ESSAYS ON NONPARAMETRIC SERIES ESTIMATION WITH APPLICATION TO FINANCIAL ECONOMETRICS

A Dissertation

by

MENG-SHIUH CHANG

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 2011

Major Subject: Agricultural Economics

ESSAYS ON NONPARAMETRIC SERIES ESTIMATION WITH APPLICATION TO FINANCIAL ECONOMETRICS

A Dissertation

by

MENG-SHIUH CHANG

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
	Qi Li
	Victoria Salin
Head of Department,	John P. Nichols

August 2011

Major Subject: Agicultural Economics

ABSTRACT

Essays on Nonparametric Series Estimation with Application to Financial Econometrics. (August 2011)

Meng-Shiuh Chang, B.S., National ChiaoTung University;
M.B.A., Tamkang University;
M.S.C., University of York
Chair of Advisory Committee: Dr. Ximing Wu

This dissertation includes two essays. In the first essay, I proposed an alternative estimator for multivariate densities. This estimator can be characterized as a transformation based estimator. The first stage estimates each marginal density separately. In the second stage, the joint density of estimated marginal cumulative distribution functions (CDF) are approximated by the exponential series estimator. The final estimate is then obtained as the product of the marginal densities and the joint density estimated in the second stage. Extensive Monte Carlo studies show the proposed estimator outperforms kernel estimators in joint density and tail distribution estimation. An illustrative example on estimating the conditional copula density between S&P 500 and FTSE 100 given Hangseng and Nikkei 225 is also discussed.

In the second essay, I extended the semiparametric model by Chen and Fan [X. Chen, Y. Fan, Estimation of copula-based semiparametric time series models, Journal of Econometrics 130 (2006) 307–335], and studied a class of univariate copula-based nonparametric stationary Markov models in which the copulas and the marginal distributions are estimated nonparametrically. In particular, I focused on the stationary Markov process of order 1 with continuous state space because it has the β-mixing property for the analysis of weakly dependent processes. The copula density functions for time series models are approximated by the series estimate on sieve spaces. In this

study, a finite dimensional linear space spanned by a sequence of power functions is treated as the sieve space where the estimation space of the copula density function is based. This sieve series estimator can be characterized as the exponential series estimator under mild smoothness conditions. By using the β-mixing properties, I showed that the copula density function approximated by the exponential series estimator for stationary first-order Markov processes has the same convergence rate as the *i.i.d.* data. The Monte Carlo simulations show that the proposed estimator outperforms the kernel estimator in the conditional density estimation, except for the Frank copula-based Markov model. In addition, the proposed estimator considerably dominates the the kernel estimator when used in the one-step-ahead forecast.

ACKNOWLEDGMENTS

I would like to express my gratitude to my committee chair, Dr. Ximing Wu, for his support, patience, and encouragement of my Ph.D study and research since 2008. His guidance helped me in the research and writing of this dissertation. Without his guideline, this dissertation would not have been completed.

My thanks also go to Dr. David Bessler, Dr. Qi Li and Dr. Victoria Salin for reading previous drafts of this dissertation and providing valuable comments that improved the contents of this dissertation.

Last but not least, I would like to thank my family for their continuous support.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

There is an increasing interest in modeling the dependence structures of financial factors by methods of copula such as market relationships and value-at-risk. For example, estimating the value-at-risk of portfolios of assets has become a usual practice in risk management. A copula is a statistical tool for modeling the multivariate dependence structure among variables in a distribution free way. In fact, it needs to solve two tasks for the use of copulas. The first starts with the model of each marginal distribution. The second step involves estimating the copula density function. However, the development of estimating the copula functions is still in its infancy. There are two approaches commonly developed for estimating copulas in econometrics. The parametric approach makes assumptions of underlying copula function and uses the maximum likelihood estimator to estimate the unknown parameters of the assumed parametric copula function. However, the parametric estimator is inconsistent if the assumed copula function is misspecified. Alternatively, one can estimate the copula functions nonparametrically. For example, the estimation based on the Kernel density estimator, a well-developed estimator for densities, has become a routine while estimating a multivariate copula function. Another popular nonparametric estimator for copula functions is the series estimator which approximates the underlying copula function in terms of a sequence of basis functions. Because of the unconstrained functional form of true copula functions, the nonparametric copula estimators are consistent under mild conditions such as smoothness. However, these two nonpara-

---

This dissertation follows the style of Journal of Economic Theory.

metric estimators for copulas take the risk of negative copula density estimate and slower convergence rate.

In this dissertation, I apply exponential series methods to estimate the copula functions. Especially, I focus on the theoretical development of multivariate density estimator which includes the copula function that captures contemporaneous dependence among each variable. Besides, I study the estimation of a class of copula-based nonparametric stationary Markov models.

In the first essay I propose an alternative estimator for multivariate densities. This estimator can be characterized as a transformation based estimator. The first stage estimates each marginal density separately. In the second stage, the joint density of estimated marginal cumulative distribution functions (CDF) is approximated by the exponential series estimator (ESE). The final estimate is then obtained as the product of the marginal densities and the joint density estimated in the second stage. We derive the convergence rate in terms of the Kullback-Leibler Information Criterion (KLIC). A second contribution of this study is to incorporate a variable selection algorithm into a sequential updating process of moment selection to overcome the curse of dimensionality. As discussed in large literature, the curse of dimensionality occurs in many nonparametric methods when a high-dimensional sample space is involved. A typical way to tackle this problem is through a principal component analysis (PCA). However, dimensionality reduction via PCA still involves all of the moments which lead to a worse estimation performance in this essay. Instead of using PCA, the method used in this essay searches for subsets of moments that best approximate the full set of moments to reduce the dimensionality, e.q. McCabe [27]. This algorithm, called Principal Variables, can identify a subset of a set of original moments. This algorithm selects moments which are optimal for a given criterion that measures how well each subset approximates the whole set. In this study, I maximize the RM criterion over all possible subsets of moments for the purpose of reducing the dimensionality. The Monte Carlo studies show that the proposed estimator outperforms the kernel density estimator and the relative performance of our method with respect to the kernel method increases with the dimensionality of sample space. Besides, I also examine the performance of estimating tail distributions. My method dominates the empirical and the kernel density estimators except for the fat-tailed case. An empirical estimation of conditional copula density of stock returns is also provided.

In economic and financial applications, one is often interested in estimating certain features of the temporal dependence of the time series. For example, Robinson [33] applies multivariate Kernel probability density and regression estimators to a univariate strictly stationary time series. This aim could be accomplished using the copula-based time series models since the temporal dependence can be characterized by the copula dependence parameter. Given the estimators of the marginal distribution and the copula dependence parameter, one can estimate the temporal dependence structure of time series models. Darsow, Nguyen, and Olsen [11] study the Markov process using the copulas. Joe [22] studies a class of stationary Markov models in terms of parametric marginal distributions and copulas. Fermanian and Scaillet [14] consider a nonparametric kernel method to estimate the copulas for time series. Moreover, Chen and Fan [6] propose a copula-based semiparametric model for the estimation of a class of stationary Markov processes.

In the second essay, the nonparametric estimation of copula-based stationary Markov Models is proposed. I extend the semiparametric model by Chen and Fan [6] and study a class of time series models in the context of the two-stage ESE in which the copulas density function and the marginal distributions are estimated nonparametrically. In particular, I focus on the stationary Markov process of order 1 with continuous state space because it has the β-mixing property for the analysis of weakly dependent processes. Since many time series models such as nonlinear ARX, nonlinear ARCH, and diffusion models may generate stationary β-mixing observations, the theory developed here is widely applicable. The copula density functions for stationary time series models are approximated by the series estimate on sieve spaces. In this study, a finite dimensional linear space spanned by a sequence of power functions is treated as the sieve space where the estimation space of the copula density function is based. This sieve series estimator can be characterized as the exponential series estimator under mild smoothness conditions. To estimate the unknown copula density function, I propose a two-stage estimator in which the first stage estimates each marginal density separately and in the second stage, the joint density of estimated marginal cumulative distribution functions (CDF) are approximated by the exponential series estimator. By using the β-mixing properties, I show that the copula density function approximated by the ESE for stationary first-order Markov models has the same convergence rate as the estimator of Wu [43] which concentrates on *i.i.d.* data. I also establish the L_2 convergence rate of the estimator of a class of stationary Markov models. I also examine the finite sample performance of the proposed estimator in two examples. In the first example, we discuss the in-sample estimation performance of the proposed estimator and then we discuss the one-step-ahead forecasting performance of the proposed estimator in the second example. The results show that our estimator outperform the kernel estimator in the conditional density estimation except for the Frank copula-based Markov model. In addition, the proposed estimator considerably dominates the kernel estimator when used in the one-step-ahead forecast.

CHAPTER II

ESTIMATION OF HIGH-DIMENSIONAL DENSITIES VIA NONPARAMETRIC COPULA

2.1. Introduction

Estimating a probability distribution plays an important role not only in the social science but in the engineering field. Many methodologies have been developed to provide an adequate estimation for a wide class of distribution. For example, the nonparametric distribution estimation, making no assumption of the distributional form, is able to capture the stylized pattern of underlying distributions. On the other hand, the parametric distribution estimation, under the correct specification of distributional form, can give an efficient estimation for the underlying distribution.

The kernel density estimation, the most popular nonparametric estimator, is an equal mixture of n kernels, centered at the n data points. It is known that the convergence rate of the kernel density estimator is restricted by the 'order' of kernel. Kernels with order higher than 2 can achieve faster rate of convergence. However, higher order kernel density estimator can produce negative density estimates. Another popular density estimator is the series estimator, which is known for its automatic adaptiveness in the sense that it can 'adapt' to the unknown smoothness of the underlying density to achieve the optimal convergence rate. However, the series estimator cannot guarantee the positiveness of density estimates either. Wu [43] proposes an alternative Exponential Series Estimator (ESE) for multivariate densities. The ESE takes the form of an exponential series and thus is strictly positive. Wu [43] demonstrates the efficacy of this density estimator. However, the ESE is defined on a bounded support and further modification is required such to fit fat-tailed distributions.

In this study, we propose an alternative estimator for multivariate densities. This estimator can be characterized as a transformation based estimator. The first stage estimates each marginal density separately. In the second stage, the joint density of estimated marginal cumulative distribution functions (CDF) are approximated by the ESE. The final estimate is then obtained as the product of the marginal densities and the joint density estimated in the second stage. Since the joint density of the marginal CDFs coincides with the copula density among the margins, the procedure can be viewed as a copula based estimator as well.

Wu [43] shows that the ESE is particularly suitable for copula density estimation because it is defined explicitly on a bounded support and free from boundary bias. However, for a high-dimensional variable, the number of moments increases exponentially, manifesting the 'curse of dimensionality'. Let d be the dimension of a random variable X. With a relatively small d, Wu [43] shows that a truncation strategy that includes all moments of the form $\prod_{i=1}^{d} X_i^{r_i}, r_i \geq 0, \sum_{i=1}^{d} r_i < m$, where m is chosen according to some information criterion, produces satisfactory results. However, this approach becomes increasingly cumbersome as d increases. For example, with $d = 3$ and $m = 4$, the full set of moments has 35 elements. For $d = 4$, the count increases to 70. Clearly, this strategy of model selection becomes quickly infeasible as d increases.

