would not have been completed.

My thanks also go to Dr. David Bessler. He gave me the insight of how to generate the research idea and apply the academic research to the real world. I also would like to thank Dr. James Richardson and Dr. Suojin Wang for reading previous drafts of this dissertation and providing many valuable comments that improved the presentation and contents of this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

TABLE Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

It is difficult to find a market practitioner or empirically orientated academic who still believes asset returns are well described by a normal distribution. Nonetheless, it would not be hard to find someone who still uses linear correlation as their sole measure of dependence between risky assets or exposures, even though linear correlation is unable, for instance, to capture the fact that equities exhibit a greater tendency to crash together than to boom together. What's more, it is not a satisfactory measure of dependence between the payouts from different options, which are in general strongly non-normally distributed.

Copulas are a powerful framework for modeling dependence between risky assets, and are applicable in both of the above problems. Copula theory has its roots in the arcane field of probabilistic metric spaces, and there they remained until the early 1990s, at which time they gained increasing attention from statisticians. From there it was a small jump to finance, and the last decade or so has witnessed an explosion in the number of papers on the application of copulas to financial problems.

Loosely speaking, the copula generalizes linear correlation as a measure of dependence in the same way that the entire distribution of returns generalizes variance as a measure of risk. If returns are normally distributed, then variance completely describes

---

This dissertation follows the style and format of *Journal of Business and Economic Statistics.*

risk, and linear correlation completely describes dependence. However, if returns are not normally distributed, then variance is not a complete description of risk, and alternative summary measures such as value-at-risk or expected shortfall may be more appropriate. Similarly, if returns are not normally distributed, then correlation no longer completely describes the dependence between the returns.

Using copulas (otherwise known as dependence functions) to construct joint distributions allows us to specify the distributions of individual returns separate from each other and separate from the dependence structure. Clearly, this dramatically increases the flexibility in specifying multivariate distributions. More importantly, it provides a more flexible framework to consider dependence between risky assets.

The commonly used parametric copulas, typically with one or two parameters, have the risk of over-simplification or misspecification bias. A d-dimensional copula is bounded in $[0,1]^d$, hence, the popular nonparametric Gaussian kernel and log-spline estimators suffer from a boundary bias problem.

In this dissertation, I apply nonparametric exponential series methods to estimate the copula function and perform copula forecasting. In particular, I focus on the correction of boundary bias using exponential series methods, present a goodness-of-fit test using exponential series methods and measure the forecasting performance of the parametric copula in financial returns. Chapter II proposes a multivariate exponential series estimator (ESE) to estimate the copula density nonparametrically. The ESE has an appealing information theoretic interpretation and attains the optimal rate of convergence for nonparametric density in Stone (1982). More importantly, it overcomes the boundary bias of copula estimation. Extensive Monte Carlo studies show the proposed estimator outperforms kernel and log-spline estimators in copula estimation. It also demonstrates two-step density estimation through an ESE copula often outperforms direct estimation of the joint density. Discussion is provided regarding application of the ESE copula to Asian stock returns during the Asian financial crisis, as well as asymmetric tail dependence.

Based on the principle of maximum entropy, Chapter III proposes a likelihood ratio statistic using a nonparametric exponential series approach. The order of the series is selected by BIC. I propose three further modifications on my test statistics: 1) instead of putting equal weight on the individual term of the exponential series, I consider geometric and exponential BIC average weights; and 2) rather than using nested sequence, I consider all subsets to select the optimal terms in the exponential series; 3) use the likelihood cross-validation to estimate the likelihood ratio statistic. My extensive Monte Carlo simulations show that the proposed tests enjoy good finite sample performances compared to the traditional methods such as the Anderson-Darling tests. In addition, this data-driven method improves upon Neyman's score test.

Chapter IV models and forecasts S&P500 index returns using Copula-VAR approach. By allowing for a rich dependence structure and more flexible marginal distributions that can better fit the features of empirical data, I compare the forecast performance of Copula-VAR model and traditional VAR and a univariate time series model. I use this approach to forecast S&P500 index returns. A modified Diebold-Mariano test of equality of mean squared forecast errors and a forecast encompassing test are applied in forecast evaluation. The findings suggest that allowing a more flexible specification in the error terms using the copula tends improve the forecast accuracy. I also demonstrate combined forecasts improved forecasts accuracy over individual models.

CHAPTER II

EXPONENTIAL SERIES ESTIMATION OF EMPIRICAL COPULAS WITH APPLICATION TO FINANCIAL RETURNS

2.1 Introduction

The modeling of multivariate distributions from the multivariate outcomes is an essential task in economic model building. Assuming data comes from some well-known distribution, the parameters in the multivariate distribution can be estimated directly and statistical or economic hypothesis can be tested accordingly. One of the commonly used candidates is the elliptic distribution because of its nice statistical properties, but the elliptic distribution is often not adequate to capture the pattern of the empirical data. This is especially true when I estimate the multivariate returns distribution or try to account for the nonlinear dependence among several assets in financial economics, see, for example, Embrechts et al. (1999). One limitation of direct estimation of the multivariate distribution is the problem of the "curse of dimensionality". As the number of random variables increases, the amount of data that I need for the multivariate density estimation to a desirable accuracy will grow exponentially. For these reasons, copulas have become popular in modeling financial and economic variables as well as actuarial risks.

The concept of copulas is introduced by Sklar (1959) and has been recognized as an effective device for modeling dependence between random variables. It allows researchers to model each marginal distribution that best fits the sample, and to estimate a copula density with some desirable features separately. All the dependence structures are summarized in the copula. In practice, the joint distribution is estimated by imposing restrictions on the specific margins and copula respectively. For example, the *t*-distribution can capture the tail heaviness in the margin but the Gaussian copula only allows symmetric dependence. Extensive treatments and discussions in the properties of copulas can be found in Nelsen (1999) and Joe (1997).

There is a growing literature on the estimation of multivariate densities using copulas. Three commonly used estimation methods of copulas are: parametric, semiparametric and nonparametric. Each method may be further sub-divided into two-step and one-step approaches. In two-step approach, each margin is estimated first and the estimated values of the CDF are used to estimate copulas in the second step. The estimated parameters (in the parametric case) are typically inefficient. Theoretically, I can also estimate the copula in one-step. The margins and the copula are estimated jointly in this approach. Although the estimated parameters (in parametric case) are efficient in this case, the one-step approach is more computationally burdensome than the two-step approach. In empirical work, I may have prior knowledge on the margins but not on the structure of the joint dependence. Therefore, the two-step approach may have an advantage over the one-step approach in terms of the computational consideration, although the estimates may be inefficient.

It is not straightforward to choose the best combination of the margins and the copula in parametric estimation. Therefore, semiparametric and nonparametric estimations have become popular in the literature recently. The main advantage of these estimation methods is to let the data determines the copula without imposing any restriction on the copula.

In semiparametric estimation, a parametric copula is specified but not the margins. The parameters in the copula function are estimated by maximum likelihood estimation. See the earlier application in Oakes (1986), Genest and Rivest (1993), Genest, Ghoudi and Rivest (1995) and more recently in Liebscher (2005) and Chen et al. (2006).

Nonparametric estimation does not assume any parametric distribution in both the margins and the copula. In this way, nonparametric estimation provides a higher degree of flexibility, since the dependence structure of the copula is not directly observable. It also illustrates a rough picture helpful to researchers for subsequent parametric estimation of the copula. In addition, the problem of misspecification in the copula can be avoided in the context of nonparametric estimation.

The earliest nonparametric estimation in copulas is due to Deheuvels (1979) who estimate the copula density based on the empirical distribution. Further work using kernel methods have been proposed by Gijbels and Mielnicnuk (1990), Fermanian and Scaillet (2003) in a time series framework and Chen and Huang (2007) with boundary correction. Recently, other nonparametric methods to estimate the copula have appeared in the literature. Sancetta and Satchell (2004) use the Bernstein polynomials to approximate the Kimeldorf and Sampson copula. Hall and Neumeyer (2006) use wavelet estimators to approximate the copula density.

The kernel is one of the popular methods in nonparametric estimations. Li and Racine (2007) present a comprehensive review of this method. In spite of its popularity, there are several drawbacks in kernel estimation. If one uses a higher order kernel estimator in order to achieve a faster rate of convergence, it can result in a negative density estimate. In addition, the support of data is often bounded with high concentration at or close to the boundary in application. The Gaussian kernel will underestimate the underlying density at the boundary. This boundary bias problem is well known in the univariate case. The problem may be even more significant in the case of multivariate bounded support variables, see Muller (1991) and Jones (1993). Many methods have been proposed to resolve this boundary bias problem of the Gaussian kernel in the literature. These methods either adopt different functional forms of kernel beyond the Gaussian kernel (for example, see Lejeune and Sarda (1992), Jones (1993) and Jones and Foster (1996)) or transform data before applying the Gaussian kernel (Marron and Ruppert 1994). Recent studies included Chen (1999), Bouezmarni and Rombouts (2007). These studies propose to use the gamma kernel and local linear kernel estimators. They found that these estimators work well under the boundary data.

Log-spline estimators are also drawn considerable attention in the literature[1] and have been studied extensively by Stone (1990). But log-spline estimators suffer from the saturation properties. If I denote s the order of the spline and the logarithm of the density defined on a bounded support has r square integrable derivatives, the fastest convergence

[1] A closely related literature is the bivariate log-spline estimator studied by Stone (1994), Koo (1996) and Kooperberg (1998).

rate of can only be achieved with $s > r$. Moreover, there is a small probability that the MLE of the estimators does not exist for the log-spline estimator.

In this paper, I employ a nonparametric estimation method for copula density, namely the exponential series estimator (ESE). This estimator is based on the method of maximum entropy density subject to a given set of moment conditions. Compared with the kernel estimator, the number of estimated parameters is largely reduced in the context of ESE for a typical copula that is a smooth function. Further, ESE does not suffer from the boundary bias problem even if the domain is restricted in [0,1] and the speed of computation can also be significantly increased. The use of ESE on copulas may also lead to the test statistics designed to measure independence between margins. Preliminary findings in this paper are: ESE provides better performance than log-spline and kernel estimator in copula density estimation. The two-step density estimation through ESE copula often outperforms direct estimation of the joint density. Although my analysis is restricted to the bivariate case, it is straightforward to generalize the result to the multivariate case.

This paper is organized as follows. Section 2.2 presents the principle of maximum entropy and its statistical properties. Section 2.3 defines the copulas and the dependence measures to be used. Section 2.4 discusses the advantages of ESE for copula estimation. Monte Carlo results are presented in Section 2.5. Section 2.6 presents a financial application on the empirical copula estimation. Section 2.7 concludes.

2.2 Maximum Entropy Density

Shannon's information entropy is defined as

$$W(f) = -\int f(x,\boldsymbol{\theta}) \log f(x,\boldsymbol{\theta}) dx, \tag{2.1}$$

W measures the randomness or uncertainty of a distribution. The maximum entropy principle suggests choosing the density that maximizes Shannon's information entropy among all the distributions that satisfy certain moment conditions. The maximum entropy density then is obtained by maximizing (2.1) subject to moment conditions:

$$\int f(x,\boldsymbol{\theta}) dx = 1, \tag{2.2}$$

$$\int \phi_k(x) f(x,\boldsymbol{\theta})\ dx = \hat{\mu}_k(x), \quad k = 1, 2, ..., m. \tag{2.3}$$

where $\hat{\mu}_k(x) = 1/n \sum_{i=1}^{n} \phi_k(x_i)$ and $\phi_k(x_i)$ is a sequence of linear independent polynomials and at least twice differentiable. The first moment condition ensures $f(x)$ be a proper density function.

The estimated density takes the form

$$f(x, \hat{\boldsymbol{\theta}}) = \exp[-\hat{\theta}_0 - \sum_{k=1}^{K} \hat{\theta}_k \phi_k(x)]$$

To ensure $f(x, \hat{\boldsymbol{\theta}})$ is proper density function,

$$\hat{\theta}_0 = \log\left[\int \exp(-\sum_{k=1}^{m} \hat{\theta}_k \phi_k(x))\ dx\right]$$

Therefore,

$$f(x,\hat{\boldsymbol{\theta}}) = \frac{\exp[-\sum_{k=1}^{m}\hat{\theta}_k \phi_k(x)]}{\int \exp[-\sum_{k=1}^{m}\hat{\theta}_k \phi_k(x)]dx}$$

Where $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \ldots, \hat{\theta}_m]$.

In general, analytical solutions for $\boldsymbol{\theta}$ cannot be obtained and nonlinear optimization is employed (see, Zeller and Highfield (1988) and Wu (2003)). To solve $\boldsymbol{\theta}$, I use the Newton's method by iteratively updating $\boldsymbol{\theta}$ according to the following equation:

$$\hat{\boldsymbol{\theta}}_{(t+1)} = \hat{\boldsymbol{\theta}}_{(t)} - \mathrm{H}^{-1}\boldsymbol{b}$$

where for the $(t + 1)th$ stage of the updating, $b_k = \int g_k(x) f(x, \hat{\boldsymbol{\theta}}_{(t)})dx - \hat{\mu}_k$ and the Hessian matrix H takes the form

$$H_{ij} = \int \phi_i(x)\phi_j(x) f(x, \hat{\boldsymbol{\theta}}_{(t)})dx$$

In fact, the maximum entropy problem and maximum likelihood approach for exponential families can be considered as a duality problem (Golan et al. 1996). The maximum entropy W multiplied by the sample size n is equivalent the negative of the maximized log-likelihood function. This implies the estimated parameters $\boldsymbol{\theta}$ are asymptotically normal and efficient. Another concept related to the maximum entropy is relative entropy, or Kullback-Leibler distance. For two densities $f(x)$ and $g(x)$, the relative entropy is defined as:

$$D(f \,\|\, g) = \int f(x) \log \frac{f(x)}{g(x)} dx \tag{2.4}$$

The relative entropy measures the closeness, or the probability distance, between two densities. The minimum relative entropy density is obtained by minimizing (2.4) subject to empirical constraints (2.2) and (2.3). Previous research has emphasized the use of minimum relative entropy to obtain smooth estimate of the density function. Barron and Sheu (1991) shows that if the logarithm of the density has r square-integrable derivatives, $\int \left| D^r \log f(x) \right|^2 < \infty$, then the sequences of density estimators $\hat{f}(x)$ converges to $f(x)$ in the sense of Kullback-Leibler distance $\int f \log(f / \hat{f})dx$ at the rate $O_p(1/m^{2r} + m/n)$ if $m \to \infty$ and $m^3/n \to 0$ as $n \to \infty$ where m is the order of polynomial and n is the sample size. If there is no estimation error due to sampling variation of the empirical moments or inefficient estimators, the convergence rate will be $O_p(1/m^{2r})$. By setting $m = n^{1/(2r+1)}$, the optimal convergence rate becomes $O_p(n^{-2r/(2r+1)})$. Practically, the choice of m is equivalent to bandwidth selection in nonparametric density estimation. In a separate paper, under suitable conditions, Wu (2007) generalizes the results of Barron and Sheu to d dimensions and shows that the optimal convergence rate is $O_p(n^{-2r/(2r+d)})$ if I set $m = n^{1/(2r+d)}$.

2.3 Copulas and Dependence Measures

According to Sklar's theorem (Sklar 1959), a d dimensional multivariate density can always written as

$$f(\boldsymbol{x}) = f_1(x_1)f_2(x_2)\ldots f_d(x_d)\,c(F_1(x_1),F_2(x_2),\ldots,F_d(x_d)) \tag{2.5}$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)$, $\boldsymbol{x}$ is a random vector, $f(x_i)$ and $F(x_i)$ are the marginal density and the cumulative distribution of x_i respectively for $i = 1,\ldots,d$. $c(u_1,u_2,\ldots,u_d) = \partial^d C(u_1,u_2,\ldots,u_d)/\partial u_1 \partial u_2 \ldots \partial u_d$ is the copula density function uniformly distributed in $[0,1]^d$ and C is the cumulative copula density function. In (2.5) I note that the decomposition of the joint density in two parts. One describes the dependence structure among the random variables in the copula function, and other describes the marginal behavior of each component. One nice property of copula is that it is invariant under increasing transformation of margins. This property is very useful in finance research. For example, the dependence structure of two asset returns does not change whether returns or logarithm of returns is used. This is not true for linear correlation, which is only invariant under the linear transformation of margins.

There are various other measures of dependence, among which Kendall's τ and Spearman's ρ are two scale free measures of dependence in the context of copula. Starting with two independent realizations (X_1, Y_1) and (X_2, Y_2) of the same pair of random variables X and Y, Kendall's τ gives the difference between the probability of concordance and the probability of discordance:

$$\tau(X,Y) = P[(X_1 - X_2)(Y_1 - Y_2) > 0] - P[(X_1 - X_2)(Y_1 - Y_2) < 0] \quad \text{for } \tau \in [-1, 1].$$

The relation between Kendall's τ and the copula is as follows:

$$\boldsymbol{\tau}(C) = 4 \int_{[0,1]^2} C(u,v)dC(u,v) - 1 \tag{2.6}$$

As a consequence, Kendall's τ depends only on the copula of (X, Y). Comparisons between results using different copula functions should be based on a common Kendall's τ.

Another useful dependence measure defined by copulas is the tail dependence. In the bivariate case the concept of tail dependence measures the dependence existing in the upper quadrant tail, or in the lower quadrant tail. If there exists a positive association (Coles, Currie, and Tawn 1999; Tawn 1998) between extreme events of X and Y, then the conditional probability $\Pr[X > F_X^{-1}(u) \mid Y > F_Y^{-1}(u)]$ is greater than 0 and decreases as $u \uparrow 1$. By definition, the upper and lower tail dependence coefficients are

$$\lambda_U = \lim_{u \to 1} \Pr\left[X > F_X^{-1}(u) \mid Y > F_Y^{-1}(u)\right] = \lim_{u \to 1} \frac{1 - 2u + C(u,u)}{1 - u} \tag{2.7}$$

$$\lambda_L = \lim_{u \to 1} \Pr\left[X < F_X^{-1}(u) \mid Y < F_Y^{-1}(u)\right] = \lim_{u \to 0} \frac{C(u,u)}{u} \tag{2.8}$$

provided that these limits exist and λ_U and $\lambda_L \in [0,1]$. The upper (lower) tail dependence coefficient quantifies the probability to observe a large (small) X, given that Y is large (small). In other words, suppose, Y is very large (small) (at the upper quantile of the distribution), the probability that X is very large (small) at the same quantile defines the tail dependence coefficient $\lambda_U\,(\lambda_L)$. If $\lambda_U\,(\lambda_L)$ are positive, the two random variables exhibits upper (lower) tail dependent. Equation (2.7) and (2.8) suggest that the tail dependence coefficient can be derived from the copula density. Furthermore, the tail dependence between *X* and *Y* is also invariant under strictly increasing transformation of *X* and *Y*.

A better interpretation of this concept in finance may be obtained if I rewrite the definition of λ_U as

$$\lambda_U = \lim_{u \to 0^+} \Pr\left[X > VaR_u(X) \mid Y > VaR_u(Y)\right]$$

where $VaR_u(X) = F_X^{-1}(1-u)$ is the value at risk. This notation implies that I have previously multiplied the return by (-1). I treat the losses as positive values. Thus, λ_U captures the dependence related to stress periods.

2.4 ESE for Empirical Copula

In the context of entropy estimation, ESE empirical copula can be understood implicitly as a uniform prior in the relative entropy concept. Hence, I am most conservative in the sense that the estimated copula is as smooth as possible given the moment conditions. There are other ways to measure the dependence structure using the entropy concept. For instance, Miller and Liu (2002) use the mutual information which is defined as

$$I(f:g)=\int \log\left[\frac{f(x_1(x_1),x_2(x_2))}{g_1(x_1)g_2(x_2)}\right]dF(x_1,x_2)$$

to measure the degree of association among the variables. But $I(f:g)$ is not invariant under the increasing transformation of the marginal.

Several features in ESE make it more appealing for the copula estimation compared to kernel and log-spline estimators. First, the number of estimated parameters is largely reduced in the context of ESE. Second, kernel and log-spline estimators suffer from the boundary bias problem even if the domain of the copula function is restricted to $[0, 1]^d$. In addition, kernel estimator is more computationally intensive compared with the ESE as the dimension of copula increases, which makes it less attractive in practice.