A second contribution of this study is to incorporate a variable selection algorithm into a sequential updating process of moment selection to overcome the dimensionality problem. We use the RM proximity indicator, e.q. Cadima, Cerdeira, and Minhoto [2], to select the subsets of moments for each order before the estimation therefore the preselected moments used for estimation in each order is a subset of a full set of moments in each order. For a given d, we impose the restriction that the moments corresponding to the marginals of the first and the second order of moments can not be replaced in the updating process, fit the data using an ESE with the second order cross moment, $m = 2$, and retain only those moments with statistically significant coefficients. We then update the first stage estimate by incorporating all RM-selected moments with $\sum_{i=1}^{d} r_i = 3$. Among those RM-selected third-order moments, we retain only those with significant coefficients. This updating process continues until we reach a pre-specified maximum order of moment, say M. In each stage, we retain all moments inherited from the previous stages and newly-incorporated moments with significant coefficients. In the end, we have M candidate estimates, and each with moments whose order is no higher than $m, m = 1, \ldots, M$. The final estimate is then chosen according to some information criterion such as the Akakie Information Criterion (AIC) or Bayesian Information Criterion (BIC).

To examine finite sample performance of the proposed estimator, we undertake two sets of experiments. The first experiment examines the performance using six sets of mixtures of multivariate normal densities. The second one examines the performance of estimating tail distributions. The proposed estimator outperforms the kernel density estimator for various examples. In addition, the relative performance of our method increases with the number of variables. An illustrative example on estimating the conditional copula density between $S\&P$ 500 and FTSE 100 given Hangseng and Nikkei 225 is also discussed.

This study proceeds as follows. Section II briefly describes the principles of nonparametric density estimator for multivariate cases and of copula method. Section III presents the two-stage transformation-based ESE and its convergence rate in terms of the Kullback-Leibler Information Criterion. An sequential updating method of moment selection is also included in Section III. Section IV discusses the results of Monte Carlo simulations of our method. Section V gives some empirical applications. The last concludes.

2.2. Multivariate Density Estimation

Let $\{X_t\}_{t=1}^n$ be a d dimensional *i.i.d.* random sample from an unknown distribution F with density f defined on the real line, $d \geq 2$. We are interested in estimating f. The parametric approach entails functional form assumptions up to a finite set of unknown parameters. Multivariate normality or more generally, the elliptical family, is commonly used due to its simplicity. Nonparametric approach provides a flexible alternative that seeks a functional approximation to the unknown density. Instead of imposing functional form assumptions, this approach allows the number of (nuisance) parameters to increase with sample size to achieve consistency. One can also combine these two approaches to balance between parsimony and goodness-of-fit. Below we briefly review various methods for multivariate density estimation, with a focus on nonparametric estimators.

2.2.1. Direct Estimation

One of the most commonly used density estimators is kernel density estimator (KDE), which takes the form

$$f_h(x) = \frac{1}{n}\sum_{t=1}^{n} K_h(X_t - x),$$

where $K_h(x)$ is a d-dimensional kernel function that peaks at $x = 0$ and h, the so-called bandwidth, controls how fast $K_h(x)$ decays as x moves away from zero. A popular choice of K is the Gaussian kernel which is the standard normal density function. For multivariate densities, product kernel is commonly used. It is well-known that the performance of KDE crucially depends on the choice of bandwidth but not on kernel function. Data-driven methods, such as cross validation, are often used for bandwidth selection e.g., Li and Racine [24].

Another popular method for density estimation is series estimation. Let $g_i, i = 1, 2, \ldots$, be a series of linearly independent real-valued basis functions defined on R. A series estimator is given by

$$f_m(x) = \sum_{i=1}^{m} \lambda_i g_i(x),$$

where m plays a role similar to bandwidth in kernel estimation and is usually determined by some data-driven methods, such as generalized cross validation. Examples of series estimators include power series, splines, and wavelets.

For a d-dimensional random variable with a r-times continuously differentiable density, both kernel and series estimators can achieve the optimal convergence rate $O_p(n^{-r/(2r+d)})$ in the L_2 norm under some regularity conditions. However, for kernel estimators to achieve a convergence rate faster than $n^{-2/(4+d)}$, one needs to use a higher order kernel, which can lead to negative density estimates. The optimal series estimator has an appealing property of automatically adapting to the smoothness of the underlying distribution, but it also shares the problem of likely negative density estimates. One of the advantages of these estimators is the linearity, which makes it easy to use cross-validation to determine their smoothing parameters and relatively straightforward to derive their asymptotic properties. But the linearity is also their weakness in the sense that their likelihood function, being a product of a sum, is complicated, and they have no sufficient statistics.

Alternatively, there are also likelihood based nonparametric density estimators. One family of estimators takes the form

$$f_m(x) = \exp(\sum_{i=1}^{m} \lambda_i g_i(x) + \lambda_0), \tag{2.1}$$

where $g_i, i = 1, \ldots, m$, are a series of linearly independent functions and $\lambda_0 \equiv \int \exp(\sum_{i=1}^{m} \lambda_i g_i(x))dx < \infty$ ensures that f_m integrates to unity. The estimation of a probability density function by sequences of exponential families, which is equivalent to approximating the logarithm of a density by a series estimator, has long been studied. Earlier studies on the approximation of log densities using polynomials include Neyman [29] and Good [15]. Transforming the polynomial estimate of log-density back to its original scale results in a density estimator in the exponential family. The maximum likelihood method provides efficient estimates of this canonical exponential family. Crain [10] establishes the existence and consistency of the maximum likelihood estimator. Zellner and Highfield [45] and Wu [42] discuss the estimation of (2.1), which typically requires nonlinear optimizations.

One obtains a nonparametric estimator in (2.1) by letting its number of terms m increase with sample size. Kooperberg and Stone [23] and Stone [38] provide in depth analyses of the log-spline density estimator, which is a special case of (2.1) with spline basis functions in its exponent. Barron and Sheu [1] establish the asymptotic properties of (2.1) for general basis functions that include power series, splines and trigonometric series in a unified framework. Wu [43] further generalizes their results to multivariate density estimation. They show that the under suitable regularity conditions, this estimator achieve the optimal rate specified in Stone [37] in terms of the Kullback-Leibler information criterion.

Following the spirit of Barron and Sheu [1], we call this family of density estimator Exponential Series Estimator (ESE) to reflect its nonparametric nature. Like a series estimator, optimal ESE adapts to the smoothness of the underlying distribution automatically. On the other hand, it is strictly positive and has a set sufficient statistics, $E[g_i(x)], i = 1, \ldots, m$, thanks to its general exponential form. In addition, ESE has an appealing information theoretic interpretation. It can be derived as the maximum entropy density by maximizing Shannon's information entropy subject to known moment constraints $E[g_i(x)] = \mu_i, i = 1, \ldots, m$, e.q. Jaynes [21].

2.2.2. Transformation-based Estimation

Transformation of variables of interest to facilitate modeling and estimation is a common practice in statistical and econometric analyses. For example, logarithmic transformation of a positive dependent variable in regression analysis sometimes mitigates heteroskedasticity. More generally, Box-Cox transformation, which nests logarithmic transformation as a limiting case, is often used to remedy deviations from normality in residuals. Although less common, transformations are also used in density estimations.

In the context of nonparametric density estimation, transformations can be used to reduce bias. Wand, Marron, and Ruppert [39] propose a transformation based kernel density estimator. They note that the usual kernel estimators with one global bandwidth work well for densities that are not far from Gaussian in shape, but can perform quite poorly when the densities deviate further from Gaussian. In a spirit close to Box-Cox transformations, they propose transformations of the data so that the density of the new data can be adequately estimated by kernel estimators with a global bandwidth. In particular, they focus on right-skewed data and the shifted power transformation family. They demonstrate that if a transformation is carefully selected, it is much more appropriate to use the typical kernel estimator with a global bandwidth on the transformed data. Consequently, the estimated density of the raw data obtained by back-transformation can have a smaller bias. Yang and Marron [44] further show that multiple families of transformations can be employed at the same time, and there can be benefits to iterating this process.

Wand, Marron, and Ruppert [39] and Yang and Marron [44] consider only parametric transformations, which reduce biases but do not improve in convergence rate. Ruppert and Cline [34] propose a smoothed empirical transformation that both reduces bias and improves convergence rate. Suppose for now X_t is a scalar. First the data are transformed to $\hat{F}(X_t)$, which is a smooth estimate of the CDF of X. The estimated density of the raw data then takes the form

$$\tilde{f}(x) = \frac{1}{n}\sum_{t=1}^{n} K_h(\hat{F}(X_t) - \hat{F}(x))\hat{f}(x),$$

where $\hat{f}(x) \equiv d\hat{F}(x)/dx$. Because $\hat{F}$ converges to a uniform distribution whose density has all derivatives equal to zero, bias of the second stage estimate is asymptotically negligible. They further show that if the bandwidths of the first and second step are chosen to be of order $n^{-1/9}$, then the squared error of $\tilde{f}$ is of order $O_p(n^{-8/9})$ as $n \to \infty$ rather than $O_p(n^{-4/5})$, the rate of an ordinary kernel estimator. This procedure can also be iterated to obtain further rate improvement, although in practice the benefits may be rather small.

Intuitively, both parametric and nonparametric transformations achieve bias reduction by choosing a transformation such that the density of the transformed data is easier to estimate in terms of, say smaller squared errors or integrated squared error. In this study, we apply the nonparametric transformation approach to multivariate density estimations. Let $\hat{F}_j$ and $\hat{f}_j$, $j = 1, \ldots, d$, be estimated marginal CDF and PDF for the jth margin of a d-dimensional data $X = [X_1, \ldots, X_d]$. The transformation based density estimator of X then takes the form

$$\tilde{f}(x) = \hat{f}_1(x_1)\cdots\hat{f}_d(x_d)\hat{c}(\hat{F}_1(x_1), \ldots, \hat{F}_d(x_d)), \tag{2.2}$$

where $\hat{c}$ is the estimate of the density of the transformed data $\left\{\hat{F}_1(x_1), \ldots, \hat{F}_d(x_d)\right\}$.

Interestingly, (2.2) can also be derived using Sklar's theorem. Let f be the density of a d-dimensional random variable, with F_j and f_j its jth marginal CDF and

PDF for $j = 1, \ldots, d$. Sklar [36] shows that the joint density can be decomposed as

$$f(X) = f_1(X_1) \cdots f_d(X_d) c(F_1(X_1), \ldots, F_d(X_d)). \tag{2.3}$$

When all marginal distributions are differentiable, the decomposition is unique. The last factor in (2.3) is termed the copula density, which completely summarizes the dependence structure among X_1 to X_d, e.g. Nelsen [28] for a general treatment of copula).

The copula decomposition allows the separation of marginal distributions and their dependence and thus facilitates construction of flexible multivariate distributions. It has also been used in multivariate analyses, especially on financial data, e.g. Patton [31]. This method has been used for density estimation as well. Hall and Neumeyer [17] shows that copula method can benefit estimation of joint densities when there are additional data for the margins. Chui and Wu [9] provide simulation evidence that two-step estimation via an empirical copula density often outperforms direct estimation of joint densities. However, both papers consider only bivariate densities. To account for application in the complex multivariate model, we propose a transformation-based estimator for the general d-dimensional case.

2.3. Transformation-based Multivariate Density Estimation

In this section we present a nonparametric transformation-based multivariate density estimation and establish its asymptotic properties. We then propose a method of model specification for the second stage estimation of the density of the transformed data, which can be viewed as an estimation of empirical copula density function.

2.3.1. The Estimator

The transformation-based estimation for an *i.i.d.* random vector $\{X_t\}_{t=1}^n$ is constructed in two simple steps. We first obtain consistent estimates of marginal densities and distributions, denoted by $\hat{F}_j$ and $\hat{f}_j$ respectively for $j = 1, \ldots, d$. Note that it is not required that $\hat{f}_j(x) = \hat{F}_j'(x)$. In fact, we can even combine smoothed estimates of marginal densities with empirical CDF's of corresponding margins.

The second step estimates the density of the transformed data $\hat{F}_t = (\hat{F}_1(X_{1t}), \ldots, \hat{F}_d(X_{dt})), t = 1, \ldots, n$. To ease notation, we define $u_t = (u_{1t}, \ldots, u_{jt})$ where $u_{jt} = \hat{F}_{jt}$ for $j = 1, \ldots, d$ and $t = 1, \ldots, n$. As discussed above, the density of $\{u_t\}_{t=1}^n$ coincides with copula density. Like an ordinary density function, one can estimate a copula density using a parametric or nonparametric method. Parametric copula density functions are usually parameterized by one or two parameters. This parsimony in functional forms imposes restrictions on the dependence structure among margins. For example, the popular Gaussian is known to exhibit zero tail dependence. Consequently, it may be inappropriate to use simple Gaussian copulae to investigate the co-movements of extreme stock returns.

Nonparametric estimation of copula densities, on the other hand, ensures consistency. However, compared with its parametric counterpart, nonparametric estimators are known to have slower convergence rates. In addition, since copula densities are defined on a bounded support, treatment of boundary bias warrants special care. Although the boundary bias problem exists for general nonparametric estimation, its consequence is particularly severe for copula density estimation. This is because unlike a lot of densities or curves that vanish at the boundaries, copula densities often spike near the boundaries and corners. For example, the dependence structure of two stock returns is often dominated by co-movements of their extreme tails, giving rise to a copula density that peaks at either end of the diagonal. In this case, a non-parametric estimate, say the kernel estimate, of the copula density without proper boundary bias correction may fail to capture the underlying dependence structure between variables.

In this study, we adopt the ESE to estimate copula density estimation. This estimator has some appealing properties that make it suitable for copula density estimation. First, the ESE copula estimator is always well defined since the copula is defined on a bounded support. Moreover, the ESE is free of the boundary bias when an optimal power series basis is used.

More notations are required for multivariate density estimations. Define a multi-index $i = (i_1, i_2, \ldots, i_d)$, and $|i| = \sum_{j=1}^{d} i_j$. Given two multi-indices i and m, $i \geq m$ indicates $i_j \geq m_j$ elementwise; when m is a scalar, $i \geq m$ means $i_j \geq m$ for all j.

As discussed above, the multivariate ESE of copula density could be derived from the maximization of the Shannon's entropy of the copula density. The multivariate ESE of copula density is obtained by maximizing entropy of the copula density. In particular,

$$H = \int_{[0,1]^d} -c(u) \log c(u) du,$$

subject to the integration to unity

$$\int_{[0,1]^d} c(u) du = 1$$

and side conditions in terms of moments

$$\int_{[0,1]^d} g_i(u)\, c(u)\, du = \hat{\mu}_i, \qquad i \in M,$$

where $\hat{\mu}_i = n^{-1} \sum_{t=1}^{n} g_i(u_t)$, $du = du_1 du_2 \cdots du_d$ and $g_i(u)$ are a sequence of linearly independent polynomials defined on $[0,1]^d$. The estimated multivariate copula density takes the form

$$c(u;\hat{\lambda}) = \exp\,(-\sum_{i\in M}\hat{\lambda}_i g_i(u) - \hat{\lambda}_0)$$

where

$$\hat{\lambda}_0 = \log\,(\int_{[0,1]^d}\exp\,(-\sum_{i\in M}\hat{\lambda}_i g_i(u))du)$$

and $M \equiv \{i : |i| > 0 \text{ and } i \leq m\}$. Given the marginal density functions, the estimated multivariate density is then estimated by $\hat{f}(X) = (\prod_{j=1}^{d}\hat{f}_j(X_j))c(\hat{F}(X);\hat{\lambda})$ where $\hat{F}(X) = (\hat{F}_1(X_1),\ldots,\hat{F}_d(X_d))$.

2.3.2. Asymptotic Properties of Two-stage Multivariate ESE

In this section, we derive the convergence rate of the proposed estimator in terms of In the first stage, we estimate the marginal distribution functions which are used as the frames of copula density function. Therefore the support of true copula density c_0 is the hypercube $[0,1]^d$. The basis functions g_i are a sequence of linear independent polynomials. We assume, without loss of generality, g_i are normalized Legendre polynomials.

Assumption 1 *The observed data* $X_1 = [X_{11}, X_{21}, \ldots, X_{d1}]$, $X_2 = [X_{12}, X_{22}, \ldots, X_{d2}]$, $\ldots$, $X_n = [X_{1n}, X_{2n}, \ldots, X_{dn}]$ *are i.i.d. continuously random samples with the joint density* $p_0(x)$*, the marginal densities* f_j *and the marginal distributions* F_j*.*

Assumption 2 *Let* $f_0(x) = \log c_0(x)$ *such that* $|f_0(x)| < \infty$ *and* $f_0(x)$ *is a member of a Sobolev space* S_m *in which* $f_0^{(r-1)}(x)$ *is absolutely continuous and* $\int_d (f_0^{(r)}(x))^2 dx < \infty$ *for* $r > d$*.* $r = \sum_{j=1}^{d} r_j$ *for nonnegative integers* r_j*. For the univariate case,* $\log f_j^{(s_j-1)}(X_j)$ *is absolutely continuous and* $\int (\log f_j^{(s_j)}(X_j))^2 dX_j < \infty$ *where* s_j *is a nonnegative integer.*

Assumption 3 $\prod_{j=1}^{d} m_j \to \infty$, *and* $(\prod_{j=1}^{d} m_j{}^3)/n \to 0$ *for nonnegative* m_j *as* $n \to \infty$. *For the univariate case,* $\tilde{m}_j \to \infty$ *and* $\tilde{m}_j^3/n \to 0$ *for* $j = 1, 2, \ldots, d$ *as* $n \to \infty$.

The proposed two-stage ESE of copula function is given by $c_{\hat{\hat{\lambda}}} = \exp(-\sum_{i \in M} \hat{\hat{\lambda}}_i \, g_i(\hat{F}_1, \ldots, \hat{F}_d) - \hat{\hat{\lambda}}_0)$. It follows $f_{\hat{\hat{\lambda}}} = -\sum_{i \in M} \hat{\hat{\lambda}}_i \, g_i(\hat{F}_1, \ldots, \hat{F}_j) - \hat{\hat{\lambda}}_0$.

Proposition 4 *If* F_j, $j = 1, \ldots, d$, *are known, the ESE of copula density,* $c_{\hat{\lambda}} = \exp(-\sum_{i \in M} \hat{\lambda}_i \, g_i(F_1, \ldots, F_j) - \hat{\lambda}_0)$, *converges to* c_0 *in the sense of KLIC with the convergence rate*

$$D(c_0 || c_{\hat{\lambda}}) = O_p(\prod_{j=1}^{d} m_j^{-2r_j} + \prod_{j=1}^{d} m_j/n)$$

Proposition 5 *Assume we estimate the marginal densities* f_j, $j = 1. \ldots, d$ *by the ESE. The multivariate ESE* $p_{\hat{\lambda}}$ *converges to* p_0 *with KLIC rate*

$$D(p_0 || p_{\hat{\lambda}}) = O_p(\max_j(\tilde{m}_j^{-2s_j} + \tilde{m}_j/n) + \prod_{j=1}^{d} m_j^{-2r_j} + \prod_{j=1}^{d} m_j/n)$$

In the following theorem, we prove the convergence rate of the two-stage ESE, $p_{\hat{\hat{\lambda}}} = (\prod_{j=1}^{d} \hat{f}_j) \, c_{\hat{\hat{\lambda}}}$, in terms of KLIC.

Theorem 6 *The two-stage ESE* $p_{\hat{\hat{\lambda}}}$ *converges to* p_0 *in the sense of KLIC with the following rate*

$$D(p_0 || p_{\hat{\hat{\lambda}}}) = O_p(\max_j(\tilde{m}_j^{-2s_j} + \tilde{m}_j/n) + \prod_{j=1}^{d} m_j^{-2r_j} + \prod_{j=1}^{d} m_j/n)$$

From Theorem 6, we know that the convergence rate of the two-stage ESE is determined by the convergent rate of marginal density estimators $D(\prod_{j=1}^{d} f_j || \prod_{j=1}^{d} \hat{f}_j)$ and the convergent rate of the ESE of copula density $D(c_0 || c_{\hat{\lambda}})$. Therefore, the convergent rate of the two-stage ESE is dominated by that of the ESE of copula density if $\prod_{j=1}^{d} m_j^{-2r_j} + \prod_{j=1}^{d} m_j/n \geq \max_j(\tilde{m}_j^{-2s_j} + \frac{\tilde{m}_j}{n})$.

2.3.3. Model Specification

In the ESE, moment criteria provide the information about the underlying density. Therefore, the number of moments used in the ESE should be large enough for an accurate estimation, particularly when the dimensionality goes high. However, too many moment criterions included in the ESE result in the dimensionality problem. To deal with this dilemma of dimensionality, a typical way is through a principal component analysis (PCA). However, dimensionality reduction via PCA still involves all of the moments which lead to a worse estimation performance in our case since the trivial moments produce extra noise. Another approach to reduce the dimensionality is to identify subsets of variables that best approximate the full set of variables, e.q. McCabe [27], Cadima and Jolliffe [3] and Cadima, Cerdeira, and Minhoto [2]. This algorithm to identify a subset of a set of original moments is to select moments which are optimal for a given criterion that measures how well each subset approximates the whole set. In this study, we maximize the RM criterion over all possible subsets of moments for a specific order to select a subset of moments ahead of the estimation for the purpose of reducing the dimensionality. The RM criterion measures the correlation between the n by p matrix Z and the n by p matrix whose columns come from projecting each of the p observed variables on Q:

$$RM = corr(Z, P_Q Z) = \sqrt{\frac{trace(Z^t P_Q Z)}{trace(Z^t Z)}} = \sqrt{\frac{\sum_{i=1}^{p} \lambda_i (r_m)_i^2}{\sum_{j=1}^{p} \lambda_j}}$$

where P_Q is the matrix of orthogonal projections on Q and λ_i is the variance of ith PC. $(r_m)_i$ measures the multiple correlations between the data set's ith PC and the q variables spanning Q. Therefore, $(RM)^2$ can be interpreted as the percentage of total variance accounted for by the q variables. It implies the maximization of RM selects the q-variable subset that maximize the same criterion as PCA.