In this section, I focus on the estimation of copula density by ESE. To make the notation traceable, bivariate formulation for the ESE empirical copula is presented. Generalization to higher dimensional cases is straightforward.

As in the univariate case, the problem is to maximize the bivariate exponential series of the copula density

$$W=-\int_{[0,1]^2} c(u,v)\log c(u,v)dudv \tag{2.9}$$

Subject to

$$\int_{[0,1]^2} c(u,v)dudv=1 \tag{2.10}$$

$$\int_{[0,1]^2} \phi_{ij}(u,v)c(u,v)dudv=\hat{\phi}_{ij}(u,v) \tag{2.11}$$

where $i=0,\ldots,n;\quad j=0,\ldots,m;$ and $i+j>0,$

u and v are uniform random variables and $\phi_{ij}(u,v)$ are a sequence of linearly independent polynomials. $\hat{\phi}_{ij}(u,v) = \frac{1}{N}\sum_{i=0}^{n}\sum_{j=0}^{m}\phi_{ij}(u,v)$ where $i + j > 0$ and N is the sample size. This estimator does not depend on the specification of the polynomials. The estimated density will be in the form

$$\hat{c}(u,v;\hat{\boldsymbol{\theta}}) = \exp[-\hat{\theta}_0 - \sum_{i=0}^{n}\sum_{j=0}^{m}\hat{\theta}_{ij}\phi_{ij}(u,v)], \quad i+j>0\,.$$

To ensure $\hat{c}(u,v;\hat{\boldsymbol{\theta}})$ is a proper density function,

$$\hat{\theta}_0 = \log\left[\int_{[0,1]^2}\exp(-\sum_{i=0}^{n}\sum_{j=0}^{m}\hat{\theta}_{ij}\phi_{ij}(u,v)dudv)\right], \quad i+j>0\,.$$

Therefore,

$$\hat{c}(u,v;\hat{\boldsymbol{\theta}}) = \frac{\exp[-\sum_{i=0}^{n}\sum_{j=0}^{m}\hat{\theta}_{ij}\phi_{ij}(u,v)]}{\int_{[0,1]^2}\exp[-\sum_{i=0}^{n}\sum_{j=0}^{m}\hat{\theta}_{ij}\phi_{ij}(u,v)]dudv} \tag{2.12}$$

where $i+j>0$ and $\boldsymbol{\theta} = [\theta_{01}, \theta_{10}, \theta_{11}, \ldots, \theta_{1m}, \ldots, \theta_{n1}, \ldots, \theta_{nm}]$.

As in the univariate case, I solve for $\boldsymbol{\theta}$ using Newton's method according to the following equation:

$$\hat{\boldsymbol{\theta}}_{(t+1)} = \hat{\boldsymbol{\theta}}_{(t)} - \mathrm{H}^{-1}\boldsymbol{b}, \tag{2.13}$$

where H is a mn by mn matrix. $\boldsymbol{b}$= $[b_{01}, b_{10}, b_{11}, \ldots, b_{1n}, \ldots, \theta_{m1}, \ldots, \theta_{mn}]$ and $b_{ij} = \int_{[0,1]^2}\phi_i(u)\phi_j(v)c(u,v;\hat{\boldsymbol{\theta}}_{(t)})dudv$.

The Hessian H is calculated as

$$H_{ij,kl} = \int_{[0,1]^2}\phi_{ij}(u,v)\phi_{kl}(u,v)c(u,v,\hat{\boldsymbol{\theta}}_{(t)})dudv.$$

In this paper, I set $\phi_{ij}(u,v) = u^i v^j$.

There is a problem associated with the choice of the order of polynomial m and n in the ESE. Wu (2007) suggests that *m* and *n* can be chosen automatically from the data. Two natural candidates are likelihood-based AIC and BIC. Preliminary experiments also suggest that both criterions provide reasonable performance in the selection of the order of polynomial.

In practice, given a data set with two variables x_1 and x_2, I can estimate the copula density as follows:

1. Estimate the marginal cumulative density functions $\hat{F}_1(x_1)$ and $\hat{F}_2(x_2)$ and denote $\hat{u} = \hat{F}_1(x_1)$ and $\hat{v} = \hat{F}_2(x_2)$.
2. Calculate $\hat{\phi}_{ij}(\hat{u},\hat{v}) = \hat{u}^i \hat{v}^j$ for $i = 0, \ldots, n;\quad j = 0, \ldots, m;$ and $i + j > 0$.
3. Solve $\hat{\boldsymbol{\theta}}$ using (2.13).
4. $\hat{c}(\hat{u},\hat{v};\hat{\boldsymbol{\theta}}) = \exp[-\hat{\theta}_0 - \sum_{i=0}^{n}\sum_{j=0}^{m}\hat{\theta}_{ij}\hat{\phi}_{ij}(\hat{u},\hat{v})],\quad$ where $i + j > 0$.

2.5 Monte Carlo Simulations

To investigate the finite sample performance of my proposed ESE copula estimator, I conduct an extensive Monte Carlo simulation study on estimation marginal densities, copula densities and bivariate densities. The *t*-distribution is commonly used in financial econometrics where the distribution of financial returns are usually fat tailed. To explore more on general densities, I also consider mixtures of univariate distribution studied by

Marron and Wand (1992). Specifically, I assume that the underlying margins are generated from the mixture of normal distributions: Gaussian, skewed-unimodal and bimodal (I assume both margins follow the same distribution). The bivariate copulas used in this study include the Gaussian copula, *t*-copula, Frank copula and Clayton copula. Each copula is able to capture a certain special dependence structure in the data. For various features of the copulas used in my study, see Chen et al. (2006). In my experiment, the dependence parameter for each type of copula is set to the corresponding Kendall's τ values 0.2, 0.4 and 0.6 in Table 2.1. The higher the value of Kendall's τ, the higher the association between two margins.

Figure 2.1 displays shapes of various copulas with Kendall's τ equal to 0.6. Note that all the copulas studied in this paper exhibited non-vanishing tails in either one or two tails which may cause the boundary bias problems (Bouezmarni and Rombouts 2007). In each experiment, the sample sizes I use are 50, 100 and 500 and each experiment is repeated 500 times. For comparison, I also estimate marginal and copula densities using the log-spline and kernel estimator. The cubic spline and product Gaussian kernel are used in this paper. The order of the polynomial is selected using the BIC in the ESE case; the smoothing parameter is chosen by the method of modified cross-validation and the number of knots is determined by $max(30,10n(2/9))$ for the log-spline estimator (Gu and Wang 2003), and the bandwidth of the kernel estimator is selected using the least square cross validation.

Since any bivariate density can be decomposed into two parts: the margins and the copula, I opt to examine the performance of the ESE, log-spline and kernel estimator for each component. Hence, I perform three sets of comparisons on the estimation of marginal density, copula density and joint density. In joint density estimation, I first estimate the margins by ESE, log-spline and kernel and estimate the copula in the second step by these three methods again. The product of the estimated margins and copula density is obtained as the estimated bivariate density.

Table 2.2 reports the mean and standard deviation of the Mean Integrated Squared Error (MISE) of the margins estimation. A basic fact is that the performance of MISE improves with the sample size. In addition, the more complicated the shape of the margins, the larger the MISE values are observed. For the sample sizes considered in this paper, the ESE outperforms the log-spline and kernel estimator in the case of Gaussian distribution. In this case, the MISE of the ESE is between 23% and 50% of kernel estimator, and from 50% to 80% in the case of log-spline estimator. This is expected since under the Gaussian margin, ESE nests the true distribution. On the other hand, for the skewed-unimodal and bimodal distribution, the log-spline estimator provides the best performance, followed by the ESE estimator. In case of the Student's *t*-distribution, ESE and log-spline estimators have comparable performance under small sample size but log-spline outperforms ESE estimator for the large sample. The simulation results suggest that: 1) log-spline estimator dominates kernel estimator in terms of MISE under the margins that I considered; 2) as long as the shape of the margins are relatively simple, ESE estimator provides better performance than the log-spline estimator in small sample. Overall, log-spline estimator provides the best performance. While for some cases, ESE estimator is comparable to or outperforms log-spline estimator. Kernel estimator is dominated by either estimator.

Table 2.3 reports the performance of the three estimators of various copula densities. Similar to the univariate case, the performance of estimators improve with the sample size. They are also positively related to the Kendall's τ. The higher the value of Kendall's τ, the higher the dependency between the margins. The copula has higher concentration close to the two tails for larger τ, and the shape of the copula become more acute near the tails in Figure 2.1. This made the boundary bias problem more significant. I also note that MISE decreases with sample size, but the decreasing rate is slower for a larger τ. For example, the MISE of Gaussian copula decreases by 60% from n = 50 to 500 when τ = 0.2; while its MISE decreases by 32% when τ = 0.6. Among the copulas I considered, Gaussian and Frank copulas show the smaller MISE values. For small τ, ESE shows slightly better performance than log-spline estimator in terms of the MISE in the copulas and sample size I considered. As the value of Kendall's τ increases, ESE outperforms log-spline estimator significantly. Again, ESE and log-spline estimator always outperforms kernel estimators in almost all the cases. The better performance of ESE can be explained by the fact that kernel estimator allocates weight outside the boundary and underestimate the underlying copula density at the tails. I also note that log-spline estimator and kernel estimator have the largest standard deviation in terms of MISE in the small sample (n = 50 and 100) and large sample respectively. I do not observe significant sample size effect on the performance among the three estimators.

Therefore, with respect to the copulas and size sample sizes I considered, I prefer the ESE in terms of the mean and standard deviation of the MISE.

So far I have assessed the relative performance of the three estimators for margins and copula estimation. In the case of margins estimation, the performance of log-spline and ESE are comparable; while ESE dominates the other two estimators in the case of copula estimation. Finally, kernel estimator is dominated by both of the estimators in both cases. Hence, I focus on the performance of ESE and log-spline estimators. Since a joint density is uniquely determined by its margins and copula, equation (2.5) immediately suggests two ways of estimating the joint density: estimate the joint density directly from the left hand side of equation (2.5) or estimate its margins and copula in two steps from the right hand side of equation (2.5). It is well known the direct estimation from the left hand side of equation (2.5) suffers from the "curse of dimensionality problem". In addition, the performance of ESE margin plus ESE copula related to log-spline margins plus ESE copula has not been assessed. In the third experiment, I am going to address the following questions. 1) How serve is the "curse of dimensionality problem" in the direct estimation comparing with the two-step estimation? 2) How is the performance of ESE margin plus ESE copula estimator compare with log-spline margins plus ESE copula in two-step joint density estimation?

Using the same setup of the margins in Table 2.2 and the copula densities in Table 2.3, I generate 4×4 = 16 joint densities. In this experiment, I set the sample size equals 50. The estimators that I consider for direct estimation are ESE, log-spline and kernel. In the two-step estimation, I first estimate the margins and then the copula. I examine two estimators here. I denote MM as ESE margins plus ESE copula estimator and LM as log-spline margins plus ESE copula estimator. In other words, I have five estimators in total to estimate each of the sixteen joint densities.

The MISE of estimated joint densities of various estimators are displayed in Table 2.4. In two step estimation, I focus on the performance of MM and LM estimators. Note that the MISE increases with Kendall's τ. This observation is similar to what I obtained in copula estimation in Table 2.3. If I move across the table, I note that Clayton copula has recorded the largest MISE as τ increases and this is true for all the margins I consider. As shown in Figure 2.1, Clayton copula has a relatively sharp tails near the boundary which complicates the estimation. Another observation is Gaussian and t margins exhibit smaller MISE in all the copulas and τ values. Again, relatively simple shapes of the margins reduce the estimation errors. Under the Gaussian margin, MM dominates LM for all Kendall's τ and the copulas that I considered. This is expected since under the Gaussian margin, the ESE nests the true margin. As the shape of the margins become more complicated or the value of τ increases, LM dominates MM in most of the cases. This illustrates the overall performance of LM is better than MM in the two-step estimation.

In the case of direct estimation, the patterns of MISE are similar to two-step estimation in terms of the margins and copulas under consideration. Except for bimodal margin, kernel is always dominated by ESE and log-spline. The MISE of kernel is approximately two times of either ESE or log-spline. But the performance of ESE and log-spline somehow does not have a notable clear cut. Under the Gaussian margin, ESE always outperforms log-spline and this is also what I observed in previous cases. The MISE under the ESE is 50% of the log-spline. In general, as the shapes of the margins become more complicated, log-spline dominates ESE.

To address the curse of dimensionality problem, I make comparison of copula estimation and direct estimation to see the extent of improvement in MISE under the copula estimation. Except for Gaussian margins, LM outperforms the other three estimators in almost all Kendall's τ and copulas I consider. Although a regular and consistent pattern cannot be observed for the difference of the MISE between LM and other three estimators, I average the MISE across the margins and calculate percentage of the MISE of LM to the MISE of the other estimators. The results of Table 2.5 indicate more than 50% of improvement is found for small τ using copula estimation, although the improvement decreases as the dependence between the variables increase. The improvements are in the order of Clayton > Frank > *t* > Gaussian in terms of copulas and K > M > L in terms of the direct estimation methods.

One of the potential benefits of copula estimation, as discussed in Hall and Neumeyer's paper (2006), is that if the margins are close to independent, the joint density can be estimated with a high degree of accuracy and the convergence rate can be improved from the two dimensional rate to that for single dimension. Their simulation results also reveal that for non-independent data, the copula still enhances the performance of estimation in finite sample compare with the direct estimation method. The simulation results based on the ESE copula estimation also shows there is substantial improvement in copula estimation in terms of the MISE as the dependence between the variables decreases. The results are consistent with Hall and Neumeyer's findings.

2.6 Empirical Application

An important question in risk management is whether the financial markets become more interdependent during financial crises. The fact that international equity markets move together more in downturns than in the upturns has been documented in the literatures, for example, see Longin and Solnik (2001), Forbes and Rigobon (2002). Hence, the concept of tail dependence plays an increasingly important role in measuring the financial contagion. If all stock prices tend to fall together as crisis occurs, the value of diversification might be overstated by ignoring the increase in downside dependence (Ang and Chen 2002). During the 1990s, several international financial crises occurred. Asian financial crisis is one of the crises that have been studied extensively in the literature. It started in Thailand with the financial collapse of Thai baht on July 2, 1997. News of the devaluation dropped the value of the baht by as much as 20%–a record low. As the crisis spread, most of Southeast Asia saw slumping currencies, devalued stock markets and asset prices.

Early studies on the dependence structure between financial assets are based on their correlations, which ignore the nonlinear dependence structure. Recent literatures used copula that capture the nonlinear dependence by assuming a parametric form of copula. They derive the corresponding value of tail dependence based on the estimated copula.

Parametric approach may be lack of flexibility and the estimated dependence will be biased if the copula is mis-specified. In this section, I model the dependence structure of the Asian stock markets returns using the ESE copula. No assumption on the dependence structure in the data is imposed. I emphasis that the results in the section are presented as an illustration of the maximum entropy estimation on copula density, rather than a detailed study of financial contagion in Asian financial crisis. Following Kim (2005) and Rodriguez (2007), I analyze the dependent structure for the Asian stock index returns by pairing with Thailand, the originator of the Asian financial crisis.

2.6.1 Data

The data used in this study are daily returns consisting of the daily stock-market indices of six East Asian countries from the DATASTREAM. When the Asian countries experienced the Asian financial crisis in 1997, these data present interesting case for the study of tail dependence as all these countries in the sample experienced a crisis of some severity during this period. Specifically, the data include Hong Kong Hang Seng (HK), Singapore Strait Times (SG), Malaysia Kuala Lumpur Composite (ML), Philippines Stock Exchange Composite (PH), Taiwan Stock Exchange Weighted (TW) and Thailand Bangkok S.E.T. (TH). The dataset covers the sample period from January 1994 to December 1998. I have altogether 1305 daily observations. I take the log-difference of each stock index and multiply by 100 to calculate the stock index returns. Following the standard practice, I assume each return series follows GARCH(1,1) model. The standardized residuals from the GARCH(1,1) models are used in the estimations.

2.6.2 Results

Table 2.6 gives summary statistics for the data. Standard deviation reveals that the most volatile market is Malaysia, followed by Thailand. All the series exhibits skewness and kurtosis relative to the Gaussian distribution. The Jarque-Bera test also demonstrates the non-normality of each series, which implies the violation of multivariate Gaussian distribution assumption. In fact, it is well known, according to Mandelbrot's (1963) paper, that most financial time series are fat-tailed. Existing studies often replaces the assumption of normality with the t distribution, which exhibits fat tails. Notice that Malaysia has the highest risk in terms of volatility, skewness, kurtosis and non-normality; while Taiwan ranks last in terms of volatility, kurtosis and non-normality. Figure 2.2 graphs the return data series for each market. Except for Taiwan, volatility increases substantially during the financial crisis in each market.

To investigate the dependence between different markets, I calculate the linear correlation and Kendall's τ. The linear correlation captures the linear dependence between two random variables. Kendall's τ is a nonlinear measure of dependence. In particular, it measures the degree of correspondence between two rankings. The estimated correlation and Kendall's τ are reported in Table 2.7.

The patterns revealed by two dependence measures are qualitatively similar. The linear correlations range from 0.19 (Taiwan and Thailand) to 0.65 (Singapore and Hong Kong) among the pairs I consider. Singapore has the highest average dependence with other countries, while Taiwan has the lowest average dependence. Although Thailand is suggested to play a trigger role in the Asian financial crisis, it only shows moderate dependence with other countries. On the other hand, the Kendall's τ ranges from 0.07 (Taiwan and Thailand) to 0.50 (Philippines and Singapore). All dependence measures rank the six markets consistently in decreasing order: Singapore, Hong Kong, Thailand, Malaysia, Philippines and Taiwan.

In Figure 2.3, I give the scatter plots corresponding to the five markets pairing with Thailand. The observations in Hong Kong, Singapore and Malaysia seem to be more clustered in upper and lower tails than Philippines and Taiwan. If I consider the occurrence of an extreme return event of one index given the return of Thailand is also extreme, this yields crude empirical estimates of lower and upper tail-dependence in the bivariate equity index returns. Table 2.8 reveals the results. The first cell in the table is 0.308, which means the probability of returns of Hong Kong being less than 5^{th} percentile given that the returns of Thailand is less than 5^{th} percentile equals 0.308. While Singapore has the strongest lower dependence, Hong Kong has the strongest upper dependence with Thailand. Philippines and Malaysia shows moderate tail dependence with Thailand and Taiwan seems to have the weakest tail dependence with Thailand. In general, the lower tail dependences are higher than the upper tail dependences. This fact is consistent with the literature that markets move together more in downturns than in upturns, which exhibits asymmetric tail dependence.

The next step is estimate the copula density. Frahm et al. (2005) show that using mis-specified parametric margins instead of nonparametric margin may lead to the wrong interpretations of dependence structure. Instead of assuming parametric margins, I estimate the margins by log-spline method in the first step since the simulation results in Table 2.2 indicates overall performance of log-spline is the best. The copula density is estimated by maximum entropy method in the second step.

Different dependence structures can be visualized by plotting their estimated copula densities along the diagonal $u = v$. Figure 2.4 shows the results. Notice that the scales in the graphs are different. Hong Kong, Singapore and Malaysia show stronger asymmetric shape with lower tail higher than upper tail; While Philippines and Taiwan have relatively symmetric shape. In the case of Hong Kong, most of the mass are concentrated in the two tails, which is suggested by the height of the estimated density with a small peak in the center of the density. Singapore and Malaysia also exhibit similar pattern with less mass concentrated on the two tails. On the other hand, Philippines and Taiwan shows a symmetric tail patterns and their densities are relatively flatter than the previous three markets.

In general, it is very difficult to find a parametric copula which can adequately explain the dependence structure based on the shape of the density. It is known that the dependence between two variables is completely captured by their copula. However, the commonly used one-parameter or two-parameter parametric copula functions only allow certain types of dependence structures. Thus the estimation of copula density by maximum entropy method provides more flexibility in modeling various dependence structures.