While using the RM condition to select the subsets which well approximate the full set of variables, one needs to choose the dimension of Q, i.e. q. In this study, we evaluate q using the following process. Suppose we already have the density function, say $\hat{f}_t$, estimated by the moments up to the order t and we want to calculate the q for the order $t+1$, say q_{t+1}. First we evaluate the correlation between the matrix of empirical moments for the order $t+1$ and the matrix of predicted moments on the basis of $\hat{f}_t$ for the order $t+1$. Denote this correlation coefficient by δ_{t+1}. Then we calculate the number q_{t+1} by $integer(\sqrt{(1-\delta_{t+1})}p_{t+1})$ where p_{t+1} is the size of moments for the order $t+1$. $(1-\delta_{t+1})$ captures the discrepancy between the matrix of empirical moments for the order $t+1$ and the matrix of predicted moments for the order $t+1$. The smaller is the discrepancy, the less is the additional information contained in the moment conditions of order $t+1$ and therefore the fewer number of additional moment conditions, calculated by $integer(\sqrt{(1-\delta_{t+1})}p_{t+1})$, is warranted in the current stage of updating.

Instead of including all feasible moment constraints at the same time, this study proposes a t-based updating process for the selection of individual moments and employs a data-driven method for the selection of the order of moments. As mentioned in Kooperberg and Stone [23], the t statistics could be used as the criterion of selection of "knots" in the logspline model. In this study, we impose the restriction that the moments corresponding to the marginals of the first and the second order of moments can not be replaced in the updating process. We outline the updating process as follows.

1. We use the polynomials corresponding the first and the second moments, say $K_1 = \{g_i(u) : |i| = 1\}$ and $K_2 = \{g_i(u) : |i| = 2\}$ in the estimation. We then drop the polynomials, matching second order cross moments, with insignificant

t statistics out of the estimation. Denote the set of remaining polynomials in the estimation by k_1. We also evaluate the density function $\hat{f}_1$, the AIC criterion AIC_1 and the BIC criterion BIC_1 based on k_1.

2. In this stage, we use the RM criterion to select the subset of polynomials corresponding to the third order of moments. Denote this subset of polynomials by $\tilde{K_3}$. We use $\tilde{K_3}$ as well as k_1 in the estimation.

   (a) We use the density function estimated in the previous stage $\hat{f}_1$ to calculate matrix correlation δ_2 so that we can get q_2 by $integer(\sqrt{(1-\delta_2)}p_2)$ where p_2 is the size of second-order moments. When q_2 is known, the selected set of polynomials, $\tilde{K_3}$, is the subset of $K_3 = \{g_i(u) : |i| = 3\}$ which maximizes the RM criterion.

   (b) We use $\tilde{K_3}$ as well as k_1 in the estimation, drop the polynomials in $\tilde{K_3}$ with insignificant t statistics. The subset of polynomials in $\tilde{K_3}$ with significant t statistics is given by k_2. The density function $\hat{f}_2$, the AIC criterion AIC_2 and the BIC criterion BIC_2 based on k_2 are also evaluated.

3. We repeat the procedures in the second step in the estimation including the following order of moments until the maximum order of moment m is reached. k_j, AIC_j and BIC_j for $j = 1, \ldots, m-1$ are obtained.

Finally, one needs to specify the optimal order of moments for the exponential series copula density estimator. Wu [43] suggested that the order of polynomials can be chosen using data-driven methods such as AIC and BIC. For example, we can use the likelihood-based AIC as the criterion to select the optimal degree of moments. Therefore the polynomials corresponding to the optimal order of moment is $\hat{k} = \{k_i, i = 1, 2, \ldots, m-1 : AIC_{\hat{k}} = \min_{i=1,2,\ldots,m-1} AIC_i\}$.

The RM-based order updating procedure has several appealing features. First, the effective number of estimated parameters in our model is usually smaller than that in Chui and Wu [9] so that the curse of dimensionality is mitigated considerably. Second, the updating process in terms of t statistics effectively removes the moments that play insignificant role in the estimation. Another important advantage of our method is that it is computation-friendly . For example, in the three dimensional case with $m = 4$, the number of moments used in the computation at the same time is 35 in Chui and Wu's method; whereas, the moments used in the computation of each step is usually less than 10.

2.4. Monte Carlo Simulation

In this section, we conduct Monte Carlo simulations to investigate the finite sample performance of the proposed estimator. We also compare the performance of the ESE with the empirical and kernel estimators on the estimation of tail of distributions. Six types of bivariate densities including uncorrelated normal, correlated normal, skewed, kurtotic, bimodal I and bimodal II, investigated in Wand and Jones [40], are used as benchmarks. Moreover, we consider mixtures of trivariate normals with similar features. The parameters for these normal mixtures are given in the appendix. The maximum order of moment is set to be four ($m = 4$). In both examples, the marginal density and distribution functions used in the ESE are estimated by the kernel density estimator (KDE). While using the KDE in the simulation, the Gaussian kernel is used and the bandwidth is chosen by the least square cross-validation (LSCV).

Our first example concerns the estimation of true density via nonparametric copula. The sample sizes are 100, 200 and 500, and each experiment is repeated for

300 times. We use both the AIC and the BIC to select the $\hat{k}$, the optimal order of moments to be incorporated in our estimator. For comparison, we also estimate the densities using the KDE. The performance is gauged by the integrated squared errors (ISE) evaluated on $[-3, 3]^d, d = 2$ and 3, with the increment of 0.15 in each dimension. Because the performance of our method under the AIC and the BIC selection of moments are similar, we only report the results using the AIC. The results are reported in Table I.

Three panels of Table I reports the ISE and the corresponding standard error for $d = 2$ and 3 respectively. For the bivariate case shown in the top panel, the performance of both estimators improves with the sample size. In our experiments, the ESE outperforms the KDE in all cases. The average ratios of the ISE between the ESE and the KDE across all six distributions are 22%, 25% and 46% when the sample sizes are 100, 200 and 500. The results for the trivariate case are reported in the bottom panel. The general pattern of performance remains the same. The corresponding average ISE ratio between the ESE and the KDE improves to 31%, 41% and 48%.

The average ISE ratio between the ESE and the KDE improves with the sample size across six distributions for $d = 2$ and 3. Also this average ratio generally improves with the dimensionality of space. This desirable result indicates that our method gets better in the relative performance with respect to the KDE as the dimensionality of a density increases.

In the financial risk management, modelling of extreme financial returns has become critical issues, e.q. Chan and Li [4]. Extreme value theory, characterizing the extremal characteristics of stationary distributions, allows us to make inference about extremal behaviors of returns. To this end, the estimation of tail index is fundamental, for which theory offers various of different approaches. In this study,

Table I. ISE of Joint Density Estimation

	N=100		N=200		N=500	
	KDE	ESE	KDE	ESE	KDE	ESE
			d=2			
uncorrelated normal	0.0103	0.0086	0.0064	0.0052	0.0035	0.0023
	(0.0047)	(0.0075)	(0.0026)	(0.0050)	(0.0012)	(0.0019)
correlated normal	0.0094	0.0081	0.0058	0.0048	0.0034	0.0028
	(0.0035)	(0.0074)	(0.0017)	(0.0033)	(0.0008)	(0.0011)
skewed	0.0223	0.0141	0.0190	0.0104	0.0146	0.0072
	(0.0061)	(0.0082)	(0.0043)	(0.0046)	(0.0029)	(0.0025)
kurtotic	0.0202	0.0153	0.0152	0.0129	0.0099	0.0010
	(0.0052)	(0.0037)	(0.0042)	(0.0023)	(0.0026)	(0.0010)
bimodal I	0.0075	0.0065	0.0049	0.0038	0.0028	0.0018
	(0.0029)	(0.0045)	(0.0017)	(0.0024)	(0.0009)	(0.0010)
bimodal II	0.0142	0.0107	0.0095	0.0065	0.0054	0.0029
	(0.0050)	(0.0080)	(0.0027)	(0.0044)	(0.0014)	(0.0015)
			d=3			
uncorrelated normal	0.0078	0.0058	0.0053	0.0028	0.0033	0.0013
	(0.0024)	(0.0041)	(0.0015)	(0.0019)	(0.0008)	(0.0007)
correlated normal	0.0058	0.0041	0.0040	0.0024	0.0026	0.0015
	(0.0014)	(0.0031)	(0.0008)	(0.0011)	(0.0005)	(0.0005)
skewed	0.0132	0.0090	0.0098	0.0053	0.0067	0.0027
	(0.0032)	(0.0064)	(0.0022)	(0.0021)	(0.0013)	(0.0008)
kurtotic	0.0218	0.0160	0.0192	0.0148	0.0146	0.0136
	(0.0027)	(0.0025)	(0.0022)	(0.0015)	(0.0017)	(0.0010)
bimodal I	0.0059	0.0042	0.0040	0.0023	0.0026	0.0011
	(0.0015)	(0.0031)	(0.0008)	(0.0014)	(0.0005)	(0.0005)
bimodal II	0.0072	0.0042	0.0051	0.0026	0.0033	0.0012
	(0.0018)	(0.0025)	(0.0011)	(0.0014)	(0.0006)	(0.0005)

NOTE:

1. KDE = Kernel estimator

2. ESE = Exponential series estimator

the second example compares the ESE with the empirical estimator and the KDE on the estimation of a tail index of a distribution. The tail index of a distribution is given by

$$T = \int_{[-\infty, q_i]^d} f(x)dx$$

where $q_i \equiv F_i^{-1}(\alpha)$, $i = 1, \cdots, d$. The benchmark densities are the same as before. The sample size is 100 and each experiment is repeated for 500 times. 5% and 10% of the marginal low-tail distribution (i.e. $\alpha = 5\%$ and 10%) are considered. The empirical estimator for tail distributions is given by $\sum_{j=1}^{n} I\{X_j \leq q_1\}$ for the univariate case. Like the previous example, we use the AIC for the specification of the ESE. In this example, the finite sample performance is measured by the mean squared errors (MSE), which is the average of the difference between the estimated tail index and the true tail index. The MSE and the corresponding standard error are shown in Table II. For d=2, the ESE outperforms the empirical estimator and KDE in terms of MSE at 5% and 10% marginal distributions. However, the ESE is slightly dominated by the KDE in the simulation of the kurtotic case at 10% marginal distribution for d=3. Finally, the MSE increases with the percentile of marginal distribution. Regarding the sample variance, the ESE dominates the others in terms of small sample variance.

For d=2, the average ratio of the MSE between the ESE and the KDE across six distributions is 63% at 5% marginal distribution and 47% at 10% marginal distribution respectively; whereas for d=3 the average MSE ratio between the ESE and the KDE is 75% and 53% respectively. The average ratio of MSE between the ESE and the KDE improves with the dimensionality of the sample space.

It has been revealed that smoothing methods usually improve the estimates of densities especially in the multivariate cases. However, there are some exceptions

Table II. ISE of Tail Distribution Estimation

percentile	5%			10%		
	EM	KDE	ESE	EM	KDE	ESE
			d=2			
uncorrelated normal	1.0320	1.7570	0.2949	2.6780	4.6047	1.1665
	(1.7724)	(2.3339)	(0.4981)	(4.0155)	(5.7700)	(1.9452)
correlated normal	1.8451	1.3867	0.9333	4.2959	3.2371	2.7643
	(2.8861)	(2.4799)	(0.8883)	(5.9669)	(4.5019)	(3.0975)
skewed	1.0076	0.7697	0.3493	3.0475	2.7579	1.8587
	(1.6747)	(1.2471)	(0.4207)	(4.2814)	(4.0362)	(2.1085)
kurtotic	1.3930	1.1677	0.6777	2.9170	2.5453	2.3564
	(2.4307)	(2.0369)	(0.6613)	(4.3901)	(4.2056)	(2.1718)
bimodal I	0.2045	0.3210	0.0483	1.0200	1.5901	0.3142
	(0.3951)	(0.5513)	(0.0904)	(1.7069)	(1.9928)	(0.5367)
bimodal II	0.2095	0.3652	0.0602	0.9720	1.5850	0.3962
	(0.4327)	(0.5585)	(0.0910)	(1.4723)	(1.9710)	(0.7425)
			d=3			
uncorrelated normal	0.1331	0.1352	0.0103	0.4535	0.6440	0.0805
	(0.4129)	(0.2539)	(0.0218)	(0.8620)	(0.8642)	(0.1819)
correlated normal	0.5059	0.2822	0.0971	1.7460	0.9812	0.5602
	(1.2809)	(0.6333)	(0.0616)	(2.5603)	(1.7711)	(0.5968)
skewed	0.1829	0.1111	0.0296	0.8814	0.5762	0.4666
	(0.3992)	(0.2475)	(0.0134)	(1.2628)	(1.0695)	(0.2811)
kurtotic	0.5730	0.3607	0.2699	1.8735	1.2931	1.4767
	(0.9976)	(0.6043)	(0.0855)	(2.8469)	(2.1702)	(0.9358)
bimodal I	0.0119	0.0080	0.0003	0.0996	0.0909	0.0076
	(0.1063)	(0.0263)	(0.0007)	(0.3228)	(0.1711)	(0.0134)
bimodal II	0.0080	0.0108	0.0004	0.1172	0.0995	0.0080
	(0.0869)	(0.0371)	(0.0008)	(0.3940)	(0.2023)	(0.0177)

NOTE:

1. EM = Empirical estimator

that the KDE performs worse than the empirical estimator. On the other hand, the ESE always outperforms the empirical estimator. Comparing between the ESE and the KDE, the ESE is substantially better than the KDE, sometimes ten times better.

2.5. Empirical Application

Starting from 1990s, Asian financial markets have become an important role in the global financial system. In the financial literature, more and more studies has appeared to address the impact of Asian financial markets on the European and American financial markets. Especially in 1998, the Asian financial crisis began in many economies in Asia raised fears of worldwide economic collapse due to financial contagion. A statistical tool which is usually used to analyze the response given some known information is the conditional probability of one market given the performance of some other markets. In this section, we applied the proposed ESE to the estimation of conditional probability density and copula density of stock return indices. We examine the conditional probability density functions of monthly stock return indices, S&P 500 (US) and FTSE 100 (UK) given marginal distributions of Hangseng (HK) and Nikkei 225 (JP) below 15%, between 15% and 30%, between 40% and 60%, between 70% and 85%, and above 85%. For more insightful information of the comovement of the US and the UK markets, the copula density function is estimated .

Monthly S&P 500 (namely Y_1), FTSE 100 (namely Y_2), Hangseng (namely Y_3) and Nikkei 225 (namely Y_4) indices are collected for February 1978 through May 2006. For each market, We calculate the rate of return R_t by $logP_t - logP_{t-1}$. To include the dynamic structure, we use a GARCH (1,1) model which assumes $R_t = \mu + u_t$ where $u_t \sim N(0, h_t)$ and $h_t = \gamma + \alpha u_{t-1}^2 + \beta h_{t-1}$ and then estimate the standardized residuals. To employ the two-stage ESE on estimate of conditional density of the US and the UK markets based on the HK and the JP markets, we estimate the marginal density and distribution functions by the KDE and the joint copula density of four indices by the ESE. The model specification of the ESE is determined by the AIC and the maximum order of moment is set to be four. The estimated conditional density function by the ESE is then given by

$$\hat{f}(y_1, y_2 | y_3 \in \tilde{F}_3^{-1}(\Delta), y_4 \in \tilde{F}_4^{-1}(\Delta)) = \frac{\hat{f}_1(y_1)\,\hat{f}_2(y_2)\,\hat{f}_3(y_3)|_{y_3 \in \tilde{F}_3^{-1}(\Delta)}\,\hat{f}_4(y_4)|_{y_4 \in \tilde{F}_4^{-1}(\Delta)}}{\hat{f}(y_3, y_4)|_{y_3 \in \tilde{F}_3^{-1}(\Delta), y_4 \in \tilde{F}_4^{-1}(\Delta)}} \cdot$$

$$\hat{c}(\hat{F}_1(y_1), \hat{F}_2(y_2), \hat{F}_3(y_3), \hat{F}_4(y_4))|_{(\hat{F}_3(y_3), \hat{F}_4(y_4)) \in \Delta}$$

where $\hat{f}_i$ and $\hat{F}_i$ are estimated marginal density and distribution functions for $i = 1, \cdots, 4$. $\tilde{F}_j^{-1}$ are empirical quantile functions for $j = 3$ and 4. $\hat{c}(\cdot)$ is the estimated copula density function by the ESE. Δ is the assigned ranges of marginal distributions of the HK and the JP markets. In this application, we focus on five different ranges: $\Delta \in (0, 15\%], (15\%, 30\%], (40\%, 60\%], (70\%, 85\%]$, and $(85\%, 1]$ which are denoted by Range 1, 2, 3, 4, and 5.

Before the estimate of conditional density functions, we investigate the conditional dependence between the US and the UK markets given the HK and the JP markets in terms of the Kendall's τ. The Kendall's τ measures the nonlinear dependence via the degree of dependence between two rankings. The Kendall's τ is defined as a function of the copula:

$$\tau(C) = 4 \int_{[0,1]^2} C(u_1, u_2) dC(u_1, u_2) - 1$$

where C is the copula function. In this study, we first estimate the copula density

Table III. Estimated Conditional Dependence Measures between $S\&P$ 500 (U.S.) and FTSE 100 (U.K.)

percentile of marginal distribution of Hangseng and Nikkei 225	Kendall's τ
$0 - 15\%$	0.2917
$15\% - 30\%$	0.3000
$40\% - 60\%$	0.3324
$70\% - 85\%$	0.3119
$85\% - 1$	0.2928

function $\hat{c}(\hat{F}_1(y_1), \hat{F}_2(y_2), \hat{F}_3(y_3), \hat{F}_4(y_4))$ by the ESE and obtain the Kendall's τ by

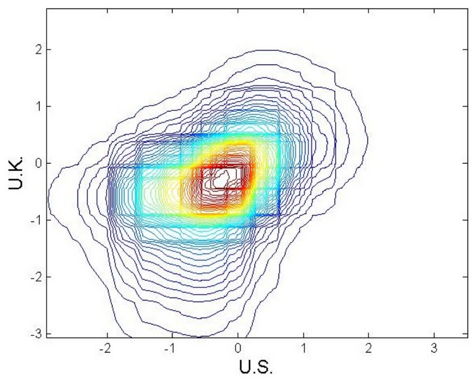
$$\hat{\tau}(C) = 4 \int_{[0,1]^2} \hat{C}(u_1, u_2, u_3, u_4)|_{(u_3,u_4)\in\Delta}\, \hat{c}(u_1, u_2, u_3, u_4)|_{(u_3,u_4)\in\Delta} du_1 du_2 - 1$$

where $\hat{C}(\cdot)$ is the copula function evaluated by the copula density function $\hat{c}(\cdot)$.

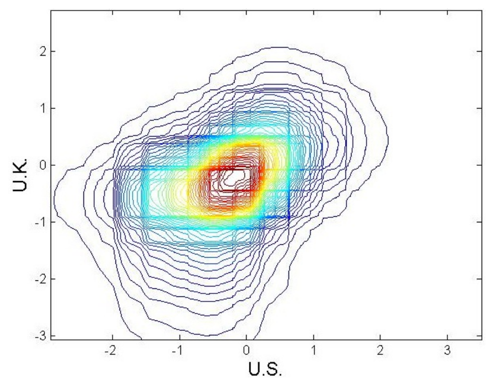
The estimated conditional Kendall's τ is reported in Table III. The conditional Kendall's τ measure remains very similar across different marginal distributions of the HK and the JP markets except $\Delta \in (40\%, 60\%]$. The conditional Kendall's τ coefficients at the low tails $\Delta \in (0, 15\%]$ and $\Delta \in (15, 30\%]$ are smaller than those at $\Delta \in (70\%, 85\%]$ and $\Delta \in (85\%, 1]$ respectively. This pattern suggests that the US market has the higher degree of dependence with the UK market when the Asian markets (e.q. HK and JP) boom. Besides, the largest conditional Kendall's τ happens at $\Delta \in (40\%, 60\%]$, which implies the US and the UK markets exhibits the highest degree of positive dependence when the HK and the JP markets remain stable.

The estimated conditional density functions of the US and the UK markets given different ranges of marginal distribution of the HK and the JP markets estimated by the ESE are shown in Figure 1. Consistent with the estimated conditional Kendall's τ, the estimated conditional density function between the US and the UK markets exhibits the highest degree of positive dependence at the middle range ($\Delta \in (40\%, 60\%]$) of marginal distributions of the HK and the JP markets by its relatively narrow profile along the diagonal. Besides, the estimated conditional densities of the US and the UK markets given low stock returns of the HK and the JP markets exhibit that there is a high possibility that the US and the UK markets simultaneously stagnate or move down when the HK and the JP markets both go down. To gain further insight into the interactions between the joint behavior of the US and the UK markets and joint behavior of the HK and the JP markets, I investigate their corresponding copula densities below.

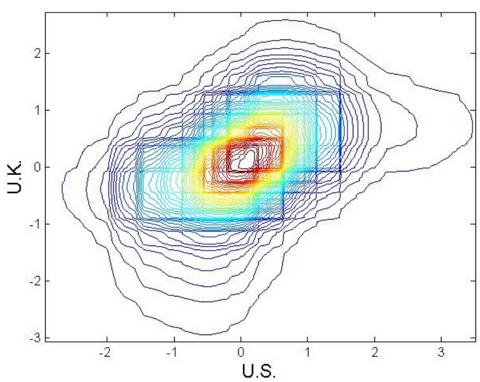
The copula densities of the US and the UK markets given different ranges of marginal distribution of the HK and the JP markets are shown in Figure 2. Except for the saddle point shape of the copula density of the US and UK markets given the middle range ($\Delta \in (40\%, 60\%]$) of marginal distributions of the HK and the JP markets, the copula densities of the US and the UK markets given the remaining ranges ($\Delta \in (0, 15\%], (15\%, 30\%], (70\%, 85\%]$, and $(85\%, 1]$) exhibit the similar bell shapes. Since all copula densities are defined on the unit square, their density values could be compared. The comovement of the US and the UK markets are likely to be influenced by the environments of the HK and the JP markets. For example, when the HK and the JP markets are down, the joint behavior of the US and the UK markets appears to be down too. On the other hand, the US and the UK stock markets both boom when the HK and the JP stock markets are thriving. In general, the joint behaviors of the US and UK markets have high correlations with those of the HK and the JP markets especially when the HK and the JP markets are at their peaks.