The lower and upper tail dependence coefficient can also be derived from the ESE copula density. One advantage of ESE copula is that an analytical solution the copula pdf is obtained. Using the definition of tail dependence, I integrate the estimated copula pdf to obtain the corresponding cumulative function. The results in Table 2.9 shows that the estimated lower tail dependences increase and upper tail dependences decrease somehow linearly in all the markets. Asymmetric tail dependences are observed in Hong Kong, Singapore and Malaysia, not in Philippines and Taiwan. Comparing with empirical tail dependence, ESE copula tends to underestimate the tails dependence at the very near tail. Still, Hong Kong exhibits the strongest lower and upper tails dependence with Tailand. The fact that lower tails dependence is stronger than the upper tails dependence still can be found.

2.7 Conclusions

This paper proposes a nonparametric estimator for copula densities. The estimator is based on multivariate exponential series estimation (ESE). ESE has an appealing information-theoretic interpretation and attains the optimal rate of convergence for nonparametric density in Stone (1982). More importantly, it overcomes the boundary bias of copula estimation. I examine the finite sample performance of the estimator in several simulations. The results show ESE outperforms the popular Gaussian kernel and log-spline estimator in copula estimation and provides superior estimates to empirical tail index in tail dependence estimation. I apply the ESE copula to estimate the dependence between Asian stock returns during the Asian financial crisis. Further research on this work can be done on different perspectives. One of the interesting directions is to derive the test for the tail asymmetries between the financial assets based on the ESE.

CHAPTER III

GOODNESS-OF-FIT VIA EXPONENTIAL SERIES LIKELIHOOD RATIOS

3.1 Introduction

For a given set of observations from a distribution with a continuous distribution function F, one important task is directed at developing procedures that are useful for drawing statistical inferences about the underlying population. Namely, one wishes to test $F(x) = F_0(x)$, where $F_0(x)$ denotes the hypothetical distribution. For a simple null hypothesis where the form of $F_0(x)$ is completely specified, many existing tests could be employed in a straightforward manner. One could also carry out tests based on the transformed sample $F_0(X_1), \ldots, F_0(X_n)$, leading to tests for uniformity on the interval [0, 1].

There is a long list of literature on goodness-of-fit tests. By far the most common formal statistical procedures for testing are those based upon the empirical distribution function, such as the Kolmogorov-Smirnov and the Cramér-von Mises tests, see for example Darling (1957). Both these tests are based upon distances of the empirical distribution function to the hypothesized distribution function. Refinements involve use of a weighting function, leading to a weighted measure of distance. Indeed for 50 years or so perhaps the preferred statistical procedure has been the weighted Cramér-von

Mises, or the Anderson-Darling statistic, of Anderson and Darling (1952). A fuller historical perspective and details of the many other procedures can be found, for example, in Conover (1999). Stephens (1974) provides a Monte Carlo comparison of the powers of those tests based upon the empirical distribution function. However, it is well-known that the powers of these empirical distribution test statistics are rather low (Stephens 1974).

The tests which based on the empirical distribution are sometimes referred to as omnibus tests, i.e. they are sensitive to almost all alternatives to the null. In the present context, this property implies that when an omnibus test fails to reject the null hypothesis, I can conclude that there is not enough evidence that the data is not generated from the null density. On the other hand, a rejection would not provide any information about the form of the density. Hence, Neyman's smooth test (1937) provides an alternative over the omnibus tests. The test requires selecting the number of components, which acts as a smoothing parameter. It can be viewed as a compromise between omnibus tests and tests whose power is focused in the direction of a specific alternative. Successive components of the smooth test can be directly related to changes in mean, variance, skewness and kurtosis. Surveys of the works on Neyman's smooth test are provided by Rayner & Best (1989) and Bera & Ghosh (2001). Further development along this line focuses on seeking the number of components in a data-driven manner, which yields the adaptive Neyman's tests for uniformity (Ledwina, 1994; Kallenberg and Ledwina, 1995; Fan, 1996).

Nonparametric kernel-based density estimation (KDE) introduced by Rosenblatt (1956) has also attracted several research efforts in tests for goodness-of-fit. As in the Neyman's smooth test, it is assumed implicitly that F possesses a probability density f, and therefore the original testing problem becomes equivalent to checking

$$H_0: f = f_0 \quad \text{against} \quad H_a: f \neq f_0$$

where f_0 corresponds to the probability density of F_0. Bickel and Rosenblatt (1973) proposed a test statistic based on the weighted L_2 distance between the kernel density estimation of f and its expected value computed under the null hypothesis, in which f_0 is fully specified. A test based on the derivative of a KDE was considered in Huang (1997). One drawback of these kernel-based tests arises from the "boundary bias" problem well-known in kernel density estimation. To improve the performance of KDE, boundary kernel functions may be employed; however the resulting test procedure becomes complicated in both implementation and asymptotic analysis.

The purpose of this paper is to introduce a nonparametric goodness of fit test based on the likelihood ratio test of Portnoy (1988). It is made nonparametric by utilizing the exponential series density estimator (ESE) of Crain (1974, 1976) and also Barron and Sheu (1991). This estimator is based on the method of maximum entropy density subject to a given set of moment conditions. Compare with the kernel estimator, the number of estimated parameters is largely reduced in the context of ESE. Furthermore, ESE does not suffer from boundary bias problem even the domain is restricted in [0, 1]. The test statistics are based on the ratio of an estimated ESE to that of the null hypothesis. It is motivated by the fact that Neyman's smooth test or score test is only an asymptotic approximation to the likelihood ratio test (Protnoy 1988). The order of the ESE is selected by BIC. I propose three further modifications on the test statistics: 1) instead of putting equal weight on the individual term of the exponential series, I consider geometric and exponential BIC average weights; and 2) rather than using nested sequence, I consider all subsets to select the optimal terms in the exponential series; 3) use the likelihood cross-validation to estimate the likelihood ratio statistic. My extensive Monte Carlo simulations show that the proposed tests enjoy good finite sample performances compared to the traditional methods such as Anderson-Darling tests. In addition, this data-driven method improves upon Neyman's score test. I conclude that the exponential series likelihood ratio test can complement the Neyman's type score test.

The plan for the rest of the chapter is as follows. The next section reviews some commonly used goodness of fit tests. Section 3.3 introduces the idea of maximum entropy and the new proposed exponential series likelihood ratio statistics. Section 3.4 shows the results of the power comparison among the tests and Section 3.5 concludes.

3.2 Reviews of Goodness-of-fit Tests

Suppose that a random sample of size n is given, I form the order statistic $X_1 < X_2 < \ldots < X_n$. I consider the empirical distribution function (EDF)

$$F_n(x) = 0 \quad if \quad x < X_1$$
$$F_n(x) = \frac{i}{n} \quad if \quad X_i < x < X_{i+1}$$
$$F_n(x) = 1 \quad if \quad X_n \leq x$$

$F_n(x)$ is a step function which is to be compared to the distribution function $F(x)$ corresponding to H_0. EDF statistics are based on the discrepancy between distribution function and the empirical distribution function of the sample of x values. These statistics include two classes: the supremum and the quadratic. For the supremum class, I concentrate on the well-known Kolmoglov-Smirnov statistics D. For the quadratic class, I concentrate on the Anderson-Darling statistic A^2 and the Cramer-Von Mises statistic W^2. The EDF statistics are consistent and unbiased. The tests discussed in this section are invariant under transformation of the random variable. Because of this feature, I can transform the distribution to the uniform distribution and restrict my discussion to the latter.

3.2.1 Probability Integral Transformation (PIT)

Let the random variable X have cdf $F(x)$. If $F(x)$ is continuous, the random variable Z produced by the transformation $Z = F(x)$ has the uniform probability distribution over the interval $0 \leq z \leq 1$. i.e.

$$g(z) = 1 \quad if \quad 0 \leq z \leq 1$$

Therefore,

$$G(z) = z$$

The underlying idea of this transformation is that the new EDF of Z, $G(z)$ is extremely simple and that it conserves the distribution of the random variables discussed in this section. Any goodness of fit test for a completely specified alternative reduces, via the probability integral transformation, to testing for uniformity.

3.2.2 Supremum Statistic

Kolmogorov proposed to use the maximum absolute difference as the test statistic.

$$D = \sup_x \left\{ | F_n(x) - \hat{F}(x) | \right\}$$

Let $z_i = \hat{F}(x_i)$, I have

$$D^+ = \max_i \left\{ \frac{i}{n} - z_i \right\}$$

$$D^- = \max_i \left\{ z_i - \frac{(i-1)}{n} \right\}$$

$$D = \max\left(D^+, D^-\right)$$

The Kolmoglov-Smirnov statistic is invariant under the PIT.

3.2.3 Quadratic Statistic

The Cramer-von Mises family of tests measures the integrated quadratic deviation of $F_n(x)$ from $F(x)$ suitably weighted by a weighting function ψ:

$$Q = n\int_{-\infty}^{\infty} [\hat{F}(x) - F_n(x)]^2 \psi(x) d\hat{F}(x)$$

With $\psi(x) = 1$, I get the Cramer-von Mises statistic W^2 and $\psi(x) = [\hat{F}(x)(1 - \hat{F}(x)]^{-1}$ leads to the Anderson-Darling statistic A^2. That is:

$$W^2 = n\int_{-\infty}^{\infty}[\hat{F}(x) - F_n(x)]^2 d\hat{F}(x)$$

$$A^2 = n\int_{-\infty}^{\infty}\frac{[\hat{F}(x) - F_n(x)]^2}{\hat{F}(x)[1-\hat{F}(x)]} d\hat{F}(x)$$

Let $z_i = \hat{F}(x_i)$, I have

$$W^2 = \sum_{i=1}^{n}\left[z_i - \frac{(2n-1)}{2n}\right]^2 + \frac{1}{12n}$$

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}[(2i-1)\log(z_i) + (2n+1-2i)\log(1-z_i)] \qquad (3.1)$$

The Anderson-Darling statistic A^2 weights strongly deviations near $z = 0$ and $z = 1$. This is justified because there the experimental deviation are small due to the constraints $[z - G_n(z)] = 0$ at $z = 0$ and $z = 1$.

3.2.4 Neyman's Smooth Test

Neyman (1937) replaced testing uniformity on [0,1] via parametric testing problem $H_0 : \theta^{(k)} = 0$ versus $H_a : \theta^{(k)} \neq 0$ in the exponential family of order *k*; That is,

$$c(\theta^{(k)})\exp\left\{\sum_{i=1}^{k}\theta_i b_i(x)\right\}$$

where $k \geq 1$, *k* is fixed, $x \in [0,1]$ and $\theta^{(k)} = (\theta_1, \ldots, \theta_k) \in R^k$. The Neyman's smooth test for testing uniformity on [0, 1] rejects null hypothesis H_0 for large values of

$$N_k = \sum_{j=1}^{k}\left[n^{-1/2}\sum_{i=1}^{n}b_j(X_i)\right]^2 \qquad (3.2)$$

where $X_1, X_2, \ldots, X_n$ are observed random variables and $b_1, b_2, \ldots$ are normalized Legendre polynomials on [0, 1]. Under the null hypothesis H_0, N_k has, an as $n \to \infty$, the central chi-squared distribution with k degrees of freedom. The most difficult problem related to this statistic is the choice of k. Recommendations in statistical literature are sometimes confusing. Some authors advocate a small number of components, whereas others show that in some situation a larger number of components is profitable. All existing suggestions concerning how to select k exploit in fact some preliminary knowledge about a possible alternative. Recently, Ledwina (1994) introduced a new data-driven method for selecting k in Neyman's smooth test of uniformity. Unlike earlier proposals, depending on alternatives of special interest, the new procedure provides an automatic choice of k, based on the data. Roughly speaking it works as follows. First, Schwarz's (1978) selection rule is applied to find a suitable dimension k, say, of an exponential family model for the data. Then Neyman's test is applied within the fitted model, resulting in the test statistics N_k. So, Schwarz's rule serves as a kind of selection, followed by the more precise instrument, being Neyman's test in the right dimension. This procedure relies on fitting the model to the data and verifying whether the difference between the model and he uniform distribution is significant. The author shows that simulated power of this data-driven version of Neyman's test also performs well in comparison with that of other tests.

3.3 New Test

3.3.1 Maximum Entropy Principle

The information entropy, the central concept of the information theory, was introduced by Shannon (1949). Entropy is an index of disorder and uncertainty. The maximum entropy (ME) principle states that among all the distributions that satisfy certain information constraints, one should choose the one that maximizes Shannon's information entropy. According to Jaynes (1957), the ME distribution is "uniquely determined as the one which is maximally noncommittal with regard to missing information, and that it agrees with what is known, but expresses maximum uncertainty with respect to all other matters."

The ME density is obtained by maximizing the entropy subject to some moment constraints. Let x be a random variable distributed with a probability density function $f(x)$, and X_1, X_2,..., X_n be an *i.i.d.* random sample of size n generated according to $f(x)$. The unknown density $f(x)$ is assumed to be continuously differentiable, positive on the interval of support (usually the real line if there is no prior information on the support of the density) and bounded. I maximize the entropy

$$\max: \quad W = -\int f(x) \log f(x) dx,$$

subject to

$$\int f(x)dx = 1,$$

$$\int g_k(x) f(x)dx = \hat{\mu}_k, \quad k = 1, 2, ..., K,$$

where $g_k(x)$ is continuously differentiable and $\hat{\mu}_k = 1/n \sum_{i=1}^{n} g_k(X_i)$. The solution takes the form

$$f(x, \hat{\boldsymbol{\theta}}) = \exp[-\hat{\boldsymbol{\theta}}_0 - \sum_{k=1}^{K} \hat{\boldsymbol{\theta}}_k g_k(x)] \tag{3.3}$$

where $\hat{\theta}_k$ is the Lagrangian multiplier associated with the k^{th} moment constraint in the optimization problem. To ensure $f(x, \hat{\boldsymbol{\theta}})$ integrates to one, I set

$$\hat{\boldsymbol{\theta}}_0 = \log\left[\int \exp[-\sum_{k=1}^{K} \hat{\boldsymbol{\theta}}_k g_k(x)]\, dx\right]$$

The maximized entropy $W = \hat{\boldsymbol{\theta}}_0 + \sum_{k=1}^{K} \hat{\boldsymbol{\theta}}_k \ \hat{\mu}_k$.

The ME density is of the generalized exponential family and can be completely characterized by the moments $Eg_k(x),\ k = 1, 2, \ldots, K$. I call these moments "characterizing moments", whose sample counterparts are the sufficient statistics of the estimated ME density $f(x, \hat{\boldsymbol{\theta}})$. A wide range of distributions belong to this family. For example, the Pearson family and its extensions described in Cobb et al. (1983), which nest the normal, beta, gamma and inverse gamma densities as special cases, are all ME densities with simple characterizing moments.

In general, there is no analytical solution for the ME density problem, and nonlinear optimization methods are required (Zellner and Highfield (1988), Ornermite and White

(1999) and Wu (2003)). I use Lagrange's method to solve this problem by iteratively updating $\boldsymbol{\theta}$

$$\hat{\boldsymbol{\theta}}_{(t+1)} = \hat{\boldsymbol{\theta}}_{(t)} - \mathrm{H}^{-1}\boldsymbol{b}$$

where for the $(t+1)^{th}$ stage of the updating, $b_k = \int g_k(x) f(x, \hat{\boldsymbol{\theta}}_{(t)}) dx - \hat{\mu}_k$ and the Hessian matrix H takes the form

$$\mathrm{H}_{k,j} = \int g_k(x) g_j(x) f(x, \hat{\boldsymbol{\theta}}_{(t)}) dx, \quad 0 \le k, j \le K.$$

The positive-definitiveness of the Hessian ensures the existence and uniqueness of the solution.

Given Equation (3.3), I can also estimate $f(x, \hat{\boldsymbol{\theta}})$ using the MLE. The maximized log-likelihood

$$l = \sum_{i=1}^{n} \log f(x_i, \hat{\boldsymbol{\theta}}) = -\sum_{i=1}^{n} [-\hat{\theta}_0 - \sum_{k=1}^{K} \hat{\theta}_k g_k(x_i)]$$

$$= -n[\hat{\theta}_0 + \sum_{k=1}^{K} \hat{\theta}_k \hat{\mu}_k] = -nW .$$

Therefore, when the distribution is of the generalized exponential family, MLE and ME are equivalent. Moreover, they are also equivalent to a method of moments (MM) estimator. This ME/MLE/MM estimator only uses the sample characterizing moments. Although the MLE and ME are equivalent in this case, there are some conceptual differences. For the MLE, the restricted estimates are obtained by imposing certain constraints on the parameters. In contrast, for the ME, the dimension of the parameter is determined by the number of moment restrictions imposed: the more moment restrictions, the more complex and at the same time the more flexible the distribution is.

To reconcile these two methods, I note that a ME estimate with the first m moment restrictions has a solution of the form

$$f(x,\hat{\boldsymbol{\theta}}) = \exp[-\theta_0 - \sum_{k=1}^{m} \theta_k g_k(x)],$$

which implicitly set θ_j, $j = m+1, m+2, \ldots,$ to be zero. When I impose more moment restrictions, say, $\int g_{m+1}(x) f(x,\boldsymbol{\theta}) dx = \hat{\mu}_{m+1}$, I let the data choose the appropriate value of θ_{m+1}. In this sense, the estimate with more moment restrictions is in fact less restricted, or more flexible. The ME and MLE share the same objective function (up to a proportion) which is determined by the moment restrictions of the maximum entropy problem. Therefore, one can regard the ME approach as a method of model selection, which generates a MLE solution.

3.3.2 The Test Statistic

I can use the ME approach for distribution tests. Consider a M dimension parameter space Θ_M. Suppose I want to test the hypothesis that $\boldsymbol{\theta} \in \Theta_m$, a subspace of Θ_M, where $m \leq M$. For $j = m, M$, let $\boldsymbol{\theta}_j$ be the MLE estimates in Θ_j, l_j and W_j be their corresponding log-likelihood and maximized entropy, I have

$$-\int f(x,\boldsymbol{\theta}_m) \log f(x,\boldsymbol{\theta}_m)\, dx$$

$$= -\int [\sum_{k=0}^{m} \theta_{m,k} g_k(x)]\ f(x,\boldsymbol{\theta}_m)\, dx$$

$$= \sum_{k=0}^{m} \theta_{m,k} \int g_k(x)\ f(x, \boldsymbol{\theta}_m)\, dx$$

$$= \sum_{k=0}^{m} \theta_{m,k} \int g_k(x)\ f(x, \boldsymbol{\theta}_M)\, dx$$

$$= \int [\sum_{k=0}^{m} \theta_{m,k} g_k(x)] \log f(x, \boldsymbol{\theta}_M)\, dx$$

$$= -\int f(x, \boldsymbol{\theta}_M) \log f(x, \boldsymbol{\theta}_m)\, dx$$

The forth equality follows because the first m moments of $f(x, \boldsymbol{\theta}_m)$ are identical to those of $f(x, \boldsymbol{\theta}_M)$. Consequently, the log-likelihood ratio

$$LR = -2(l_m - l_M)$$

$$= 2n(W_m - W_M)$$

$$= -2n[\int f(x, \boldsymbol{\theta}_m) \log f(x, \boldsymbol{\theta}_m)\, dx - \int f(x, \boldsymbol{\theta}_M) \log f(x, \boldsymbol{\theta}_M)\, dx]$$

$$= 2n[\int f(x, \boldsymbol{\theta}_M) \log f(x, \boldsymbol{\theta}_M)\, dx - \int f(x, \boldsymbol{\theta}_m) \log f(x, \boldsymbol{\theta}_m)\, dx]$$

$$= 2n \int f(x, \boldsymbol{\theta}_M) \log \frac{f(x, \boldsymbol{\theta}_M)}{f(x, \boldsymbol{\theta}_m)}\, dx,$$

which is the Kullback-Leibler distance between $f(x, \boldsymbol{\theta}_M)$ and $f(x, \boldsymbol{\theta}_m)$ multiplied by twice of the sample size. Consequently, if the true model $f(x, \boldsymbol{\theta}_M)$ nests $f(x, \boldsymbol{\theta}_m)$, the quasi-MLE estimate $f(x, \boldsymbol{\theta}_m)$ minimizes the Kullback-Leibler statistics between $f(x, \boldsymbol{\theta}_M)$ and $f(x, \boldsymbol{\theta}_m)$, as shown in White (1982).