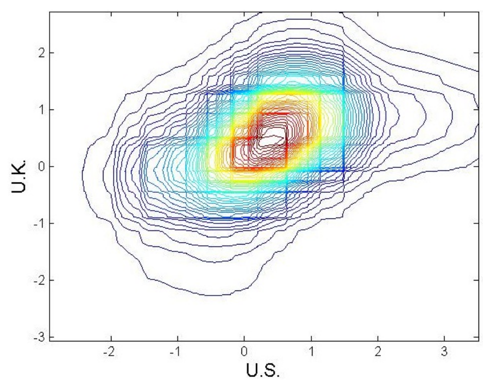
(a) Range 1



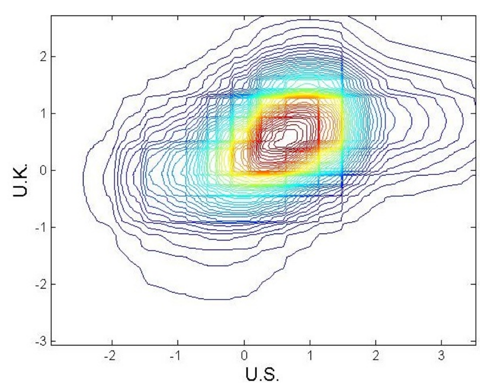
(b) Range 2



(c) Range 3

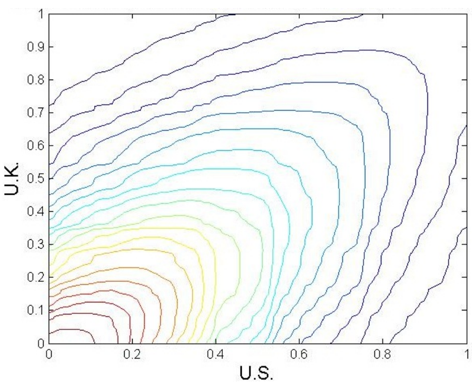


(d) Range 4

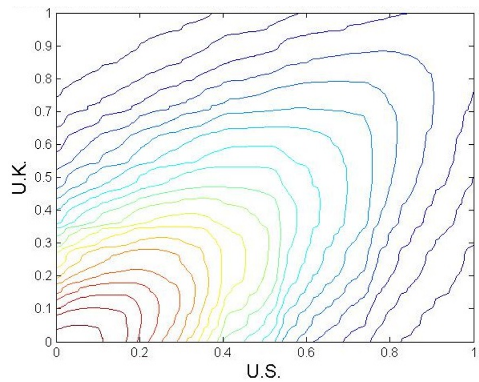


(e) Range 5

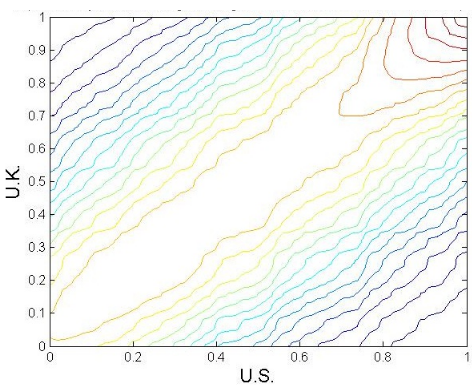
Fig. 1. Contours of conditional density estimation of S&P 500 and FTSE 100 given different ranges of marginal distributions of Hangseng and Nikkei 225



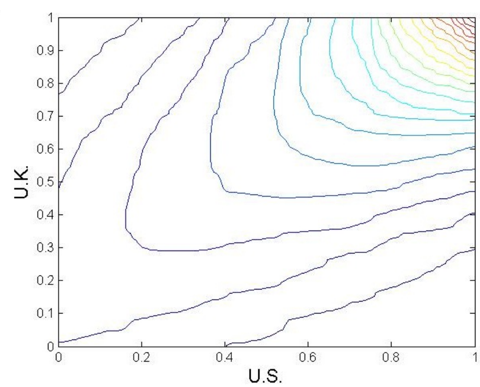
(a) Range 1



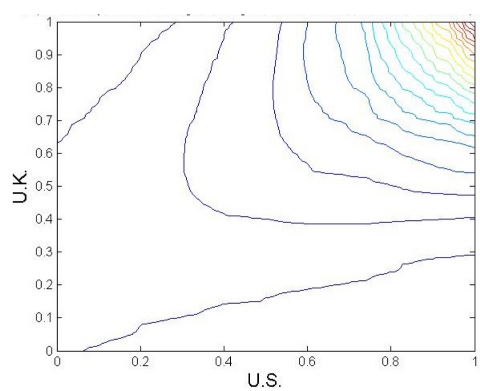
(b) Range 2



(c) Range 3



(d) Range 4



(e) Range 5

Fig. 2. Contours of conditional density estimation of S&P 500 and FTSE 100 given different ranges of marginal distributions of Hangseng and Nikkei 225

The examination of the conditional densities provides important insight into the joint behavior of the stock markets given the behavior of other markets. For example, when the HK and the JP markets are thriving or declining, the conditional densities of the US and the UK markets reveal an information that can not be found in the unconditional density. The US and the UK markets are more likely to be influenced by the environments of the HK and the JP markets when the HK and the JP markets undergo a boom than when the HK and the JP markets suffer from a recession. The asymmetric effects between the joint behavior of the US and the UK markets and that of the HK and the JP markets imply the US and the UK markets are more robust to the recession than the HK and the JP markets.

2.6. Conclusion

This study proposes a two-stage multivariate exponential series estimator via a copula density. The marginal density and distribution functions of each variable are estimated in the first stage and the joint copula density, in terms of the estimated marginal distributions, is then approximated by the exponential series estimation. Finally, the joint density is obtained by the product of marginal densities and the joint copula density estimated in the second stage. A sequential updating method of moment selection is incorporated to select informative moments.

I examine the finite sample performance of the estimator in two experiments. I first investigate the performance using various multivariate normal mixtures. Besides, I estimate the tail distribution. The results show that our method considerably outperforms the kernel estimator in the density estimation. In addition, the proposed method provides superior estimates to the kernel and empirical density estimators in the tail distribution except for the kurtotic case. Finally, I apply our method to estimate the conditional copula density between $S\&P$ 500 and FTSE 100 given Hangseng and Nikkei 225.

CHAPTER III

EXPONENTIAL SERIES ESTIMATION OF COPULA-BASED FIRST ORDER MARKOV PROCESS

3.1. Introduction

The measure of risk in financial portfolio analysis directly relies on the measure of dependence among assets within the portfolio. For instance, the distribution of the return on a portfolio relies not only on the distribution of individual asset but on the dependence between assets, which could be captured by a copula function. A copula is a statistical tool for modelling the multivariate dependence structure among variables without any distribution constraint. By Sklar's theorem, the model of a multivariate density function can be separately identified by the individual marginal densities and the copula density function. Embrechts, McNeil, and Straumann [13] provides general reviews of application of copulas on financial analysis.

The number of papers on copula theory in analyzing contemporaneous dependence between several time series has grown enormously in recent years but the model of temporal dependence of a univariate time series has less been addressed in the literature. Robinson [33] applies multivariate Kernel probability density and regression estimators to a univariate strictly stationary time series. Darsow, Nguyen, and Olsen [11] study the Markov process using the copulas. Joe [22] studies a class of stationary Markov models in terms of parametric marginal distributions and copulas. Fermanian and Scaillet [14] consider a nonparametric kernel method to estimate the copulas for time series. Moreover, Chen and Fan [6] propose a copula-based semiparametric model for the estimation of a class of stationary Markov processes.

Many nonparametric methods such as splines and wavelets can be regarded as examples of sieve extremum estimation. The method of sieve, Grenander [16] , maximizes an empirical criterion function over a sequence of approximating spaces which are called sieve spaces. The sieve space is chosen as a dense of the underlying parameter space and its size is allowed to increase with the sample size. Shen and Wong [35] derived the convergent rates for general sieve M-estimation. Newey and Powell [30] established the consistency of sieve minimum distance estimates. For weakly dependent data, White and Wooldridge [41] establish consistency of sieve extremum estimates. Chen and Shen [7] derive the convergence rates of sieve extremum estimates and root-n asymptotic normality of sieve extremum estimates for stationary β-mixing observations. Chen [5] obtains the rate of sieve extremum estimates for both *i.i.d.* and weakly dependent data.

Another nonparametric estimators, called Exponential Series Estimator (ESE), has also drawn considerable attention in the literature, e.q. Barron and Sheu [1], Marsh [26] and Wu [43]. The ESE is based on the method of Maximum Entropy density subject to a given set of moment conditions. Compared with other nonparametric estimators, the effective number of parameters is largely reduced while using the ESE for smooth functionals. Due to several appealing properties, the ESE is increasingly used in the estimation and hypothesis testing in terms of density functions. However, there is no study using the ESE on analyzing time series models, especially on stationary Markov processes.

In this paper, I extend the semiparametric model by Chen and Fan [6] and study a class of time series models in the context of the two-stage ESE in which the copulas density function and the marginal distributions are estimated nonparametrically. In particular, I focus on the stationary Markov process of order 1 with continuous state space because it has the β-mixing property for the analysis of weakly dependent processes. The copula density functions for time series models are approximated by the series estimate on sieve spaces. In this study, a finite dimensional linear space spanned by a sequence of power functions is treated as the sieve space where the estimation space of the copula density function is based. This sieve series estimator can be characterized as the Exponential Series Estimator under mild smoothness conditions. By using the β-mixing properties, I show that the copula density function approximated by the ESE for stationary first-order Markov models has the same convergence rate as the estimator of Wu [43]. It helps us establish the consistency of the proposed two-stage estimator. To examine finite sample performance of the proposed estimator, I undertake two sets of experiments. In the first example, I discuss the in-sample estimation performance of the proposed estimator and then I discuss the one-step-ahead forecasting performance of the proposed estimator in the second example. The results show that our estimator outperform the kernel estimator in the conditional density estimation except for the Frank copula-based Markov model. In addition, the proposed estimator considerably dominates the kernel estimator when used in the one-step-ahead forecast.

The rest of this paper is organized as follows. In Section II, I introduce the exponential series estimator for the time series models. Also, the consistency of the estimator is verified. A sequential updating method of moment selection is also included in Section II. In Section III, I conduct Monte Carlo simulation to examine the finite sample performance of the proposed estimator. Section IV concludes with discussions.

3.2. Two-step Estimation of Copula-based Markov Models of Order 1

In this section, I present a two-stage estimator for a class of univariate copula-based nonparametric time series models and derive its convergent rate. I also demonstrate the method of model specification for the second stage estimation of the copula density function. Assume $\{Y_t\}$ is a stationary Markov process of order 1 with continuous state space. The joint distribution function of Y_{t-1} and Y_t, say $G(y_0, y_1)$, can be used to determine to probabilistic property of Y_t. By Sklar's [36] theorem, the joint distribution $G(y_0, y_1)$ can be interpreted by the marginal distribution of Y_t, $F(Y_t)$ and the copula density function of Y_t and Y_{t-1}, $c(F(y_{t-1}), F(y_t))$. Therefore, to estimate $G(y_0, y_1)$, I can estimate the marginal distributions $F(y_t)$ and then estimate $c(F(y_{t-1}), F(y_t))$ based on the estimated marginal distributions $\hat{F}(y_t)$.

3.2.1. First Step of Estimation of Copula-based Markov Models of Order 1

In this stage, I can use any nonparametric model to estimate the marginal density $f(y_t)$ and the invariant distribution $F(y_t)$. In this study, I use the nonparametric kernel estimators for $f(y_t)$ and $F(y_t)$ which are given by

$$\hat{f}(y_t) = \frac{1}{(T-1)h} \sum_{j=2}^{T} k(\frac{Y_j - y_t}{h}),$$

and

$$\hat{F}(y_t) = \frac{1}{(T-1)} \sum_{j=2}^{T} G(\frac{Y_j - y_t}{h})$$

where $G(y_t) = \int_{-\infty}^{y_t} k(v)dv$ and $k(v)$ is a second order Gaussian kernel function. h denotes the bandwidth which is selected based on the least square cross-validation method.

3.2.2. Second Step of Estimation of Copula-based Markov models of Order 1

After estimating the marginal distributions, I can estimate the copula density function $c(\hat{F}(y_{t-1}), \hat{F}(y_t))$ which characterizes the scale-free temporal dependence property. To ease notation, I define $u_t = (u_{0t}, u_{1t})$ where $u_{0t} = \hat{F}(y_{t-1})$ and

$u_{1t} = \hat{F}(y_t)$ for $t = 2, \dots, T$. In this paper, I propose to use an alternative nonparametric estimator which is the sieve series estimator for $c(\hat{F}(y_{t-1}), \hat{F}(y_t))$. In particular, this study uses finite-dimensional linear spaces spanned by the power functions as sieve spaces such that the sieve series estimator coincides with the Exponential Series Estimator (ESE) in Wu [43]. It turns out that I use the ESE for $c(\hat{F}(y_{t-1}), \hat{F}(y_t))$. The ESE can be naturally derived from the method of Maximum Entropy subject to a given set of moment constraints.

$$H = \int_{[0,1]^2} -c(u) \log c(u) du,$$

subject to the integration to unity

$$\int_{[0,1]^2} c(u) du = 1$$

and side conditions in terms of moments

$$\int_{[0,1]^2} g_i(u)\, c(u)\, du = \hat{\mu}_i, \qquad i \in M,$$

where $\hat{\mu}_i = n^{-1} \sum_{t=1}^{n} g_i(u_t)$, $du = du_1 du_2$ and $g_i(u)$ are a sequence of linearly independent polynomials defined on $[0,1]^2$. The estimated bivariate copula density takes the form

$$c(u; \hat{\lambda}) = \exp\left(-\sum_{i \in M} \hat{\lambda}_i g_i(u) - \hat{\lambda}_0\right)$$

where

$$\hat{\lambda}_0 = \log\left(\int_{[0,1]^d} \exp\left(-\sum_{i \in M} \hat{\lambda}_i g_i(u)\right) du\right)$$

and $M \equiv \{i : |i| > 0 \text{ and } i \leq m\}$.

Since analytical solutions for λ cannot be obtained, I need a nonlinear optimization, namely Newton's method, to solve for λ by iteratively updating

$$\hat{\lambda}_{t+1} = \hat{\lambda}_t - H^{-1} b$$

where the gradient

$$b_i = \hat{\mu}_i - \int g_i(u)c(u;\hat{\lambda}_t)du$$

and the Hessian matrix takes the form

$$H_{ij} = \int g_i(u)g_j(u)c(u;\hat{\lambda}_t)du.$$

Given the marginal density functions, the estimated bivariate density is then estimated by $\hat{g}(y_t|y_{t-1}) = \hat{f}(y_t)c(\hat{F}(y_{t-1}), \hat{F}(y_t); \hat{\lambda})$.

3.2.3. Asymptotic Properties

The goal of this study is to estimate the conditional density of Y_t given Y_{t-1} nonparametrically via $g^*(y_t|y_{t-1}) = f^*(y_t) \cdot c(F^*(y_{t-1}), F^*(y_t))$. Therefore the convergence rate of this conditional density estimator comes from the maximum of the convergence rate of the estimator of $c(F^*(y_{t-1}), F^*(y_t))$ and the convergence rate of the estimator of $f^*(y_t)$. Using bivariate ESE for the copula density, I start with the L_2 convergence rate $||c - \hat{c}||_2$ for weakly dependent observations.

I first estimate the marginal distribution functions which are used as the frames of copula density function. I can use existing nonparametric estimators such as the kernel , the spline and the series estimators. In this study, I employ the exponential series estimator of Barron and Sheu [1] for the estimation of marginals. Besides, the support of true copula density c_0 is the hypercube $[0,1]^2$. The basis functions ϕ_i are a sequence of linear independent polynomials. I assume, without loss of generality, ϕ_i are normalized Legendre polynomials.

Assumption 7 $\{Y_t\}_{t=1}^n$ *is a stationary first order Markov process generated by* $(F^*(\cdot), C(\cdot,\cdot))$ *where* F^* *is the true distribution function which is absolutely continuous with respect to Lebesgue measure.* $C(\cdot,\cdot)$ *is the true copula function for* (Y_{t-1}, Y_t) *with* *unknown parameter α^*. This copula function is absolutely continuous with respect to Lebesgue measure on $[0,1]^2$ and is neither the upper nor the lower bound of Frechet-Hoeffding boundaries.*

Note that under Assumption 7, $\{U_t : U_t \equiv F^*(Y_t)\}$ is a stationary Markov process of order 1 in which the joint distribution of U_{t-1} and U_t is given by $C(u_0, u_1)$ where $F^*(Y_{t-1}) = u_0$ and $F^*(Y_t) = u_1$.

Proposition 8 *If the copula density function is positive on $(0,1)^2$ and there are constants $\tilde{\lambda} \in [0,1)$, $0 < a, d < \infty$, a norm-like function $\Lambda(\cdot) \geq 1$ and a small set K such that $\int_0^1 \Lambda(u) \cdot c(U_{t-1}, u)du \leq \Lambda(U_{t-1}) - a \cdot [\Lambda(U_{t-1})]^{\tilde{\lambda}} + d \cdot I_k(U_{t-1})$ where $c(\cdot,\cdot)$ is the copula density associated with $C(\cdot,\cdot)$, then under Assumption 7, $\{Y_t\}$ is β-mixing with $\beta_t \leq \beta_0 \cdot (1+t)^{\tilde{\lambda}/(\tilde{\lambda}-1)}$ for some $\beta_0 > 0$.*

Given Assumption 7, Proposition 8 presents that a stationary first order Markov process $\{Y_t\}$ can be regarded as a β-mixing process with polynomial decay rate $\beta_t \leq \beta_0 \cdot (1+t)^{-\tilde{\lambda}/(-\tilde{\lambda}+1)}$. In this study, I assume $\{Y_t\}$ converges fast enough in the sense that $\tilde{\lambda} > 2/3$. Therefore the $\{Y_t\}_{t=1}^n$ satisfies Condition $A.1$ of Chen and Shen [7].

I next present the L_2 convergence rate of two-stage ESE of copula density for *i.i.d.* observations. The results here are based on Wu [43]. To make difference, I use X to represent the *i.i.d.* random variable, which is different from the weakly dependent random variable, Y.

Assumption 9 *The observed data $X_1 = [X_{11}, X_{21}]$, $X_2 = [X_{12}, X_{22}]$, $\ldots$, $X_n = [X_{1n}, X_{2n}]$ are i.i.d. continuous random samples with the joint density $f_0(x)$, the marginal densities f_j and the marginal distributions F_j.*

Assumption 10 *By Sklar's theorem, I have $f_0(x) = (\prod_{j=1}^{2} f_j(x_j))\, c_0(F_1(x_1), F_2(x_2))$. Let $p_0(x) = \log c_0(x)$ such that $|p_0(x)| < \infty$ for all x in the support of f_0 and $p_0(x)$*

is a member of a Sobolev space W_2^r in which $p_0^{(r-1)}(x)$ is absolutely continuous and $\int (p_0^{(r)}(x))^2 dx < \infty$ for $r > 2$. $r = r_1 + r_2$ for nonnegative integers r_1 and r_2 ,and $p_0^{(r)}(x) = \partial p_0^r(x)/\partial x_1^{r1} \cdot \partial x_2^{r_2}$.

Assumption 10 ensures that f_0 is bounded away from zero and infinity on its support. Let $m = (m_1, m_2)$ for nonnegative integers m_1 and m_2. Define $M = \{i : |i| > 0 \text{ and } i \leq m\}$. The proposed two-stage bivariate ESE of copula density takes the form

$$c_{\hat{\hat{\theta}}} = \exp(-\sum_{i \in M} \hat{\theta}_i\, \phi_i(\hat{F}_1(x_1), \hat{F}_2(x_2)) - \hat{\theta}_0)$$

where $\hat{F}_1(x_1)$ and $\hat{F}_2(x_2)$ are the exponential series estimators for $F_1(x_1)$ and $F_2(x_2)$, and the normalization term $\hat{\theta}_0 = \log[\int\int \exp(-\sum_{i\in M} \hat{\theta}_i\, \phi_i(\hat{F}_1(x_1), \hat{F}_2(x_2)))dx_1 dx_2] < \infty$. It follows that the log density $p_{\hat{\hat{\theta}}}(x) = -\sum_{i\in M} \hat{\theta}_i\, \phi_i(\hat{F}_1(x_1), \hat{F}_2(x_2)) - \hat{\theta}_0$.

Assumption 11 *$\prod_{j=1}^{2} m_j \to \infty$, and $(\prod_{j=1}^{2} m_j{}^3)/n \to 0$ as $n \to \infty$.*

Proposition 12 *If F_1,and F_2, are known, the ESE of copula density, $c_{\hat{\theta}} = \exp(-\sum_{i\in M} \hat{\theta}_i\, \phi_i(F_1, F_2) - \hat{\theta}_0)$, converges to c_0 in L_2 norm with the convergence rate*

$$||c_0 - c_{\hat{\theta}}||_2 = O_p(\prod_{j=1}^{2} m_j^{-r_j} + \sqrt{m_1 \cdot m_2/n}).$$

Now I establish the L_2 convergence rate of the two-stage ESE for copula density, $||c_0 - c_{\hat{\hat{\theta}}}||_2$.