Under the null hypothesis of uniformity, the maximum entropy $W_m = 0$. The test statistics of uniformity becomes

$$LR = -2nW_M \xrightarrow{d} \chi^2(M) \tag{3.4}$$

My proposed test is implemented in two steps. First I fix the upper bound of *m*, denoted by *M* and determine the optimal dimension m of the polynomial by BIC criterion, where $m \le M$ and BIC is calculated as

$$BIC_k = 0.5k \log n - l_k \quad where\ k = 1,\ldots,M$$

Select the model with index *I* defined by

$$I = [m, 1 \le m \le M : BIC_m = \min_{1 \le k \le M} BIC_k]$$

The second step is to calculate the likelihood ratio statistic $-2nW_m$, where *m* is the optimal dimension of the polynomial selected from BIC criterion in the first step. Reject H_0 of uniformity if $-2nW_m > C_{m,\alpha,n}$, where $C_{m,\alpha,n}$ is the simulated critical value for the test statistic of dimension *m* at significance level α with sample size *n*.

3.3.3 Weighting Functionality

The above specification assume that I place equal weight on $\boldsymbol{\theta}_k$, where $k = 1,\ldots,M$. I also propose to place different weighting schemes on $\boldsymbol{\theta}_k$. The first one is geometric weighting. Denote θ_{ki}^* be the weighted value of θ_{ki} for $i = 1,\ldots,k$ and $k = 1,\ldots,M$. θ_{ki}^* is calculated as follows:

$$\hat{\theta}_{ki}^{*} = \begin{cases} \frac{1}{k}\sum_{t=1}^{k} \hat{\theta}_{ti} & i = 1,\dots,k, \quad k = 1,\dots,M \\ \\ 0 & if \;\; i > k \end{cases} \tag{3.5}$$

The larger the value of *i*, the less weight I place on θ_{ki}.

The second weighting scheme is BIC weighting. The weighted BIC is related to the Bayesian Model Averaging (BMA). There is a large literature on BMA; see the review by Hoeting et. al. (1999). The weighting of $\hat{\theta}_{ki}$ is defined as:

$$w_{ki} = \frac{\exp(-0.5BIC_i)}{\sum_{j=1}^{k}\exp(-0.5BIC_j)} \tag{3.6}$$

3.3.4 BIC with All Subsets

Section 3.3.2 considers nested sequence models $\{1,\dots,M\}$ where *M* is the maximum order of the polynomial. In fact, if I define S^* to be a set with all the subset inside $S = \{1,\dots,M\}$, it may be considered as the other way to select the terms in the polynomials. Define analogously, select the model with index *I* defined by

$$I = [m^*, : BIC_{m^*} = \min_{k \in S^*} BIC_k] \tag{3.7}$$

BIC is calculated based the m^* as shown in section 3.3.2.

3.3.5 Maximum Likelihood Cross-Validation

This approach yields a density estimate which has an entropy interpretation, being that the estimate will be close to the actual density in the Kullback-Leibler sense. Likelihood cross-validation uses the leave-one-out strategy and find the density estimate at each x_i based on the other (n - 1) observations leaving out x_i. Thus one can perform maximum likelihood cross-validation by maximizing

$$l = \log L = \sum_{i=1}^{n} \log \hat{f}_{-i}(x_i) \tag{3.8}$$

where $\hat{f}_{-i}(x_i)$ is the leave-one-out maximum entropy estimator of $f(x_i)$. Likelihood cross-validation method may work well for a range of standard distribution or thin tailed distribution.

3.3.6 Relating LR Test to Neyman's Smooth Test

As Portnoy (1988) prove that the exponential series likelihood ratio statistics and Neyman's smooth statistics are asymptotically equivalent, i.e.

$$LR = -2nW = \sum_{j=1}^{k} \left[n^{-\frac{1}{2}} \sum_{i=1}^{n} b_j(X_i) \right]^2 + o_p(1)$$

That means, smooth statistic is only an approximation of likelihood ratio statistic under the small sample. In addition, Claeskens and Hjort (2004) prove that these two statistics do not have similar performances unless the alternative density is close to f_0. Hence, I suspect likelihood ratio statistic may outperform the Neyman's smooth statistic in terms of the power comparison under small sample.

3.4 A Simulation Study

3.4.1 Critical Values of the Test Statistic

To implement my test statistic proposed in Section 3, critical values of the test statistic should be determined from the null distribution. However, the test statistic employs the exponential series estimators only has the asymptotic distribution. Thus, critical values of the test statistic for the finite sample are determined by the means of Monte Carlo simulation.

To estimate the percentiles of the null distribution, 10000 random samples were generated from the standard uniform distribution for various values of sample size. For each $n \leq 100$, the critical value $C_{m,\alpha,n}$ of the test statistic $-2nW_m$ was calculated as follows. First fix n and α, and set the upper bound of m, denoted by M. Second, generate random numbers with size n from uniform distributed 10000 times. Third, choose the optimal value of m ($m \leq M$) by BIC criterion and calculate the corresponding likelihood ratio statistic $-2nW_m$ for each simulation. $C_{m,\alpha,n}$ is the (1-α) percentile of likelihood ratio statistics up to m. For example, if m = 2, the critical value is the (1-α) percentile of likelihood ratio statistics which include the order m = 1 and m = 2.

Table 3.1 displays the simulated critical values of the test statistics for subsequent values of K for n = 20, 50 and 100 at 5% significance level. This table shows that the critical values various as the sample size increases. For n = 20, likelihood ratio statistic has less variability in terms of the critical values compare with the Neyman's smooth statistic. Size distortion is more significant in Neyman's smooth statistic at K = 1 for the all the sample sizes that I considered ($\chi^2(1)$ = 3.841 at 5% significance level). As the sample size increases, both of the test statistics approach to $\chi^2(1)$ distribution and the critical values also become smaller for both test statistics.

3.4.2 Power Analysis

To compare the performance of the test statistics, I performed Monte Carlo simulations against the following 3 sets of alternatives:

1. $g_1(x) = \frac{1}{4}(x^{-\frac{1}{2}} + (1-x)^{-\frac{1}{2}})$

2. $g_2(x) = 1 + d\cos(\pi f x)$

   With the following parameters

   1. $d = .5, f = 1$

   2. $d = .5, f = 2$

   3. $d = .7, f = 4$

   4. $d = .9, f = 5$

3. $g_3(x) = c(\theta^{(s)}) \exp\left[\sum_{j=1}^{s} \theta_j b_j(x)\right]$

With the following parameters

1. $s = 1, \theta^{(1)} = .4$

2. $s = 2, \theta^{(2)} = (0, -.5)$

3. $s = 2, \theta^{(3)} = (.2, -.3)$

4. $s = 3, \theta^{(3)} = (0, 0, .6)$

5. $s = 4, \theta^{(4)} = (0, -.5, 0, -.2)$

6. $s = 4, \theta^{(4)} = (.1, .2, -.3, -.4)$

7. $s = 5, \theta^{(5)} = (0, 0, 0, 0, .5)$

8. $s = 8, \theta^{(8)} = (0, 0, 0, 0, 0, 0, 0, -.7)$

Alternative 1 (Figure 3.1) has the density mass cluster at the two tails. Alternatives 2 (Figure 3.2) with different *d* and *f* parameters capture the height and the periodic patterns of the densities. Alternatives 3 provide a flexible way to characteristic more general shape of the density function from low to high frequency by alternating the order of the polynomials and $\boldsymbol{\theta}$ (see page 1003, Ledwina 1994).

Alternatives 1 to 3 have been considered in Ledwina (1994) which compared several tests of uniformity. In particular, she focused on the Neyman's smooth test with BIC as criteria for order selection. This statistic performs well in comparison with other tests of uniformity.

A total of 10,000 random samples of size *n* equal to 20 and 50 were generated from each of the 3 set of alternative distributions. In Table 3.2 and Table 3.3, I compared the powers of my proposed test statistics for $n = 20$ and 50 and the number of added terms is allowed to grow to 10. I considered different versions of my test statistics: Uniform weighting, geometric weighting, BIC weighting and likelihood cross-validation, and compare with the corresponding Neyman's counterpart. Respective power of Anderson-Darling test (AD) is also presented.

From the results of Table 3.2 and Table 3.3, I observe that the performance of my proposed tests varied with the alternative distribution as well as the sample sizes. I am interested in the overall performance of my test statistics rather than the performance of the test statistics for a particular alternative distribution. For a small sample size (Table 3.2), Neyman's statistic loses power in comparison with the likelihood ratio statistic under all the weighting schemes and specification. In general, uniform weighting showed higher power compare with other specifications. In the first setting which density mass cluster at the two tails, geometric weighting performs worse than other tests in the experiment. In the second setting, BIC weighting and likelihood CV performs better in comparison with other specification consistently. In the third setting, geometric weighting and likelihood CV specification loses power under the high frequency alternative. Anderson-Darling statistic is powerful for some alternatives, for some others it has considerably smaller power than its competitors.

Power increases as the sample size increases. Differences in power are not pronounced for large sample size (Table 3.3). Power of the likelihood ratio test and Neyman's smooth test are comparable. In addition, the effects of weighting and specification have relatively smaller impact on the performance of my test statistics.

The main conclusion are the likelihood ratio test demonstrates higher power compare with the Neyman's smooth test under small sample and that difference weighting schemes do not improve on the uniform weighting scheme for the alternatives densities that I consider in the sample.

Table 3.4 compares the likelihood ratio test based on the nested sequence and all subset specification. When I consider both nested model sequence and all subset specification, the number of terms is allowed to grow to 6. The particular choices of where to cut off the series are not of much importance for power behavior (Ledwina, 1994). The results show that nested sequence specification has higher power than the all subset specification regardless of the sample size. The difference in terms of power is more pronounced as the sample size increases.

3.5 Conclusions

Score test is one of the most popular methods for goodness of fit test. Motivated by the fact that the score test is only the asymptotic approximation to the likelihood ratio test (Portnoy 1988), I introduced a nonparametric goodness of fit test which based on the exponential series estimator (Crain 1974, 1976) combined with the likelihood ratio test. My test statistic is based on the ratio of an estimated density to that of the null hypothesis. I cooperate with the BIC criterion for the order selection of the series estimator. The novelty of the test is that it has an appealing information theoretic interpretation and it overcomes the boundary bias of the kernel estimation for the estimated density.

I propose three further modifications on the test statistics: First, instead of putting equal weight on the individual term of the exponential series, I consider geometric and exponential BIC average weights; second, I consider all subsets to select the optimal terms in the exponential series; third, I also use the likelihood cross-validation method to estimate the likelihood ratio.

The extensive Monte Carlo simulations show that my proposed test has good small sample performance compared to the traditional methods such as Anderson Darling test. Uniform weighting scheme dominates other specifications. All subset specification is dominated by the nested sequence model. Finally, this data-driven method improves upon Neyman's score test. I conclude that the exponential series likelihood ratio test can complement the Neyman's type score test.

CHAPTER IV

FORECASTING EQUITY RETURNS: COPULAS, VARs AND COMPOSITES

4.1 Introduction

Modeling the joint distribution of financial asset is a very important issue in derivative pricing, risk management and portfolio allocations. In practice, models based on multivariate normal assumption (the so-called mean-variance approach) are widely used for their simplicity. However, several extreme events have made financial practitioners worried more about the nonlinear dependence of their portfolio returns and the joint normal specification fails to provide a good approximation for the dependence in general. Recently, a few works have devoted to the modeling of the joint distribution, while keeping the univariate distribution as flexible as possible, and to see how these models could be better for representing the nonlinear behavior, than the joint normal specification (Breymann et al. 2003). These works apply copula methodology, which allows for a decomposition of the joint distribution into marginal distributions and a copula function.

Copula functions are well studied objects in the statistical literature. These functions have been introduced to model a joint distribution once the marginal distributions are known. In fact, the essential idea of the copula approach is that a joint distribution can be factored into the marginals and a dependence function called copula. The dependence relationship is entirely determined by the copula, while scaling and shape (mean, standard deviation, skewness and kurtosis) are completely determined by the marginals. When multivariate normal distribution is rejected by the data, the copula may be used as an important alternative to represent the dependence in joint distributions. The evidence that copula based models fit financial data better than joint normality can be found in many recent works. However, research work dealing with financial asset forecasts which combines multivariate time series with copula are still in its starting. Patton (2007) provides a comprehensive review on the applications of copulas in modeling of financial time series.

In this perspective, I apply here a copula-VAR approach for index returns forecasting. The first contribution of the paper is a joint empirical analysis, by means of a copula-VAR model on S&P500 index return, federal fund rate and Commodity Research Bureau's (CRB) commodity spot index. The second contribution of the paper is the comparison between the copula-VAR model and the classical VAR model. In this perspective, I perform forecasting exercises considering 1-step-ahead forecasts to 4-steps-ahead forecasts on S&P500 index return, showing that when the margins are not normally distributed the copula approach yields better forecasts than the classical VAR model.

The rest of paper is organized as follows. Section 4.2 provides a brief literature review on forecasting. Section 4.3 describes the copula-VAR model, while an empirical application of the concerning the S&P500 index return is introduced in Section 4.4. An extensive forecasting exercise with the comparison among the alternative models are presented in Section 4.5. Section 4.6 concludes.

4.2 Literature Review on Forecasting

In the last 40 years, primarily due to the work of Box and Jenkins (1970), univariate autogressive moving-average (ARMA) model has been quite successful in time-series forecasting. One of the main problems with this type of model is that it does not explicitly take into account the influence of other observable variables known or suspected to be related to the series of interest. Vector autoregression (VAR) model represent one approach toward incorporating additional information and accounts for the dynamic nature of the data. It has been profoundly used in economic and finance researches such as policy analysis and forecasting. However, researchers have typically found that unrestricted VAR models do not forecast well (Kling and Bessler 1985). Classical VAR models impose assumptions that the disturbance terms across the equations are independent, identically and normally distributed. But there is a widely accepted observation that many economic and financial data are non-normally distributed. They exhibit fat-tails, skewness, and recent work suggests that some also exhibit "asymmetric dependence", where some pairs of variables are more highly correlated during negative movements than positive movements.

This observation raises two important problems: the construction of alternative, more palatable, density specifications, and the description and analysis of dependence between disturbance terms in a more general manner, as when the joint distribution of the variables of interest is non-elliptical the correlation coefficient is no longer sufficient to describe the dependence structure. More generally, the distribution of the asset returns will depend on the univariate distributions of the individual assets in the portfolio and on the dependence between each of the assets, which is captured by a function called copula.

Actually, copula is popularly using in economics and finance in the last decade: see, for example in Embrechts et al. (2003), Patton (2004, 2005) and Granger et al. (2006). Nelsen (2006) and Joe (1997) provide detailed introductions to copulas and their statistical and mathematical foundations, while Cherubini et al. (2004) focus primarily on applications of copulas in mathematical finance and derivatives pricing. In the multivariate time series context, Patton (2006) allows for time allows for time variation in the conditional copula by allowing the parameter(s) of a given copula to vary through time. Jondeau and Rockinger (2006) employ a similar strategy. Rodriguez (2007), on the other hand, considers a regime switching model for conditional copulas, in the spirit of Hamilton (1989). Chollete (2005), Garcia and Tsafack (2007) and Okimoto (2006) employ a similar modeling approach, with later author finding that the copula of equity returns during the low mean-high variance state is significantly asymmetric while the high mean-low volatility state has a more symmetric copula. Lee and Long (2006) combine copulas with multivariate GARCH models in an innovative way: they use copulas to construct flexible distributions for residuals from a multivariate GARCH model, employing the GARCH model to capture the time-varying correlation, and the copula to capture any dependence remaining between the conditionally uncorrelated standardized residuals.

Besides using individual model to forecast, Bates and Granger (1969) introduced the concept and methodology of forecast combination. It is well known that a linear combination of forecasts can outperform individual forecasts. Surveys and discussions of some subsequent developments and applications by Clemen (1989), Granger (1989), and Diebold and Lopez (1996) testify to the widespread interest that has developed this topic. The combination methods proposed by Bates and Granger require very simple calculations to determine appropriate weights, and moreover many studies have found that such straightforward methods generate high-quality forecasts. One would expect combination to be most effective when composites are formed from forecasts generated by widely divergent methodologies. For example, Montgomery et al. (1998) combine unemployment rate forecasts from nonlinear time-series models with median forecasts from a survey of professional economic forecasters. However, gains have also been reported from the combination of forecasts from apparently quite similar sources. For example, Newbold and Granger (1974), Bessler and Brandy (1981), Winkler and Makridakis (1983) and Granger and Ramanathan (1984) demonstrated benefits from combining alternative forecasts obtained through different univariate time-series methods. Many studies analysing practical applications of combination are briefly summarized in Clemen (1989).

4.3 Copula-VAR Model

To describe the interactions of returns and conditional variances in VAR model, I consider a general Copula Vector Autoregressive model, I define a VAR(p) model with dimension M and an error term $\sqrt{h_{l,t}}\eta_{l,t}$ as follows:

$$y_{l,t} = \beta_0^l + \sum_{i=1}^{M}\sum_{k=1}^{p}\beta_{i,k}^l y_{i,t-k} + \sqrt{h_{l,t}}\mu_{l,t} \tag{4.1}$$

where $l = 1,\dots,M$.

$\mu_{l,t}$ is the standardized innovations, $E(\mu_{l,t}) = 0$ and $\text{var}(\mu_{l,t}) = 1$. $\sqrt{h_{l,t}}$ can be constant or time-varying like in GARCH (1,1) models:

$$h_{l,t} = \alpha_0^l + \alpha_1^l(\mu_{l,t-1}\sqrt{h_{l,t-1}})^2 + \beta_1^l h_{l,t-1} \tag{4.2}$$

where $l = 1,\dots,M$.

From the Sklar's theorem (1959), I can express the conditional joint distribution $G_t(\mu_{1,t},\dots,\mu_{n,t};\boldsymbol{\theta})$ with parameters vector $\boldsymbol{\theta}$ can be expressed as follows:

$$(\mu_{1,t},\dots,\mu_{n,t}) \sim G_t(\mu_{1,t},\dots,\mu_{n,t};\boldsymbol{\theta}) = C_t(F_{1,t}(\mu_{1,t};\alpha_1),\dots,F_{n,t}(\mu_{n,t};\alpha_n);\boldsymbol{\rho}) \tag{4.3}$$

The joint distribution G_t of a vector of innovations $\mu_{l,t}$ is the copula $C_t(.;\boldsymbol{\rho})$ of the cumulative distribution functions of the copula from the innovations marginals $F_{1,t}(\mu_{1,t};\alpha_1),\dots,F_{n,t}(\mu_{n,t};\alpha_n)$, where $\boldsymbol{\rho}, \alpha_1,\dots,\alpha_n$ are the copula and marginals parameters, respectively.

4.4 Examples of Copula

Many copula families have been applied in financial literature to model the joint distributions of financial data. Their properties are well studied and their usefulness in representing the distribution tails is widely documented (see e.g. Embrechts et al. 2003 and Jondeau and Rockinger 2006). In the rest of this section, I consider a few examples of copulas which are popular in these types of application.

4.4.1 Elliptical Copulas

The class of elliptical distributions provides useful examples of multivariate distributions because they share many of the tractable properties of the multivariate normal distribution. Furthermore, they allow to model multivariate extreme events and forms of non-normal dependencies. Elliptical copulas are simply the copulas of elliptical distributions (see Fang et al. (1990) for a detailed treatment of elliptical distributions).

I present two copulas belonging to the elliptical family that are commonly used in empirical applications: the Gaussian and Student's *t*-copula. By applying Sklar's theorem and using the relationship between the distribution and the density function, I can derive their density functions. The copula of the multivariate Gaussian distribution is the Gaussian copula, and its probability density function is:

$$C(\Phi(x_1),\ldots,\Phi(x_n)) = \frac{f^{normal}(x_1,\ldots,x_n)}{\prod_{i=1}^{n} f_i^{normal}(x_i)} = |\Sigma|^{-1/2}\exp\left(-\tfrac{1}{2}\varsigma'(\Sigma^{-1}-I)\varsigma\right) \tag{4.4}$$

where $\varsigma = (\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_n))'$ is the vector of univariate Gaussian inverse distribution functions, $u_i = \Phi(x_i)$, while Σ is the correlation matrix.

The main properties of Gaussian copula are i) the symmetry, since the same correlation coefficient matrix Σ determines the dependence of both lower and upper tails; ii) a weak tail dependence. Gaussian copula may be used to define random variables with a Gaussian dependence structure, and nonnormal marginal distributions, featuring for instance fat tails.

The copula of the multivariate Student's t distribution is the Student's *t* copula, and its density function is:

$$C(t_{v_C}(x_1),\ldots,t_{v_C}(x_n)) = \frac{f^{Student}(x_1,\ldots,x_n)}{\prod_{i=1}^{n} f_i^{Student}(x_i)} = |\Sigma|^{-1/2}\frac{\Gamma(\frac{v_c+n}{2})}{\Gamma(\frac{v_c}{2})}\left[\frac{\Gamma(\frac{v_c}{2})}{\Gamma(\frac{v_c+1}{2})}\right]^n \frac{(1+\frac{\varsigma'\Sigma^{-1}\varsigma}{v_c})^{-(v_c+n)/2}}{\prod_{i=1}^{n}(1+\frac{\varsigma_i^2}{v_c})^{-(v_c+1)/2}} \tag{4.5}$$

where $\varsigma = (t_{v_c}^{-1}(u_1),\ldots,t_{v_c}^{-1}(u_n))'$ is the vector of univariate Student's *t* inverse distribution function, v_c are the degrees of freedom, $u_i = t_{v_c}(x_i)$, and Σ is the correlation matrix. The *t* copula dependence structure introduces an additional parameter compared with the Gaussian copula, namely the degrees of freedom v_c. Increasing the value of v_c decreases the tendency to exhibit extreme co-movements. It allows for joint fat tails and an increased probability of joint extreme events compared with the Gaussian copula. Unlike Gaussian copula, the closed form expression for *t*-copula is not available and it can only be calculated numerically.

Both Gaussian copula and *t*-copula are important examples of elliptical copula family and their dependence features are characterized by the correlation matrix Σ.

4.4.2 Archimedean Copulas

A drawback of elliptical copulas described above is that they feature symmetric dependence only. In many finance and insurance applications, however, it is reasonable to believe that big losses are more likely to take place together than big gains. Such asymmetries cannot be well represented by elliptical copulas. Another important class of copulas is the Archimedean copulas. Contrary to elliptical copulas, they allow for a great variety of dependence structures. An Archimedean copula is defined by Joe (1997)

$$C_{\Psi}(u_1, \ldots, u_n) = \Psi^{-1}[\sum_{i=1}^{n} \Psi(u_i)]$$

where Ψ is a convex, decreasing and positive function on [0,1], with $\Psi(0) = 1$; $\Psi(1) = 0$.

Among the different Archimedean copulas, Clayton and Gumbel copula are most commonly used. The Clayton copula corresponds to $\Psi(x) = x^{-\alpha_c} - 1$ and is given by:

$$C(u_1,\ldots,u_n)=\left(\sum_{i=1}^{n}u_i^{-\alpha_c}-n+1\right)^{-1/\alpha_c}$$

where α_c is the copula parameter and $0<\alpha_c<\infty$, whereas the density is given by

$$C(u_1,\ldots,u_n)=\alpha_c^n\frac{\Gamma(\frac{1}{\alpha_c}+n)}{\Gamma(\frac{1}{\alpha_c})}\left(\prod_{i=1}^{n}u_i^{-\alpha_c-1}\right)\left(\sum_{i=1}^{n}u_i^{-\alpha_c}-n+1\right)^{-\frac{1}{\alpha_c}-n} \tag{4.6}$$

t-copula allows for joint extreme events, but not for asymmetries. If one believes in the asymmetries in equity return dependence structures reported by for instance Longin and Solnik (2001) and Ang and Chen (2002), the *t*-copula may also be too restrictive to provide a reasonable fit. Then, the Clayton copula, which is an asymmetric copula, exhibiting greater dependence in the negative tail than in the positive, might be a better choice.

Gumbel copula corresponds to $\Psi(x)=(-\ln(x))^{\alpha_G}$ and is defined as:

$$C(u_1,\ldots,u_n)=\exp\left[\left(-\sum_{i=1}^{n}(-\log u_i)^{\alpha_G}\right)^{1/\alpha_G}\right]$$

for $\alpha_G\geq 1$ and the corresponding density function is

$$\begin{aligned}c(u_1,\ldots,u_n)=C(u_1,\ldots,u_n)(\prod_{i=1}^{n}u_i)^{-1}(\sum_{i=1}^{n}(-\log u_i)^{\alpha_G})^{-n+n/\alpha_G}(\prod_{i=1}^{n}\log u_i)^{\alpha_G-1}\\ \times\left[1+(\alpha_G-n+1)(\sum_{i=1}^{n}(-\log u_i)^{\alpha_G})^{-1/\alpha_G}\right]\end{aligned} \tag{4.7}$$

The Gumbel copula is also an asymmetric copula, as opposite to the Clayton copula, exhibits greater dependence in the positive tail than in the negative.

4.5 Two Stage Estimations for Margins and Copula

Pure parametric models may impose too strong restrictions with a risk of misspecification, while pure non-parametric models are more flexible, but can require too many observations to get a reasonable fit. In this section, instead, I consider a semi-parametric specification of the joint distribution, in which the marginal distributions are left unspecified, whereas the copula is parameterized. Such a semiparametric model can be estimated by using a two stage procedure. In a first step the marginal distributions are estimated by their sample counterparts. Then, these first step estimations are used to approximate the ranks corresponding to the observed returns. Finally, the copula parameters are estimated by maximizing the log-likelihood function applied on the estimated ranks. This procedure initially suggested by Genest et al. (1995) is detailed below.

For each variable $X_1, \ldots, X_n$ with T observations, I estimate the corresponding marginal cdfs by nonparametric kernel method, which is given by:

$$\hat{F}_i(x_i) = \frac{1}{Th}\sum_{j=1}^{T} K\left(\frac{x_i - x_{ij}}{h}\right)$$

where h is the bandwidth parameter and K is the second order Gaussian kernel function. In practice, h is selected optimally based on the least square cross-validation method.

These estimated $\hat{F}_i(x_i)$, where $i = 1, \ldots, n$ are approximately uniformly distributed on [0,1] (when T is large). Computing the multivariate transformed observations $[\hat{F}_1(x_1), \ldots, \hat{F}_n(x_n)]$ is the first step for the copula fitting. Then, several copula families can be estimated from these multivariate series of estimated cdf. For this purpose I can maximize the log-likelihood function of the transformed data:

$$L(\theta; X) = \sum_{i=1}^{T} \log c(\hat{F}_1(x_{1i}), \ldots, \hat{F}_n(x_{ni}); \theta) \tag{4.8}$$

The log-likelihood estimator $\hat{\theta}$ that maximizes the above likelihood is consistent and asymptotically normal, when T tends to infinity. However, it is not semi-parametrically efficient (see e.g., Genest et al., 1995).

4.6 Empirical Application

The data used in this empirical study are the S&P500 index return obtained from EconStats and the effective federal funds rate obtained from the Federal Reserve Bank of St. Louis. In order to control for the effects of inflation on the monetary policy innovation by including the Commodity Research Bureau’s (CRB) commodity spot index. The sample data is daily from May 26, 1981 to December 31, 2007, with number of observations equal 6940. Note that, I use the first 6910 observations for model estimation and leave the last 30 observations for forecasts comparison. The same data was used by Crowder (2006) to examine the role of stock market in the Federal Reserve policy rule using VAR model. The stock returns are calculated as the annualized log change in the daily closing price of the S&P500 index. The data are plotted in Figure 4.1.

Table 4.1 gives summary statistics for the data. Standard deviation reveals that the volatility of the series is in the order of Federal funds rates, S&P500 index returns and CRB. All the series exhibits skewness and kurtosis relative to the Gaussian distribution. The Jarque-Bera test also demonstrates the non-normality of each series, which implies the violation of multivariate Gaussian distribution assumption. Notice that S&P500 has the highest risk in terms of skewness, kurtosis and non-normality.

To investigate the dependence between different markets, I calculate the linear correlation and Kendall's τ. The linear correlation captures the linear dependence between two random variables. Kendall's τ is a nonlinear measure of dependence. In particular, it measures the degree of correspondence between two rankings. The estimated correlation and Kendall's τ are reported in Table 4.2.

The patterns revealed by two dependence measures are qualitatively similar. Although the magnitude of all the pairs of correlations are very small, they are significant at 1% level since the sample size is large. The linear correlation range from -0.0373 (Federal rate and CRB) to 0.0024 (S&P500 and CRB) among the pairs I consider. On the other hand, the Kendall's τ ranges from -0.0298 (Federal rate and CRB) to 0.0019 (S&P500 and CRB). In addition, S&P500 and Federal fund rate shows an inverse relationship which is consistent with the monetary theory: A tighter monetary policy raises the federal fund rate, which increases the discount rate, in turn causing stock prices to decline.

To start with the copula-VAR model, I examine the time series property of the data. A careful analysis of the levels and of the first log-differences of the three series shows that S&P500 returns is I(0), while federal funds rate and CRB price index are both I(1) variables in the sample period. The first step is to filter the original data using VAR for linear correlation and GARCH (1,1) model for the volatility in each series. I used the Bayesian information criterion (BIC) to select the most appropriate lag order of the VAR model. The optimal choice resulted to be a model with four lags.

The next step is to use the filtered series to estimate the cdf of the three variables using kernel methods. Here I adopt a second-order Gaussian kernel density function and the optimal bandwidth is obtained by least square cross-validation. The Gaussian copula, the *t*-copula, the Clayton copula and Gumbel copula are then estimated by the two-step procedure for the three transformed series.

Comparisons between a set of competing copula-based models can be done via statistical criteria. Likelihood ratio tests are often used, see Vuong (1989) and Rivers and Vuong (2002) for example. But these tests are often criticized by their low power against the alternative (Malevergne and Sornette 2006). Alternatively, information criteria, such as Schwarz's Bayesian Information (BIC) is used to penalize models with more parameters.

Table 4.3 provides the loglikelihood and the BIC values using the maximum likelihood approach on (4.8). The smaller the BIC value or the larger the likelihood value is, the better is the goodness-of-fit. For the four models considered, the Clayton copula model provides the best fit according to the BIC and the log-likelihood criterion. Note that the BIC criteria put a heavy penalty on the normal and *t*-copula. One of the reasons why Clayton copula dominates other copula is that I am dealing with the financial data and most of the financial data are found to exhibits greater dependence in the negative tail than in the positive, as reported by Longin and Solnik (2001) and Ang and Chen (2002). This is exactly one of the important features for Clayton copula.

4.7. Forecasts and Forecast Evaluation

Based upon the model specifications determined earlier, I re-cursively generate 30 one-step-ahead forecasts, 29 two-steps-ahead, 28 three-steps-ahead, and 27 four-steps-ahead forecasts for the models that I consider in this paper.

First, one-step-ahead forecast $\hat{x}_{t+1}$ is derived from the model (4.1) from the actual value of x using information available up to time t. This forecast value is then used to forecast two-steps-ahead $\hat{x}_{t+2}$, and so on.

The forecast errors are calculated as the difference between the forecast and the actual index returns. While criticisms exist over the use of MSE-type criteria in forecast comparison (e.g. Clements and Hendry 1993), I choose mean squared forecast error (MSFE) as the criterion for its simplicity and popularity. A summary of forecast performance of the three models is given in Table 4.4.

The MSFEs of the copula-VAR model are 0.287, 0.381, 0.498 and 0.457 for one up to four-steps-ahead forecasts, respectively. They are all less than the MSFEs of the classical VAR and ARIMA models. The allowance of more flexibility on the dependence structure also seems to increase the accuracy of index returns forecasts in all four forecasting horizon's. If I compare VAR and ARIMA models, the gain in forecasts is not that much for VAR model. The other two variables in the VAR model do not help much in terms of the forecasting accuracy.

Of course, forecasts can also be formally compared in a number of other ways. It is possible that two sets of forecasts are visually different from each other, but they may not be so statistically. One way is to test the equality of forecast mean squared errors or some other measure of economic loss. In an important contribution to the literature on forecast evaluation, Diebold and Mariano (1995) proposed a formal, yet rather intuitive, test procedure. For a pair of *h*-steps-ahead forecast error $(e_{it}, i = 1, 2, t = 1, \ldots, T)$, the quality of forecast is to be judged on some specific function *g{e)* of the forecast error *e* (MSFE is often used). Following Harvey et al. (1997), the null hypothesis of equality of expected forecast performance is:

$$E[g(e_{1t}) - g(e_{2t})] = 0$$

Defining a new series by

$$d_t = g(e_{1t}) - g(e_{2t})$$

The Diebold-Mariano test statistic is then

$$DM = [\hat{V}(\bar{d})]^{-1/2}\bar{d}$$

where $\bar{d}$ is the sample mean of d_t, $\hat{V}(\bar{d})$ is the sample variance of $\bar{d}$ which asymptotically can be estimated by $T^{-1}[\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k]$, where γ_k is the k^{th} autocovariance of d_t, and can be estimated by $T^{-1}\sum_{t=k+1}^{T}(d_t - \bar{d})(d_{t-k} - \bar{d})$. Under the null hypothesis, this statistic has an asymptotic standard normal distribution. The simulation evidence shows, however, that this test statistic could be seriously oversized in the case of multi-steps-ahead forecasts ($h \geq 2$). Harvey et al. (1997) suggested two modifications: One is to adjust the degree of freedom,

$$MDM = \left[\frac{T + 1 - 2h + T^{-1}h(h-1)}{T} \right]^{1/2} DM \tag{4.9}$$

And the second adjustment is to compare the statistic with critical values for the Student's *t*-distribution with (T-1) degrees of freedom, rather than from the standard normal distribution.

The modified Diebold-Mariano test results are given in Table 4.5. Noticeably, the hypothesis of equality of forecast errors from the pairwise comparison of the models cannot be rejected at 0.10 significance level in all cases except in one-steps-ahead forecasts for the copula-VAR model vs VAR and ARIMA models. This, together with relatively low *p*-values in the higher horizons means to point to the different performance between copula-VAR model and other two models. In addition, The *p*-values associated with VAR-ARIMA pair are high, implying that these two models statistically perform similarly well in the horizons that I consider.

To move beyond the equality tests, a more stringent requirement would be that the competing forecasts embody no useful information absent in the preferred forecasts. This is the basic idea of forecast encompassing. Encompassing is closely related to composite forecasting and is essentially a type of conditional misspecification analysis (Hendry, 1995, chapter 4). Denote two forecast error series by e_{it}, $i = 1, 2$ as before, and the composite forecast error by ε_t, a white noise term, and write

$$e_{1t} = \lambda(e_{1t} - e_{2t}) + \varepsilon_t \tag{4.10}$$

The null hypothesis is $\lambda = 0$. When the null is true, according to Chong and Hendry (1986), the first forecast encompasses the second. The actual test involves an ordinary least squares regression of e_{it} on $(e_{1t} - e_{2t})$. A *t*-test of $\hat{\lambda}$ is used as my test for encompassing.

Table 4.6 shows the encompassing test results in probability form. The null hypothesis is that the forecasts of generated from a model (in a column) encompass the forecasts of a model in the row. For example, the first entry in the second row is 0.04, and VAR model encompasses the copula-VAR model at 0.1 significance level is rejected. Both of the VAR and ARIMA models does not encompass copula-VAR model in one and two-steps-ahead forecast. Starting at three-steps-ahead forecast, no encompassing evidence is found. That means, the copula-VAR model has some advantage over the VAR and ARIMA model in one- and two-steps-ahead forecasts. In general, these encompassing test results are consistent with the findings of the equality tests.

Follows Granger and Ramanathan (1984), Table 4.7 considers three alternative approaches to obtaining linear combinations of forecasts. Method A minimizes the sum of squared errors of forecasts to obtain the weights; Method B is the case in which weights are constrained to sum to unity; Method C has no restrictions on the weights, but a constant term is added. Combination of three forecasts is better than any combination of pair. The results also demonstrate that Method C is consistently better, in terms of mean square error, than either Methods A or B. It is interesting to note that Method A which is unconstrained, has weights adding nearly to one, but this is not true for Method C, where the weights on the forecasts add to substantially less than one.

4.8 Conclusions

In this paper I analyzed data on S&P500 index returns in an effort to provide evidence of whether allowing a more flexible dependence structure on the disturbance terms can improve the forecasting ability of classical VAR model. I consider a trivariate copula-VAR model with four lag on the VAR model and Clayton copula. While in many cases, the copula-VAR improves the model's forecasting performance. I also demonstrate that the combined forecasts outperform individual forecasts.

Future research may explore this subject in the following ways. First, use more general formulation of the copula in terms of the number of parameters; second, experiment with data with different frequency; third, examine the longer forecasting horizons if necessary.

CHAPTER V

SUMMARY

This dissertation has attempted to apply the nonparametric exponential series estimation (ESE) to model the copula function and performing the goodness-of-fit test. In addition, I apply the copula-VAR model to make forecast on the stock index returns.

In the first essay I propose a multivariate exponential series estimator (ESE) to estimate the copula density nonparametrically. Extensive Monte Carlo studies show the proposed estimator outperforms the popular Gaussian kernel and log-spline estimator in copula estimation and provides superior estimates to empirical tail index in tail dependence estimation. It also demonstrates two-step density estimation through an ESE copula often outperforms direct estimation of the joint density. I apply the ESE copula to estimate the dependence between Asian stock returns during the Asian financial crisis. The results show that there exist asymmetric higher lower tail dependences among some of the Asian countries during the Asian financial crisis which is consistent with the literatures that the financial assets show stronger dependency during the financial crisis. Based on this work, further researches can be done on the derivation for the test of the tail asymmetries between the financial assets based on the ESE.

In the second essay I propose a likelihood ratio statistic using a nonparametric exponential series approach. The order of the series is selected by BIC. I propose three further modifications on my test statistics based on the weighting, sub-setting and likelihood cross-validation estimation method on the proposed test statistic. My extensive Monte Carlo simulations show that the proposed tests enjoy good finite sample performances compared to the traditional methods such as the Anderson-Darling tests. Uniform weighting scheme dominates other specifications. All subset specification is dominated by the nested sequence model. In addition, this data-driven method improves upon Neyman's score test under small sample. I conclude that the exponential series likelihood ratio test can complement the Neyman's type score test.

In the last essay, I analyzed data on S&P500 index returns in an effort to provide evidence of whether allowing a more flexible dependence structure on the disturbance terms can improve the forecasting ability of classical VAR model. I consider a trivariate copula-VAR model with four lag on the VAR model and Clayton copula. In addition, I compare the forecast performance of Copula-VAR model, a classical VAR model and a univariate time series model. In many cases, the copula-VAR improves the model's forecasting performance based on the mean forecast square error. The findings suggest that allowing a more flexible specification in the error terms using the copula tends improve the forecast accuracy. I also demonstrate that the combined forecasts outperform individual forecasts. Future researches may explore this subject in three directions. First, use more general formulation of the copula in terms of the number of parameters; second, experiment with data with different frequency; third, examine the longer forecasting horizons if necessary.

REFERENCES

Anderson, T. W., and Darling, D. A. (1952), "Asymptotic Theory of Certain 'Goodness-Of-Fit' Criteria Based on Stochastic Processes," *Annals of Mathematical Statistics,* 23, 193-212.

Ang, A., and Chen, J. (2002), "Asymmetric Correlation of Equity Portfolios," *Journal of Financial Economics,* 63, 442-294.

Barron, A., and Sheu, C. H. (1991), "Approximation of Density Functions by Sequences of Exponential Families," *Annals of Statistics*, 19, 1347-1369.

Bates, J. M., and Granger, C. W. J. (1969), "The Combination of Forecasts," *Operational Research Quarterly*, 26, 325-329.

Bera, A. K., and Ghosh, A. (2001), "Neyman's Smooth Test and Its Applications in Econometrics," in *Handbook of Applied Econometrics and Statistical Inference,* eds. A. Ullah, A. Wan and A. Chaturvedi, New York: Marcel Dekker, pp. 170-230.

Bessler, D. A., and Brandy, J. A. (1981), "Forecasting Livestock Prices with Individual and Composite Methods," *Applied Economics*, 13, 513-522.

Bickel, P. J., and Rosenblatt, M. (1973), "On Some Global Measures of the Deviation of Density Function Estimates," *Annals of Statistics*, 1, 1071-1095.

Bouezmarni, T., and Rombouts, J. (2007), "Nonparametric Density Estimation for Multivariate Bounded Data," Working paper, Centre Interuniversitaire sur le Risque, les Politiques Economiques et l' Emploi, Montreal, Canada.

Box, G. E. P., and Jenkins, G. M. (1970), *Time Series Analysis, Forecasting and Control*, San Francisco. CA: Holden-Day.

Breymann, W., Dias, A., and Embrechts, P. (2003), "Dependence Structures for Multi variate High-Frequency Data in Finance," *Quantitative Finance*, 3, 1-14.

Chen, S. (1999), "A Beta Kernel Estimation for Density Functions," *Computational Statistics and Data Analysis*, 31, 131-145.

Chen, S. X., and Huang, T. (2007), "Nonparametric Estimation of Copula Functions for Dependence Modeling," Unpublished manuscript, Department of Statistics, Iowa State University, Ames, Iowa, U.S.A.

Chen, X. H., Fan, Y. Q., and Tsyrennikov, V. (2006), "Efficient Estimation of Semiparametric Multivariate Copula Models," *Journal of American Statistical Association*, 101, 1228-1241.

Cherubini, U., Luciano, E., and Vecchiato, W. (2004), *Copula Methods in Finance*, New York: Wiley.

Chollete, L. (2005), "Frequent Extreme Events? A Dynamic Copula Approach," Working paper, Norwegian School of Economics and Business.

Chong, Y. Y., and Hendry, D. F. (1986), "Econometric Evaluation of Linear Macroeconomic Models," *Review of Economic Studies*, 53, 671-690.

Claeskens, G., and Hjort, N. L. (2004), "Goodness of Fit via Non-parametric Likelihood Ratios," *Scandinavian Journal of Statistics*, 31, 487–513.

Clemen, R. T. (1989), "Combining Forecasts: A Review and Annotated Bibliography," *International Journal of Forecasting*, 5, 559-583.

Clements, M. P., and Hendry, D. F. (1993), "On the limitations of Comparing Mean Square Forecast Errors," *Journal of Forecasting*, 12, 617-637.

Cobb, L., Koppstein, P., and Chen, N. (1983), "Estimation and Moment Recursion Relations for Multimodal Distributions of the Exponential Family," *Journal of American Statistical Association*, 8, 124-30.

Coles, S. G., Currie, J., and Tawn, J. A. (1999), "Dependence Measures for Extreme Value Analysis," Unpublished manuscript, Department of Mathematics and Statistics, Lancaster University.

Conover, W. J. (1999), *Practical Non-parametric Statistics*, 3rd edition, New York: Wiley.

Crain, B.R. (1974), "Estimation of Distributions Using Orthogonal Expansions," *Annals of Statistics*, 2, 454–463.

——— (1976), "More on Estimation of Distributions Using Orthogonal Expansions," *Journal of American Statistical Association*, 71, 741–745.

Crowder, W. J. (2006), "The Interaction of Monetary Policy and Stock Returns," *The Journal of Financial Research*, 29, 523-535.

Darling, D. A. (1957), "The Kolmogorov-Smirnov, Cramer-von Mises Tests," *Annals of Mathematical Statistics*, 28, 823-838.

Deheuvels, P. (1979), "La Function de Dependance Empirique et Ses Proprietes. Un Test Non Paramtrique d'independence," *Académie Royale de Belgique, Bulletin de la Classe des Sciences*, 65, 274-292.

Diebold, F. X., and Lopez, J. A. (1996), "Forecast Evaluation and Combination," in *Handbook of Statistics*, v.14, ed. G. S. Maddala and C. R. Rao, Amsterdam: North Holland, pp. 241-268.

Diebold, F. X., and Mariano, R. S. (1995), "Comparing Predictive Accuracy," *Journal of Business and Economic Statistics*, 13, 253-263.

Embrechts, P., Lindskog, F., and McNeil, A. (2003), "Modelling Dependence with Copulas and Applications to Risk Management," in *Handbook of Heavy Tailed Distributions in Finance*, eds. S. Rachev, Amsterdam: Elsevier, pp. 329-384.

Embrechts, P., McNeil, A., and Straumann, D. (1999), "Correlation: Pitfalls and Alternatives," *Risk*, 5, 69-71.

Fan, J. (1996), "Test of Significance Based on Wavelet Thresholding and Neyman's Truncation," *Journal of the American Statistical Association*, 91, 674-688.

Fang, K.T., Kotz, S., and Ng, K. W. (1990), *Symmetric Multivariate and Related Distributions*, New York: Chapman and Hall.

Fermanian, J. D., and Scaillet, O. (2003), "Nonparametric Estimation of Copulas for Time Series," *Journal of Risk*, 5, 25-54.

Forbes, K., and Rigobon, R. (2002), "No Contagion, Only Interdependence: Measuring Stock Market Co-movements," *Journal of Finance*, 57, 2223-62.

Garcia, R., and Tsafack, G. (2007), "Dependence Structure and Extreme Comovements in International Equity and Bond Markets," Working paper, Université de Montreal.

Genest, C., Ghoudi, K., and Rivest, L. (1995), "A Semiparametric Estimation Procedure of Dependence Parameters in Multivariate Families of Distributions," *Biometrika*, 82, 543-552.

Genest, C., and Rivest, L. P. (1993), "Statistical Inference Procedures for Bivariate Archimedean Copulas," *Journal of American Statistical Association*, 88, 1034-1043.

Gijbels, I., and Mielnicnuk, J. (1990), "Estimating the Density of a Copula Function," *Communications in Statistics A*, 19, 445-64.

Golan, A., Judge, G., and Miller, D. (1996), *Maximum Entropy Econometrics: Robust Estimation with Limited Data*, New York: John Wiley and Sons Press.

Granger, C. W. J. (1989), "Combining Forecasts – Twenty Years Later," *Journal of Forecasting*, 8, 167-173.

Granger, C. W. J., Patton, A. J., and Terasvirta, T. (2006), "Common Factors in Conditional Distributions for Bivariate Time Series," *Journal of Econometrics*, 132, 43-57.

Granger, C. W. J., and Ramanathan, R. (1984), "Improved Methods of Combining Forecasts," *Journal of Forecasting*, 3, 197-204.

Gu, C., and Wang, J. (2003), "Penalized Likelihood Density Estimation: Direct Cross-Validation and Scalable Approximation," *Statistica Sinica*, 13, 811-826.

Hall, P., and Neumeyer, N. (2006), "Estimating a Bivariate Density When There Are Extra Data on One or Both Components," *Biometrika*, 93, 439-450.

Hamilton, J. D. (1989), "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357-384.

Harvey, D. I., Leybourne, S. J., and Newbold, P. (1997), "Testing the Equality of Prediction Mean Square Errors," *International Journal of Forecasting*, 13, 281-291.

Hendry, D. F. (1995), *Dynamic Econometrics*, Oxford: Oxford University Press.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999), "Bayesian Model Averaging: a Tutorial," *Statistical Science*, 14, 382-417.

Huang, L. S. (1997), "Testing Goodness-Of-Fit Based on a Roughness Measure," *Journal of the American Statistical Association*, 92, 1399-1402.

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, London: Chapman Hall.

Jondeau, E., and Rockinger, M. (2006), "The Copula-GARCH Model of Conditional Dependencies: An International Stock Market Application," *Journal of International Money and Finance*, 25, 827-853.

Jones, M. (1993), "Simple Boundary Correction for Kernel Density Estimation," *Statistical Computing*, 3, 135-146.

Jones, M., and Foster, P. (1996), "A Simple Nonnegative Boundary Correction Method for Kernel Density Estimation," *Statistica Sinica*, 6, 1005-1013.

Kallenberg, W. C. M., and Ledwina, T. (1995), "Consistency and Monte-Carlo Simulation of a Data Driven Version of Smooth Goodness-Of-Fit Tests," *Annals of Statistics*, 23, 1594-1608.

Kim, Y. (2005), "Dependence Structure in International Financial Markets: Evidence from Asian Stock Markets," Unpublished manuscript, Department of Economics, University of California at San Deigo.

Kling, J. L., and Bessler, D. A. (1985), "A Comparison of Multivariate Forecasting Procedures for Economic Time Series," *International Journal of Forecasting*, 1, 5-24.

Koo, J. Y. (1996), "Bivariate B-splines for Tensor Logspline Density Estimation," *Computational Statistics and Data Analysis*, 21, 31–42.

Kooperberg, C. (1998), "Bivariate Density Estimation with An Application to Survival Analysis," *Journal of Computational and Graphical Statistics*, 7, 322-341.

Ledwina, T. (1994), "Data-Driven Version of Neyman's Smooth Test," *Journal of the American Statistical Association*, 89, 1000-1005.

Lee, T. H., and Long, X. (2006), "Copula-Based Multivariate GARCH Model with Uncorrelated Dependent Standardized Returns," Working paper, Department of Economics, University of California at Riverside.

Lejeune, M., and Sarda, P. (1992), "Smooth Estimators of Distribution and Density Functions," *Computational Statistics and Data Analysis*, 14, 457-471.

Li, Q., and Racine, J. (2007), *Nonparametric Econometrics: Theory and Practice*, Ewing, NJ: Princeton University Press.

Liebscher, E. (2005), "Semiparametric Density Estimators Using Copulas," *Communications in Statistics A*, 34, 59-71.

Longin, F., and Solnik, B. (2001), "Extreme Correlation of International Equity Markets," *Journal of Finance*, 56, 69-676.

Malevergne, Y., and Sornette, D. (2006), *Extreme Financial Risks: From Dependence to Risk Management*, New York: Springer-Verlag.

Mandelbrot, B. (1963), "New Methods in Statistical Economies," *Journal of Political Economy*, 71, 421-440.

Marron, J., and Ruppert, P. (1994), "Transformations to Reduce Boundary Bias in Kernel Density Estimation," *Journal of the Royal Statistical Society, Series B*, 56, 653-671.

Marron, J., and Wand, P. (1992), "Exact Mean Integrated Squared Error," *The Annals of Statistics*, 20, 712-736.

Miller, D., and Liu, W. H. (2002), "On the Recovery of Joint Distributions from Limited Information," *Journal of Econometrics*, 107, 259-274.

Montgomery, A. L., Zarnowitz, V., Tsay, R. S., and Tiao, G. C. (1998), "Forecasting the U.S. Unemployment Rate," *Journal of the American Statistical Association*, 93, 478-493.

Muller, H. (1991), "Smooth Optimum Kernel Estimators Near Endpoints," *Biometrika*, 78, 521-530.

Nelsen, R. B. (2006), *An Introduction to Copulas*, 2nd edition, New York: Springer-Verlag.

Newbold, P., and Granger, C. W. J. (1974), "Experience with Forecasting Univariate Time Series and the Combination of Forecasts," *Journal of the Royal Statistical Society, Series A*, 137, 131-146.

Neyman, J. (1937), "Smooth Test for Goodness of Fit," *Skandinavisk Aktuarietidskrift*, 20, 150-199.

Oakes, D. (1986), "Semiparametric Inference in a Model for Association in Bivariate Survival Data," *Biometrika*, 73, 353-361.

Okimoto, T. (2006), "New Evidence of Asymmetric Dependence Structure in International Equity Markets: Further Asymmetry in Bear Markets," Working paper, Department of Economics, Yokohama National University, Yokohama, Japan.

Ornermite, D., and White, H. (1999), "An Efficient Algorithm to Compute Maximum Entropy Densities," *Econometric Reviews*, 18, 141-67.

Patton, A. J. (2004), "On the Out-of-Sample Importance of Skewness and Asymmetric Dependence for Asset Allocation," *Journal of Financial Econometrics*, 2, 130-168.

——— (2005), "Estimation of Copula Models for Time Series of Possibly Different Lengths," *Journal of Applied Econometrics*, 21, 147-173.

——— (2006), "Modeling Asymmetric Exchange Rate Dependence," *International Economic Review*, 47, 527-556.

——— (2007), "Copula-Based Models for Financial Time Series," in *Handbook of Financial Time Series*, eds. T.G. Andersen, R.A. Davis, J. P. Kreiss and T. Mikosch, New York: Springer Verlag, forthcoming.

Portnoy, S. (1988), "Asymptotic Behavior of Likelihood Methods for Exponential Families when the Number of Parameters Tends to Infinity," *The Annals of Statistics*, 16, 356-366.

Rayner, J. C. W., and Best, D. J. (1989), *Smooth Tests of Goodness of Fit*, Oxford: Oxford University Press.

Rivers, D., and Vuong, Q. (2002), Model Selection Tests for Nonlinear Dynamic Models, *The Econometrics Journal*, 5, 1-39.

Rodriguez, J. C. (2007), “Measuring Financial Contagion: A Copula Approach,” *Journal of Empirical Finance*, 14, 401-423.

Rosenblatt, M. (1956), “Remarks on Some Non-Parametric Estimates of a Density Function,” *Annals of Mathematical Statistics*, 27, 832-837.

Sancetta, A., and Satchell, S. (2004), “The Bernstein Copula and Its Applications to Modeling and Approximations of Multivariate Distributions,” *Econometric Theory*, 20, 535-562.

Schwarz, G. (1978), “Estimating the Dimension of a Model,” *Annals of Statistics*, 6, 461-464.

Shannon, C. E. (1949), *Mathematical Theory of Communication*, Urbana: University of Illinois Press.

Sklar, A. (1959), “Fonctions De Repartition a n Dimensionset Leurs Mrges,” *Publications de l’ Institute de Statistique de l’Universite de Paris*, 8, 229-231.

Stephens, M. A. (1974), “EDF Statistics for Goodness of Fit and Some Comparisons,” *Journal of the American Statistical Association*, 69, 730-737.

Stone, C. (1982), “Optimal Global Rates of Convergence for Nonparametric Regression,” *The Annals of Statistics*, 10, 1040-1053.

———— (1990), “Large-sample Inference for Log-spline Models,” *The Annals of Statistics*, 18, 717-741.

——— (1994), "The Use of Polynomial Splines and Their Tensor Products in Multivariate Function Estimation," *The Annals of Statistics*, 22, 118-184.

Tawn, J. (1988), "Bivariate Extreme Value Theory: Models and Estimation," *Biometrika*, 75, 397-415.

Vuong, Q. (1989), "Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses," *Econometrica*, 57, 307-333.

White, H. (1982), "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1-26.

Winkler, R. L., and Makridakis, S. (1983), "The Combination of Forecasts," *Journal of the Royal Statistical Society, Series A*, 146, 150-157.

Wu, X. (2003), "Calculation of Maximum Entropy Densities with Application to Income Distribution," *Journal of Econometrics*, 115, 347-354.

——— (2007), "Exponential Series Estimator of Multivariate Density," Working paper, Department of Agricultural Economics, Texas A&M University.

Zellner, A., and Highfield, R.A. (1988), "Calculation of Maximum Entropy Distribution and Approximation of Marginal Posterior Distributions," *Journal of Econometrics*, 37, 195-209.

APPENDIX A

Table 2.1 Parameter Values of Copulas Corresponding to Kendall's Tau

Kendall's Tau	Gaussian	*t*	Frank	Clayton
0.2	0.309	0.309	1.86	0.50
0.4	0.588	0.588	4.16	1.33
0.6	0.809	0.809	7.93	3.00

Table 2.2 MISE of Univariate Distribution (the Values Are Multiplied by 100)

		Gaussian		Skewed-Unimodal		Bimodal		*t*	
n=50	ESE	0.637	(1.018)	1.494	(1.961)	2.337	(1.265)	0.690	(0.621)
	Log-spline	0.772	(0.749)	1.238	(1.514)	1.509	(3.819)	0.756	(0.494)
	Kernel	1.542	(3.776)	1.982	(4.352)	1.804	(3.866)	1.382	(2.210)
n=100	ESE	0.284	(0.131)	0.751	(0.517)	1.328	(0.823)	0.381	(0.122)
	Log-spline	0.359	(0.107)	0.594	(0.324)	0.860	(0.362)	0.387	(0.123)
	Kernel	0.715	(0.394)	1.015	(0.885)	1.047	(0.652)	0.815	(0.670)
n=500	ESE	0.056	(0.003)	0.186	(0.017)	0.437	(0.019)	0.161	(0.011)
	Log-spline	0.111	(0.010)	0.138	(0.014)	0.212	(0.016)	0.089	(0.007)
	Kernel	0.231	(0.034)	0.295	(0.063)	0.305	(0.039)	0.218	(0.033)

Table 2.3 MISE of Copula Density Estimation

Kendall's Tau	Copula	*n*	ESE		Log-spline		Kernel	
0.2	Gaussian	50	0.164	(0.0071)	0.170	(0.0112)	0.233	(0.0135)
		100	0.107	(0.0014)	0.120	(0.0041)	0.171	(0.0024)
		500	0.065	(0.0001)	0.076	(0.0002)	0.099	(0.0002)
	t	50	0.210	(0.0194)	0.244	(0.0420)	0.270	(0.0165)
		100	0.139	(0.0014)	0.160	(0.0051)	0.201	(0.0032)
		500	0.098	(0.0001)	0.104	(0.0005)	0.127	(0.0006)
	Frank	50	0.172	(0.0117)	0.168	(0.0161)	0.221	(0.0114)
		100	0.103	(0.0020)	0.105	(0.0019)	0.170	(0.0046)
		500	0.058	(0.0001)	0.069	(0.0001)	0.104	(0.0001)
	Clayton	50	0.217	(0.0061)	0.236	(0.0194)	0.274	(0.0115)
		100	0.160	(0.0025)	0.188	(0.0189)	0.214	(0.0052)
		500	0.113	(0.0001)	0.118	(0.0008)	0.146	(0.0095)
0.4	Gaussian	50	0.240	(0.0080)	0.350	(0.0397)	0.381	(0.0223)
		100	0.180	(0.0024)	0.270	(0.0150)	0.293	(0.0070)
		500	0.131	(0.0001)	0.174	(0.0027)	0.199	(0.0059)
	t	50	0.345	(0.0139)	0.455	(0.0829)	0.480	(0.0415)
		100	0.263	(0.0024)	0.349	(0.0200)	0.371	(0.0059)
		500	0.217	(0.0005)	0.251	(0.0047)	0.288	(0.0007)
	Frank	50	0.215	(0.0164)	0.292	(0.0267)	0.329	(0.0187)
		100	0.142	(0.0031)	0.215	(0.0101)	0.235	(0.0048)
		500	0.090	(0.0002)	0.125	(0.0009)	0.157	(0.0009)
	Clayton	50	0.574	(0.0190)	0.664	(0.1138)	0.683	(0.0369)
		100	0.498	(0.0039)	0.555	(0.0669)	0.577	(0.0509)
		500	0.364	(0.0012)	0.410	(0.0130)	0.453	(0.0154)
0.6	Gaussian	50	0.484	(0.0155)	0.780	(0.0489)	0.805	(0.0545)
		100	0.401	(0.0047)	0.654	(0.0087)	0.645	(0.0166)
		500	0.328	(0.0014)	0.518	(0.0012)	0.532	(0.0005)
	t	50	0.721	(0.0315)	1.014	(0.0897)	1.072	(0.0812)
		100	0.629	(0.0087)	0.859	(0.0230)	0.865	(0.0209)
		500	0.451	(0.0039)	0.692	(0.0087)	0.721	(0.0070)
	Frank	50	0.302	(0.0202)	0.551	(0.0547)	0.585	(0.0648)
		100	0.212	(0.0051)	0.400	(0.0084)	0.408	(0.0123)
		500	0.142	(0.0003)	0.237	(0.0034)	0.281	(0.0004)
	Clayton	50	1.632	(0.0331)	1.917	(0.0872)	2.004	(0.0955)
		100	1.493	(0.0169)	1.695	(0.0543)	1.750	(0.0619)
		500	1.105	(0.0043)	1.409	(0.0040)	1.619	(0.0075)

Table 2.4 MISE of the Joint Density Estimation

				Copula			
	Kendall's Tau			Gaussian	*t*	Frank	Clayton
Marginal Distribution	0.2	Gaussian	MM	0.377	0.364	0.300	0.365
			LM	0.464	0.518	0.544	0.635
			M	0.389	0.403	0.359	1.075
			L	0.556	0.726	0.603	1.437
			K	1.211	1.136	1.156	2.070
		Skewed Unimodal	MM	0.993	0.983	0.890	1.029
			LM	0.513	0.468	0.481	0.483
			M	1.059	1.021	1.157	1.247
			L	0.891	1.003	1.009	1.486
			K	1.593	1.815	1.713	2.415
		Bimodal	MM	1.040	1.055	0.960	1.334
			LM	0.435	0.473	0.457	1.153
			M	1.409	1.421	1.327	2.610
			L	0.850	0.925	0.871	1.204
			K	1.083	1.218	1.099	2.419
		T	MM	0.362	0.291	0.332	0.325
			LM	0.268	0.244	0.245	0.278
			M	0.392	0.429	0.387	0.388
			L	0.465	0.599	0.574	0.567
			K	1.030	1.290	1.100	1.424
	0.4	Gaussian	MM	0.616	0.752	0.637	0.972
			LM	0.836	0.784	0.704	1.070
			M	0.415	0.467	1.174	1.535
			L	0.823	0.798	1.297	1.739
			K	1.492	1.517	1.522	2.659
		Skewed Unimodal	MM	1.269	1.175	1.152	1.348
			LM	1.028	1.090	1.071	1.213
			M	1.677	1.469	1.578	1.520
			L	1.086	1.173	1.335	1.844
			K	1.821	1.957	2.108	2.865
		Bimodal	MM	1.239	1.358	1.152	1.657
			LM	1.081	1.066	1.034	1.533
			M	1.648	1.656	1.686	2.844
			L	1.140	1.149	1.089	1.607
			K	1.448	1.432	1.594	2.893
		T	MM	0.626	0.704	0.563	0.807
			LM	0.626	0.712	0.571	0.872
			M	0.972	1.064	1.229	1.116
			L	0.703	0.900	0.866	0.846
			K	1.342	1.410	1.612	1.828

Table 2.4 (Continued)

			Copula			
Kendall's Tau			Gaussian	*t*	Frank	Clayton
0.6	Gaussian	MM	1.202	1.126	1.050	2.409
		LM	1.201	1.302	1.080	2.662
		M	0.563	0.694	2.039	2.088
		L	1.351	1.449	2.359	2.743
		K	2.700	2.520	3.057	3.373
	Skewed Unimodal	MM	1.832	1.874	1.924	2.252
		LM	1.786	1.880	1.428	2.205
		M	2.056	2.288	2.539	2.236
		L	1.849	2.087	3.643	2.462
		K	2.818	3.219	4.185	3.718
	Bimodal	MM	2.002	2.169	1.784	3.465
		LM	1.703	1.901	1.569	3.183
		M	2.343	2.584	2.161	3.566
		L	1.558	1.981	1.638	3.319
		K	2.044	2.592	2.228	3.687
	T	MM	1.341	1.287	1.588	2.217
		LM	0.841	1.080	0.906	1.934
		M	1.520	1.645	1.644	2.361
		L	1.168	1.344	1.426	2.251
		K	2.094	2.367	2.782	3.022

Table 2.5 Average MISE for Different Copula and Kendall's Tau

Tau		Copula							
		Gaussian		*t*		Frank		Clayton	
		MISE	Proportion	MISE	Proportion	MISE	Proportion	MISE	Proportion
0.2	LM	0.420		0.426		0.432		0.637	
	M	0.812	52%	0.819	52%	0.808	53%	1.330	48%
	L	0.691	61%	0.813	52%	0.764	57%	1.173	54%
	K	1.229	34%	1.365	31%	1.267	34%	2.082	31%
0.4	LM	0.892		0.913		0.845		1.172	
	M	1.178	76%	1.164	77%	1.417	63%	1.754	51%
	L	0.938	95%	1.005	89%	1.147	78%	1.509	59%
	K	1.526	58%	1.579	57%	1.709	52%	2.561	35%
0.6	LM	1.383		1.541		1.246		2.496	
	M	1.621	85%	1.803	77%	2.096	66%	2.563	54%
	L	1.481	93%	1.715	81%	2.266	61%	2.694	51%
	K	2.414	57%	2.674	52%	3.063	45%	3.450	40%

NOTE:
1. Proportion means the ratio of MISE of LM to the ratio of MISE of the other direct estimation methods.
2. MISE is the average across the marginal distributions.

Table 2.6 Descriptive Statistics for the Stock Indices Returns

	Hongkong	Singapore	Malaysia	Philippines	Taiwan	Thailand
Mean	-0.0131	-0.0329	-0.1127	-0.0574	-0.0103	-0.1118
Median	0.0046	-0.0139	-0.0280	0.0000	0.0000	-0.0533
Maximum	17.2702	15.3587	23.2840	13.3087	6.3538	16.3520
Minimum	-14.7132	-10.6621	-37.0310	-11.0001	-11.3456	-15.8925
Std. Dev.	1.9668	1.6484	2.7410	1.9546	1.5864	2.5656
Skewness	0.2615	0.6875	-1.2933	-0.0172	-0.4601	0.5974
Kurtosis	13.8503	16.1685	42.0452	8.1313	7.2913	9.4746
Jarque-Bera	6411.5	9524.6	83196.3	1430.7	1046.6	2355.2
P-value	<1.00e-04	<1.00e-04	<1.00e-04	<1.00e-04	<1.00e-04	<1.00e-04

Table 2.7 Correlation Matrix of the Stock Indices Returns

	Hongkong	Singapore	Malaysia	Philippines	Taiwan	Thailand
Hongkong		0.6526	0.3797	0.3816	0.2488	0.3867
Singapore	0.3597		0.4104	0.2581	0.2797	0.5126
Malaysia	0.2816	0.4783		0.3020	0.1993	0.3862
Philippines	0.1954	0.5080	0.2069		0.1999	0.4133
Taiwan	0.1214	0.1312	0.0992	0.0979		0.1886
Thailand	0.2205	0.2900	0.2726	0.2069	0.0669	
Average Dependence						
Linear	0.4099	0.4863	0.3491	0.3609	0.2233	0.3775
Kendall	0.2357	0.2899	0.2541	0.1930	0.1033	0.2114

NOTE: Upper triangle is the linear correlation and the lower triangle is the Kendall's Tau.

Table 2.8 Empirical Tail Dependence for Bivariate Standardized Returns

	Hong Kong		Singapore		Malaysia		Philippines		Taiwan	
Percentile	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
5%	0.308	0.242	0.354	0.167	0.262	0.242	0.231	0.182	0.169	0.106
4%	0.288	0.283	0.346	0.132	0.269	0.151	0.250	0.170	0.154	0.113
3%	0.282	0.225	0.359	0.125	0.179	0.100	0.179	0.150	0.128	0.150
2%	0.346	0.185	0.346	0.148	0.077	0.000	0.192	0.074	0.115	0.111
1%	0.231	0.143	0.231	0.214	0.000	0.000	0.000	0.000	0.000	0.143

NOTE:

1. Upper = empirical conditional upper tail probability.
2. Lower = empirical conditional lower tail probability.

Table 2.9 Estimated Tail Dependence for Bivariate Standardized Returns

	Hong Kong		Singapore		Malaysia		Philippines		Taiwan	
Percentile	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
5%	0.314	0.257	0.296	0.209	0.205	0.198	0.181	0.187	0.143	0.145
4%	0.259	0.211	0.237	0.167	0.162	0.156	0.159	0.162	0.121	0.119
3%	0.198	0.167	0.168	0.123	0.134	0.129	0.126	0.121	0.097	0.092
2%	0.135	0.121	0.119	0.087	0.089	0.081	0.087	0.089	0.078	0.076
1%	0.075	0.061	0.062	0.048	0.065	0.057	0.056	0.058	0.046	0.048

NOTE:

1. Lower = Estimated upper tail dependence.
2. Upper = Estimated lower tail dependence.

Table 3.1 5% Critical Values of Test Statistics for Subsequent Values of K (Uniform Weighted)

n=20		1	2	3	4	5	6	7	8	9	10
	LR	3.721	5.287	6.428	7.156	7.548	7.687				
	Neyman	3.705	5.367	6.537	7.213	7.617	7.854				

n=50		1	2	3	4	5	6	7	8	9	10
	LR	3.817	5.460	6.114	6.121	6.123	6.405	6.406	6.411	6.411	6.411
	Neyman	3.788	5.317	5.869	6.016	6.090	6.168	6.172	6.182	6.182	6.182

n=100		1	2	3	4	5	6	7	8	9	10
	LR	3.827	5.213	5.347	5.420	5.420	5.420	5.420	5.420	5.420	5.420
	Neyman	3.812	5.047	5.231	5.367	5.341	5.342	5.342	5.342	5.342	5.342

NOTE:
1. LR statistic is calculated from equation (3.4).
2. Neyman statistic is calculated from equation (3.2).

Table 3.2 Monte Carlo Power Estimate x 100, Tests of Size 0.05 with Sample Size Equal to 20

	(1)		(2)		(3)		(4)		(5)
n=20	Geometric Weighted Neyman	Geometric Weighted LR	Uniform Weighted Neyman	Uniform Weighted LR	BIC Weighted Neyman	BIC Weighted LR	Likelihood CV Neyman	Likelihood CV LR	Anderson-Darling
g1	30	39	44	49	43	44	43	43	45
g2_1	38	38	16	18	37	37	34	37	42
g2_2	20	22	27	30	31	33	25	26	17
g2_3	15	18	22	25	23	26	23	23	10
g2_4	29	31	32	32	33	35	31	32	8
g3_1	41	43	42	46	41	43	35	39	45
g3_2	39	43	45	50	39	41	34	38	25
g3_3	26	31	43	49	41	45	31	37	15
g3_4	39	45	48	52	38	41	39	40	30
g3_5	36	41	43	47	37	39	36	37	10
g3_6	39	43	47	51	39	42	36	39	20
g3_7	16	22	38	41	37	40	20	24	9
g3_8	12	16	29	31	26	28	14	17	12

NOTE:
1. Anderson-Darling statistic is defined in equation (3.1).
2. Geometric weighting is defined in equation (3.5).
3. BIC weighting is defined in equation (3.6).
4. Likelihood CV is defined in equation (3.8).

Table 3.3 Monte Carlo Power Estimate x 100, Tests of Size 0.05 with Sample Size Equal to 50

	(1)		(2)		(3)		(4)		(5)
n=50	Geometric Weighted Neyman	Geometric Weighted LR	Uniform Weighted Neyman	Uniform Weighted LR	BIC Weighted Neyman	BIC Weighted LR	Likelihood CV Neyman	Likelihood CV LR	Anderson-Darling
g1	79	78	84	85	86	87	85	86	68
g2_1	75	76	73	73	73	74	73	74	70
g2_2	51	52	55	56	62	61	41	40	35
g2_3	40	39	46	47	46	47	31	32	18
g2_4	62	64	65	65	65	66	62	64	15
g3_1	81	81	80	80	81	80	81	81	79
g3_2	84	85	86	84	85	86	84	85	50
g3_3	60	61	60	58	62	61	83	83	30
g3_4	88	89	90	90	90	91	84	85	60
g3_5	77	75	73	71	74	75	71	70	18
g3_6	78	81	75	75	76	76	72	70	43
g3_7	38	39	49	49	51	50	40	41	20
g3_8	29	28	44	44	47	46	28	30	25

NOTE:

1. Anderson-Darling statistic is defined in equation (3.1).
2. Geometric weighting is defined in equation (3.5).
3. BIC weighting is defined in equation (3.6).
4. Likelihood CV is defined in equation (3.8).

Table 3.4 Likelihood Ratio Test Based on the Nested Sequence and All Subset Specifications

	n=20		*n*=50	
	Uniform Weighted Likelihood Ratio	All Subset Likelihood Ratio	Uniform Weighted Likelihood Ratio	All Subset Likelihood Ratio
g1	28	20	39	29
g2_1	25	18	39	27
g2_2	21	19	37	31
g2_3	17	14	38	30
g2_4	19	15	43	40
g3_1	20	17	41	28
g3_2	19	16	44	37
g3_3	14	12	31	24
g3_4	28	24	56	48
g3_5	19	15	40	35
g3_6	20	16	48	41
g3_7	15	11	33	35
g3_8	13	9	26	15

NOTE: All subsets specification is defined in equation (3.7).

Table 4.1 Descriptive Statistics for the Data

	S&P500	Fund rate	CRB
Mean	0.0549	5.9756	5.5748
Median	0.0307	5.5200	5.5526
Maximum	13.8051	22.3600	6.0633
Minimum	-36.3001	0.8600	5.3260
Std. Dev.	1.6159	3.1822	0.1386
Skewness	-1.7835	1.0879	1.0791
Kurtosis	44.3913	5.3916	4.7264
Jarque-Bera	499020.3	3022.527	2209.015
p-value	<1.00e-04	<1.00e-04	<1.00e-04

Table 4.2 Correlation Matrix of the Stock Indices Returns

	S&P500	Federal Fund	CRB
S&P500		-0.0025	0.0019
Federal Fund	-0.0037		-0.0298
CRB	0.0024	-0.0373	

NOTE: Lower triangle is the linear correlation and the upper triangle is the Kendall's Tau.

Table 4.3 The BIC and Log Likelihood of the Four Copulas

Model	Log Likelihood	Number of Parameters	BIC
Symmetric copulas			
Normal	7.7853	4	-0.20516
Student's *t*	8.2347	5	2.737397
Asymmetric copulas			
Clayton	10.4398	1	-17.0386
Gumbel	5.3803	1	-6.91955

NOTE:
1. Normal copula is defined in equation (4.4).
2. Student's *t* copula is defined in equation (4.5).
3. Clayton copula is defined in equation (4.6).
4. Gumbel copula is defined in equation (4.7).

Table 4.4 Mean Squared Forecast Errors by Horizon

	One Step	Two Steps	Three Steps	Four Steps
ARIMA	0.343	0.421	0.587	0.479
VAR	0.351	0.409	0.508	0.515
Copula-VAR	0.287	0.381	0.498	0.457

Table 4.5 Test Results for the Equality of Forecast MSFE

	One Step	Two Steps	Three Steps	Four Steps
ARIMA vs VAR	0.21	0.38	0.37	0.41
VAR vs Copula-VAR	0.08	0.13	0.16	0.19
Copula-VAR vs ARIMA	0.09	0.14	0.12	0.24

NOTE: Each entry is the *p*-value of modified Diebold-Mariano test (equation 4.9) on the quality of mean squared forecast errors (MSFE).

Table 4.6 Forecast Encompassing Tests

	One Step		
	Copula-VAR	VAR	ARIMA
Copula-VAR	1	0.31	0.21
VAR	0.04	1	0.31
ARIMA	0.07	0.23	1
	Two Steps		
	Copula-VAR	VAR	ARIMA
Copula-VAR	1	0.26	0.27
VAR	0.12	1	0.38
ARIMA	0.09	0.41	1
	Three Steps		
	Copula-VAR	VAR	ARIMA
Copula-VAR	1	0.48	0.34
VAR	0.14	1	0.27
ARIMA	0.16	0.21	1
	Four Steps		
	Copula-VAR	VAR	ARIMA
Copula-VAR	1	0.28	0.26
VAR	0.12	1	0.34
ARIMA	015	0.25	1

NOTE:
Each entry is the *p*-value of the null hypothesis that a model (in a column) encompasses another model (in a row) from Equation 4.10. For example, the first entry in the second row is 0.08, and VAR model encompasses the copula-VAR model at 0.05 signficance level. All diagonal entries have values 1 for obvious reasons.

Table 4.7 Combined Forecasts for the One-Step Ahead Forecasts

Forecast	Sum of Squared Errors	Weights for			
		Constant	VAR	Copula-VAR	ARIMA
Original		--	1.00	--	--
VAR	10.53	--	--	1.00	--
Copula-VAR	8.61	--	--	--	1.00
ARIMA	10.29				
Combined Method A (Unconstrained, No Constant Term)					
All Three	5.51	0.00	0.32	0.48	0.20
VAR & Copula-VAR	7.20	0.00	0.32	0.68	0.00
Copula-VAR & ARIMA	6.80	0.00	0.00	0.73	0.27
ARIMA & VAR	7.95	0.00	0.53	0.00	0.47
Combined Method B (No Constant, Weight Sum to 1)					
All Three	5.84	0.00	0.34	0.45	0.21
VAR & Copula-VAR	7.51	0.00	0.29	0.71	0.00
Copula-VAR & ARIMA	6.23	0.00	0.00	0.69	0.31
ARIMA & VAR	8.09	0.00	0.48	0.00	0.52
Combined Method C (Unconstrained, With Constant)					
All Three	5.12	11.67	0.30	0.42	0.19
VAR & Copula-VAR	6.58	12.34	0.30	0.64	0.00
Copula-VAR & ARIMA	5.34	6.84	0.00	0.72	0.11
ARIMA & VAR	7.01	5.47	0.52	0.00	0.38

APPENDIX B

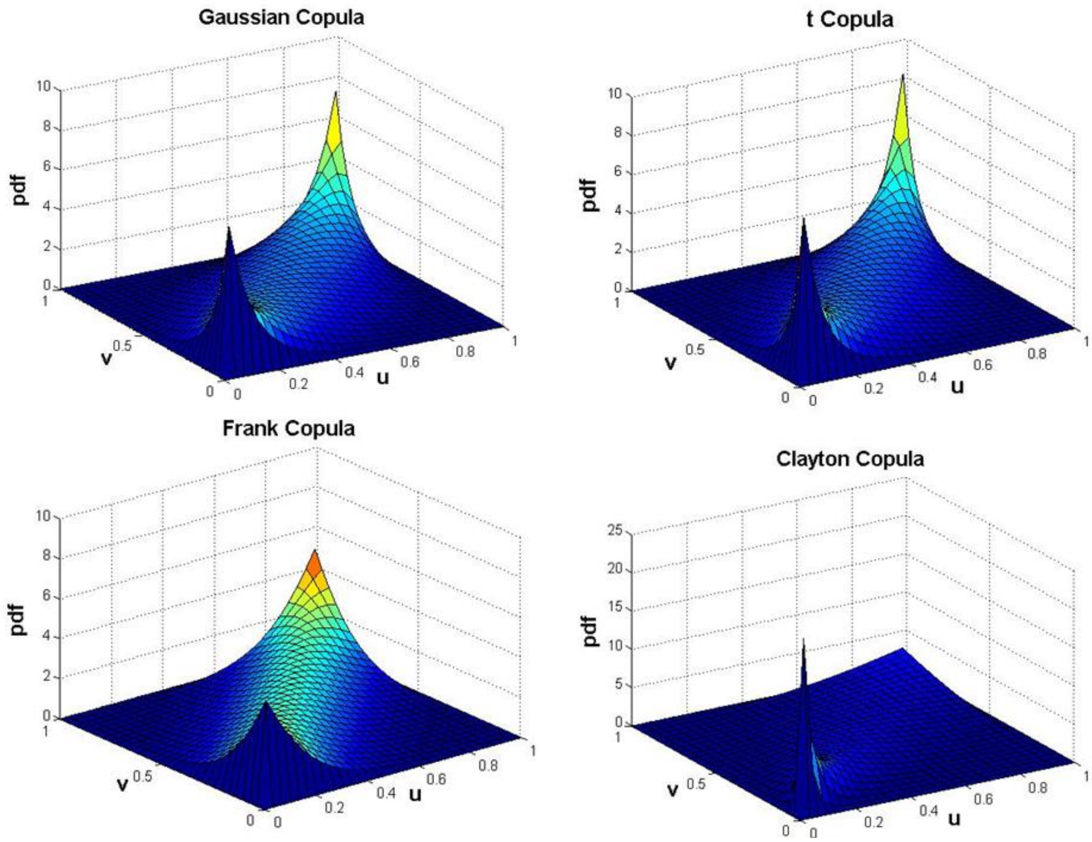

Figure 2.1 Parametric Copulas with Dependence Parameters Corresponding to Kendall's Tau Equal to 0.6

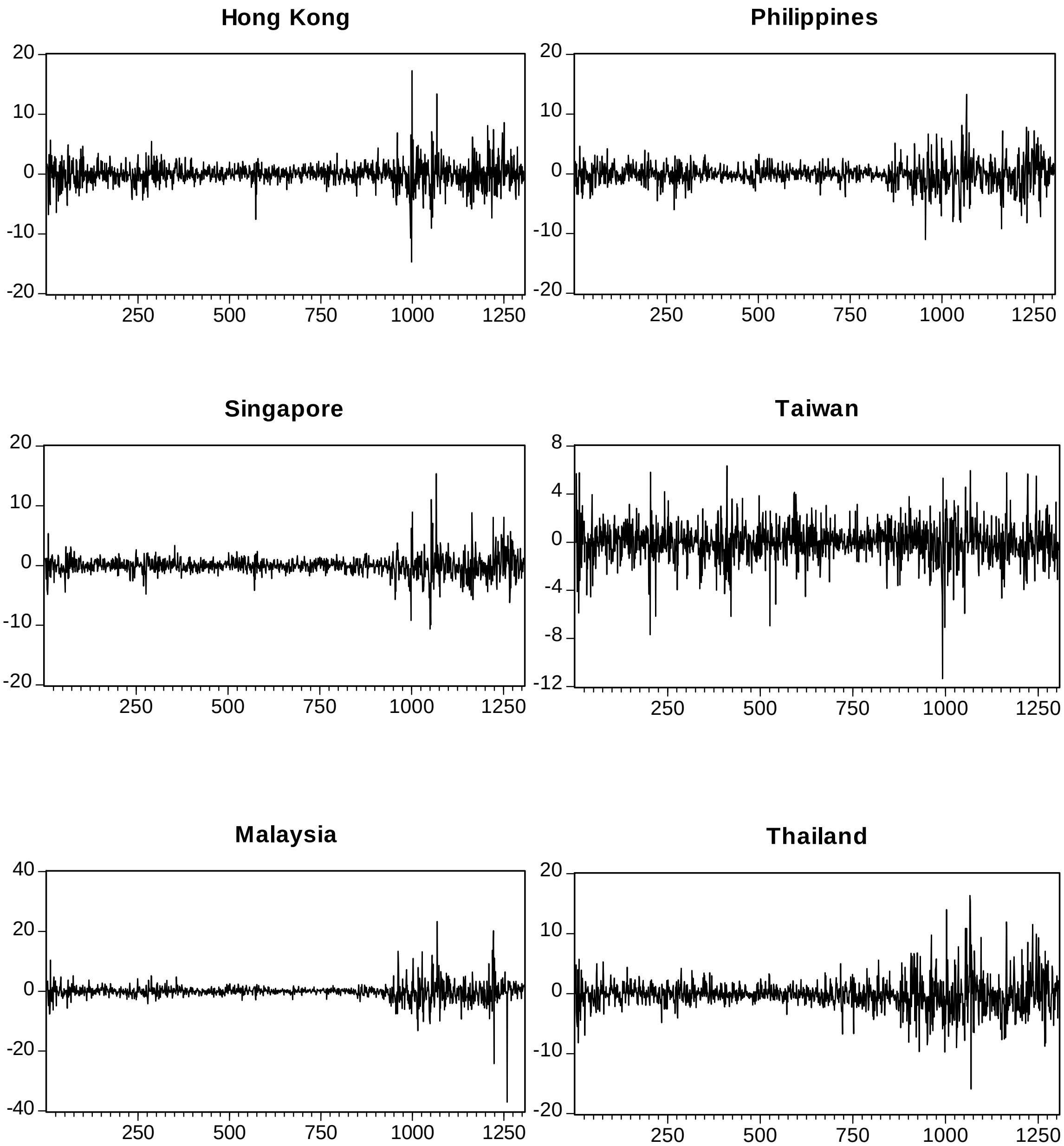

Figure 2.2 Asian Stock Indices Returns

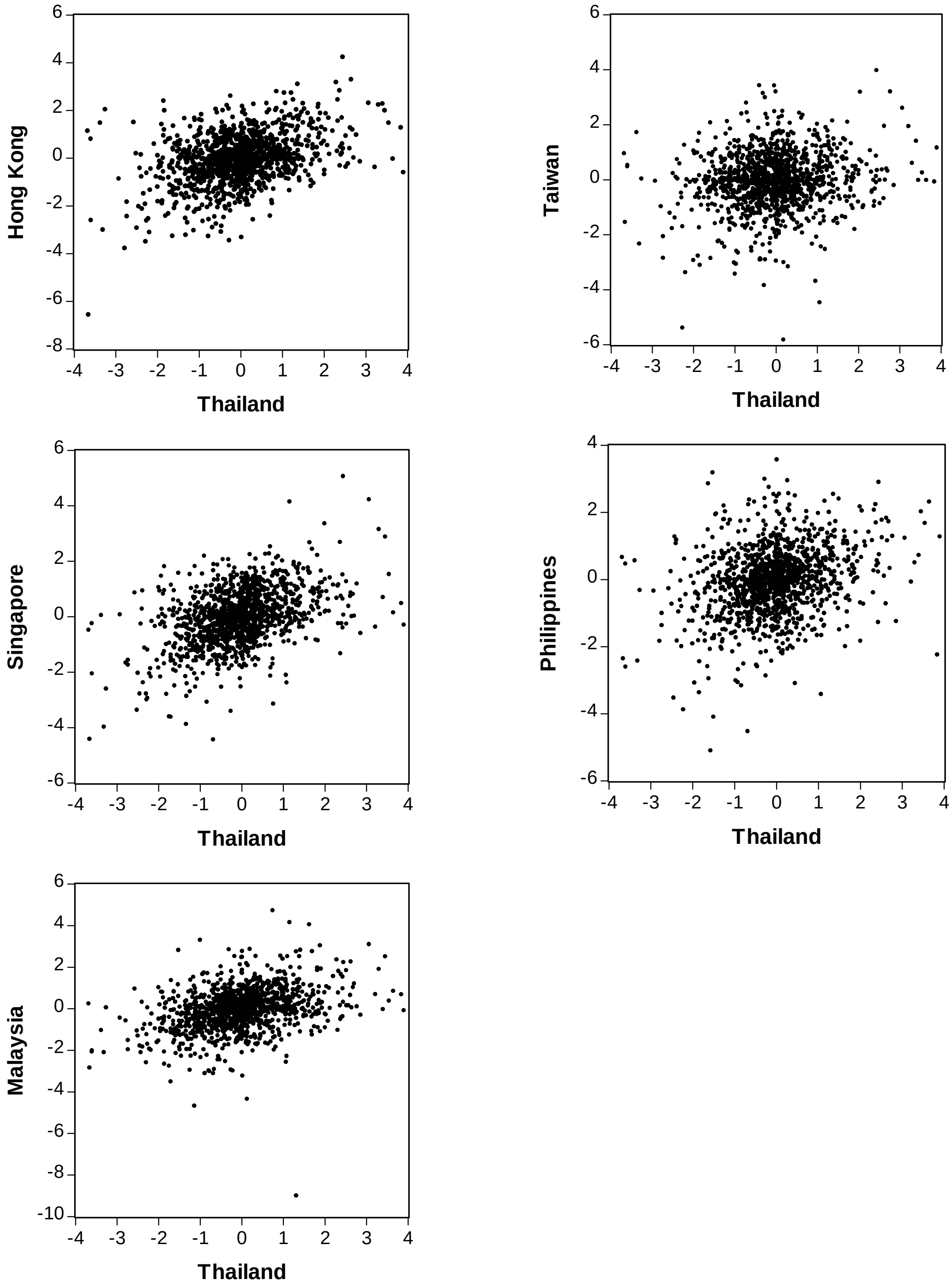

Figure 2.3 Scatter Plots of the Standardized Stock Indices Returns in Various Asian Countries

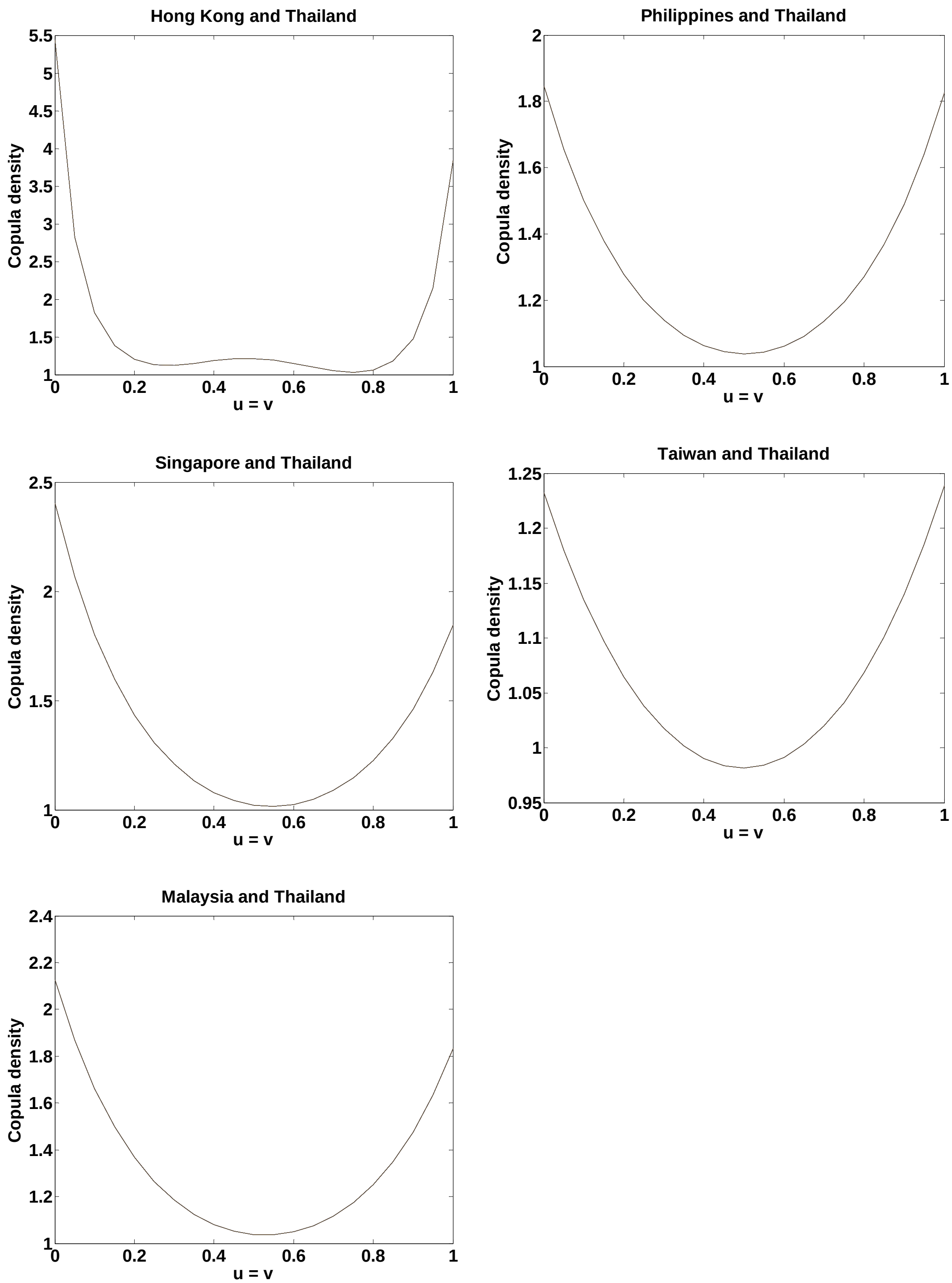

Figure 2.4 The Estimated Copula Density on the Diagonal for Various Asian Countries Pairing with Thailand

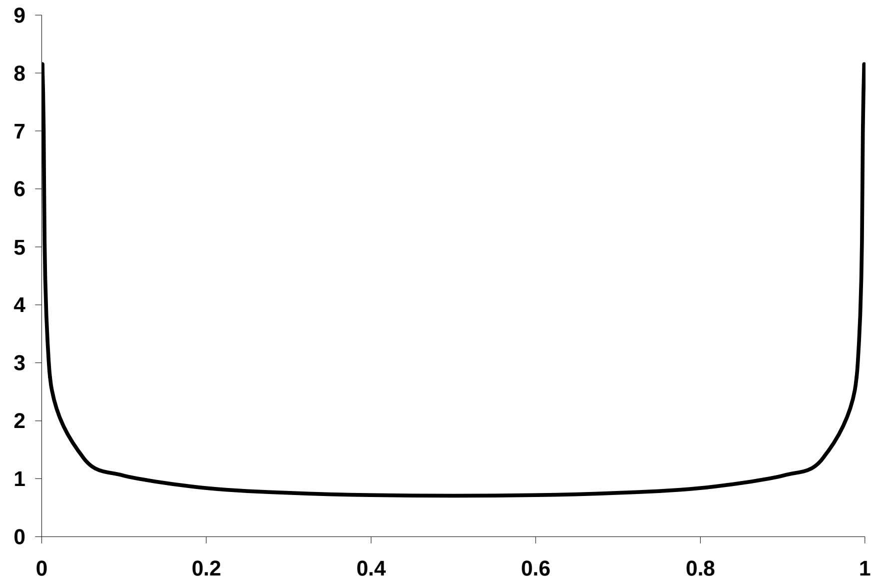

Figure 3.1 Graph of g_1

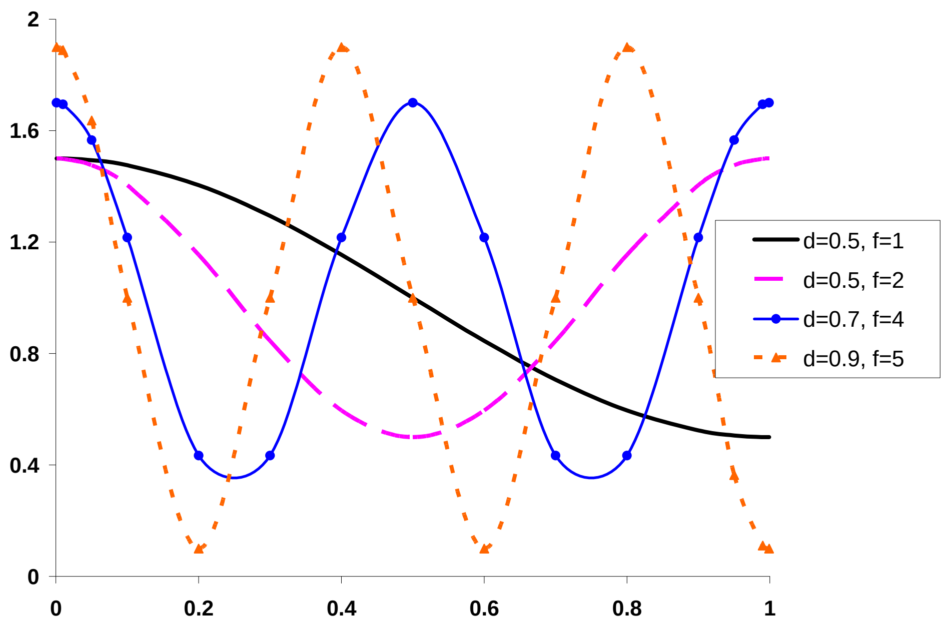

Figure 3.2 Graphs of g_2

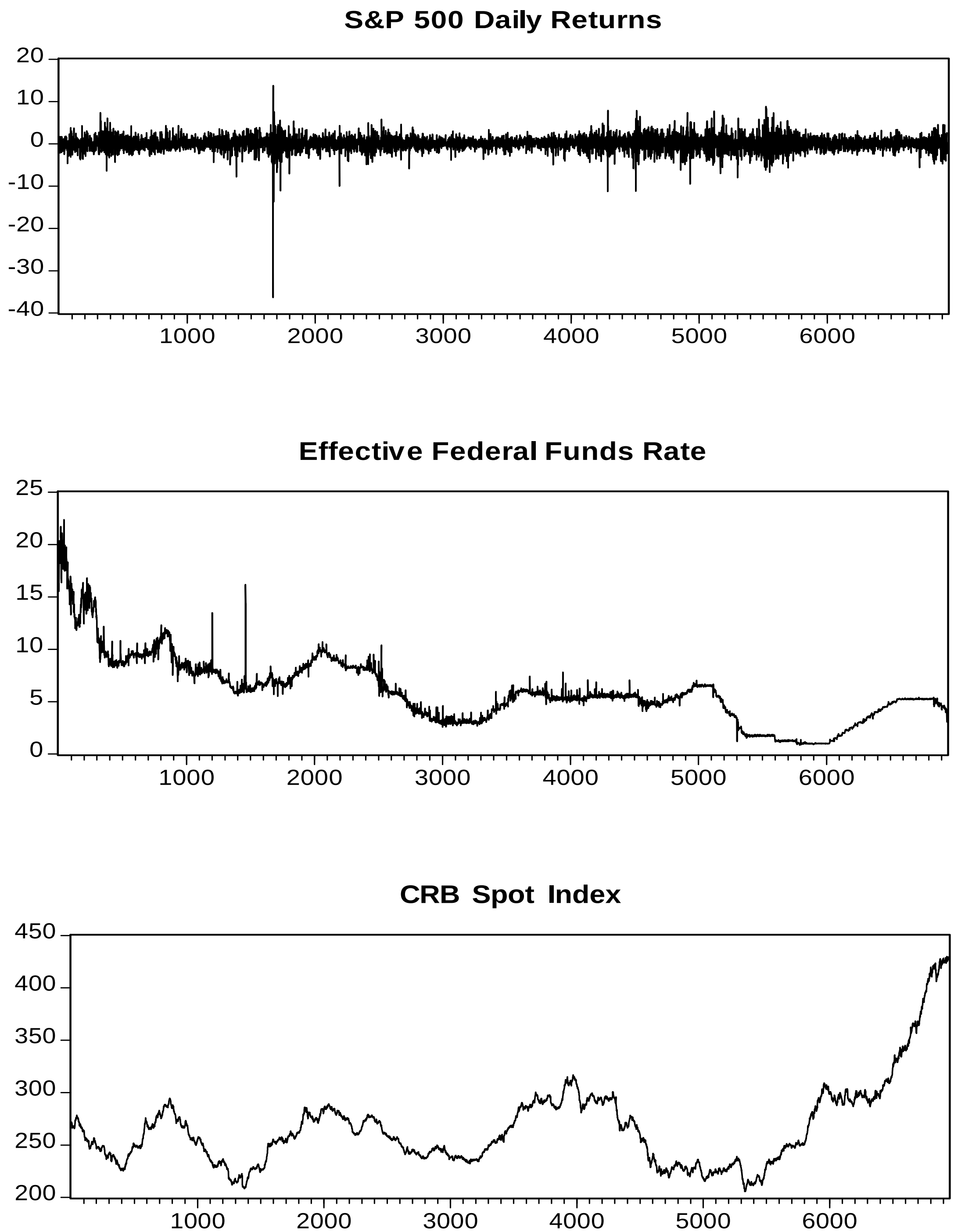

Figure 4.1 Data from S&P 500 Daily Returns, Effective Federal Funds Rate and CRB Spot Index: May 26, 1981 to December 31, 2007

VITA

Name:	Chin Man Chui
Address:	Room 2120, Tun Man House, Oi Man Estate, Kowloon, Hong Kong
Email Address:	chinman168@neo.tamu.edu
Education:	B.A., Quantitative Analysis for Business, City University of Hong Kong, 1993
	M.S., Statistics, University of Hong Kong, 1998
	M.S., Business Economics, Chinese University of Hong Kong, 2000