Proposition 13 *The two-stage ESE for copula density $c_{\hat{\hat{\lambda}}}$ converges to c_0 in L_2 norm with the following rate*

$$||c_0 - c_{\hat{\hat{\theta}}}||_2 = O_p(\prod_{j=1}^{2} m_j^{-r_j} + \sqrt{m_1 \cdot m_2/n}).$$

As aforementioned, the approximation of copula density c_0 takes the form

$$c_\theta = \exp(-\sum_{i\in M}\theta_i\,\phi_i(F_1(x_1),F_2(x_2)) - \log[\int\int \exp(-\sum_{i\in M}\theta_i\,\phi_i(F_1(x_1),F_2(x_2)))dx_1dx_2]).$$

After taking logarithm, I have

$$\log c_\theta = -\sum_{i\in M}\theta_i\,\phi_i(F_1(x_1),F_2(x_2)) - \log[\int\int \exp(-\sum_{i\in M}\theta_i\,\phi_i(F_1(x_1),F_2(x_2)))dx_1dx_2].$$

Since $\log c_0 \in W_2^r$ for $r > 2$, it implies that based on Sobolev embedding theorem, a real-valued $\log c_0$ is a p-smooth function if $r > p+1$. Let $h_\theta(x) \equiv -\sum_{i\in M}\theta_i\,\phi_i(F_1(x_1),F_2(x_2))$. Therefore $h_\theta \in \Theta_n$ which is a linear subspace of the space of finite many polynomial basis functions h_θ. Θ_n can increase with n. In the literature of sieve estimation, Θ_n is called a finite-dimensional linear sieve space. Hence $\log c_\theta$ can be written as $h_\theta(x) - \log(\int \exp(h_\theta)dx)$. Furthermore, $f_0(x) = \log c_0(x)$ is given by $h_0(x) - \log(\int \exp(h_0(x))dx)$.

The log-likelihood evaluated at a single observation Z is given by

$$l(h_\theta, Z) = h_\theta(Z) - \log(\int_Z \exp(h_\theta(x))dx).$$

Stone [38] showed that $l(h_\theta, Z)$ is concave and $E_n(l(h_\theta, Z))$ is strictly concave in $h_\theta \in \Theta_n$. Therefore I next present that the bivariate ESE for copula density is a case of the series estimator for the concave extended linear model, e.q. Huang [19].

Assumption 14 *Assume $r > p+1$ such that h is p-smooth if it is p_1-times continuously differentiable on the support and $D^\alpha(h)$ satisfies a Hölder condition with exponent $p_2 \in (0,1]$ for all α with $[\alpha] = p_1$ where $p = p_1 + p_2$.*

Proposition 15 *Suppose Assumption 9, 10, 11 and 14 and Proposition 12 hold. Let $\rho_{2n} \equiv \inf_{h\in\Theta_n} ||h_0 - h||_2 = m_1^{-r_1}\cdot m_2^{-r_2}$. Then the series estimator $\hat{h}$ for h_θ exists* *uniquely with probability approaching one as $n \to \infty$ and*

$$||h_0 - \hat{h}||_2 = O_p(\sqrt{\frac{m_1 \cdot m_2}{n}} + m_1^{-r_1} \cdot m_2^{-r_2}).$$

The sieve estimator for the concave extended linear model is a special case of Theorem 3.2 of Chen [5] by taking $\delta_n \asymp \sqrt{(m_1 \cdot m_2)/n}$ and $||\pi_n\theta_0 - \theta_0|| \asymp \rho_{2n}$ where δ_n measures the complexity of the sieve space and $||\pi_n\theta_0 - \theta_0||$ measures the approximation rate. See Chen [5] for details.

In the following theorem, I show that $||c_0 - c_{\hat{\theta}}||_2$ keeps the same for both β-mixing data and *i.i.d.* observations.

Theorem 16 *Suppose Proposition 8, 12, 13, 15 hold.* $||c_0 - c_{\hat{\theta}}||_2 = O_p(\sqrt{\frac{m_1 \cdot m_2}{n}} + m_1^{-r_1} \cdot m_2^{-r_2})$ *for β-mixing data.*

In the end of this section, I derive the convergence rate of the estimator of conditional density $||g^*(y_t|y_{t-1}) - \hat{g}(y_t|y_{t-1})||_2$.

By Sklar's theorem, I have $\hat{g}(y_t|y_{t-1}) = \hat{f}(y_t) \cdot c(\hat{F}(y_{t-1}), \hat{F}(y_t); \hat{\theta})$. Using Theorem 1 in Barron and Sheu [1], I have $||f^* - \hat{f}||_2 = O_p(\sqrt{\frac{\tilde{m}_1}{n}} + \tilde{m}_1^{-\tilde{r}})$. It follows, using Theorem 16,

$$||g^*(y_t|y_{t-1}) - \hat{g}(y_t|y_{t-1})||_2 = O_p((\sqrt{\frac{m_1 \cdot m_2}{n}} + m_1^{-r_1} \cdot m_2^{-r_2}) + (\sqrt{\frac{\tilde{m}_1}{n}} + \tilde{m}_1^{-\tilde{r}})).$$

So $||g^*(y_t|y_{t-1}) - \hat{g}(y_t|y_{t-1})||_2 = O_p(\sqrt{\frac{m_1 \cdot m_2}{n}} + m_1^{-r_1} \cdot m_2^{-r_2})$ if $(\sqrt{\frac{m_1 \cdot m_2}{n}} + m_1^{-r_1} \cdot m_2^{-r_2}) \geq (\sqrt{\frac{\tilde{m}_1}{n}} + \tilde{m}_1^{-\tilde{r}})$ where $\tilde{m}_1$ is the dimension of polynomial family in the univariate case and assume $\log f^*(y_t)$ has $\tilde{r}$ square-integrable derivatives, see Barron and Sheu [1] for details. Note that $\tilde{m}_1$ does not necessarily equal to m_1.

3.2.4. Model Specification

Instead of including all feasible moment constraints at the same time, this study adopts a t-based updating process for the selection of individual moments and employs a data-driven method for the selection of the order of moments. In this study, I impose the restriction that the polynomials corresponding to the first and the second order of moments can not be replaced in the updating process since the first and the second order of moments are sufficient statistics of Gaussian distribution. The process starts with all polynomials corresponding to the second-order moment, $K_2 = \{g_i(u) : |i| = 2\}$ as well as the polynomials matching the first-order moment, $K_1 = \{g_i(u) : |i| = 1\}$. I delete the cross moments inside K_2 which go with the insignificant t statistics. The set of remaining polynomials is denoted by k_1 where $k_1 = K_1 \cup K_2 \backslash \{x\}$ and $\{x\}$ is the set of second-order cross moments without significant coefficients. The AIC and BIC based on k_1, called AIC_1 and BIC_1 are evaluated, too. In the next step, the polynomials associated with the third-order moment, i.e. $K_3 = \{g_i(u) : |i| = 3\}$, as well as the set of polynomials $\{k_1\}$ are used in the ESE of copula density. Also, the same deletion process of polynomials is taken to leave behind those having insignificant t statistics within K_3. Let k_2 denote the set of residual polynomials at this step where $k_2 = (\{k_1, g_i(u)\} : g_i(u) \subset K_3$ with significant coefficients). Repeat this procedure until the maximum order of moment, m, is reached, so are k_m, AIC_m and BIC_m. Finally, one needs to specify the optimal order of moments for the exponential series copula density estimator.In this study, a data-driven method, AIC, is used as the criterion to select the optimal degree of moments.

3.3. Monte Carlo Simulation

In this section, I conduct two Monte Carlo simulations to address the finite sample performance of the proposed estimator. In the first example, I discuss the in-sample estimation performance of the proposed estimator and then I discuss the one-step-ahead forecasting performance of the proposed estimator in the second example. In each example, I also evaluate the performance of the kernel density estimator for comparison. The underlying copula density functions used in our simulation follow a part of setup organized in the simulation of Chen, Wu, and Yi [8]. Four types of bivariate copula density functions including Gaussian, Frank, Clayton, and Gumbel are used as benchmarks. In our experiments, the dependence parameter for each type of copula is set such that their corresponding Kendall's τ are $\pm$ 0.8 and $\pm$ 0.5 for Gaussian copula, $\pm$ 0.833 and $\pm$ 0.5 for Frank copula, and 0.5, 0.714, 0.833 and 0.857 for Clayton and Gumbel copulas. It has been revealed that the performance of nonparametric copula estimation is not satisfactory when the marginal distribution has fat tails. In our study, the copula density function is estimated by the ESE so that the performance of the proposed estimator needs to be paid attention when the marginal distribution of the time series is fat tailed. To address this point, this study uses Student's t distribution with three degrees of freedom as the distribution function of the time series since Student's t distribution with three degrees of freedom is fat tailed relatively to the standard normal distribution. The marginal distributions, used as the arguments for the copula density, and the marginal densities are estimated by the kernel density estimator (KDE). While estimating the kernel densities in the simulation, the second order Gaussian kernel is used and the bandwidth is chosen by the least square cross-validation (LSCV).

In order to simulate a strictly stationary first-order Markov process from a bivariate copula with the marginal distribution F^*, this study follows the steps suggested in Chen, Wu, and Yi [8], which are outlined as follows.

1. First I generate a sequence of $i.i.d.$ random variables with uniform distribution, say $\{\Pi_t\}_{t=1}^n$.

2. Set $\Omega_1 = \Pi_1$ and $\Omega_t = C_{2|1}^{-1}[\Pi_t|\Omega_{t-1}]$. Note that $C_{2|1}[\cdot|u] \equiv \frac{\partial}{\partial u}C(u,\cdot)$ is the conditional distribution of the transformed variable $U_t \equiv F^*(Y_t)$, given $U_{t-1} = u$ and $C_{2|1}^{-1}[q|u]$ is the qth conditional quantile of U_t, given $U_{t-1} = u$.

3. Set $Y_t = F^{*-1}(\Omega_t)$ for $t = 1, \ldots, n$.

Our first example concerns the in-sample estimation of true conditional density. I generate two sequences of time series which have length of 300 and 500 respectively, but I delete the first 200 observations so the actual sample sizes are 100 and 300. Each experiment is repeated for 50 times. I use both the AIC and the BIC to select the actual moments to be incorporated in our estimator. For comparison, I also estimate the conditional densities using the KDE. The in-sample performance is gauged by the average of conditional integrated squared errors on y_{t-1} (ACISE), which is given by

$$\sum_{y_{t-1}\in Y_t} CISE(y_{t-1})/(T-1)$$

where $CISE(y_{t-1})$ is defined as

$$\int_{k_1\in y_t} (\hat{g}(Y_t = k_1|y_{t-1}) - g(Y_t = k_1|y_{t-1}))^2 dk_1.$$

The ACISE is evaluated on $[-10, 10]^2$ with the increment of 0.2 in each dimension. Because the performance of our method under the AIC and the BIC selection of moments are similar, I only report the results using the AIC. The results are reported in Table IV. Four panels of Table IV report the ACISE and the corresponding standard error for Gaussian, Frank, Clayton and Gumbel copulas.

Table IV. ACISE of In-sample Conditional Density Estimation

		KDE(N=100)	ESE(N=100)	KDE(N=300)	ESE(N=300)
Gaussian copula	τ = -0.8	0.0494	0.0254	0.0488	0.0249
		(0.0106)	(0.0023)	(0.0095)	(0.0030)
	τ = -0.5	0.0192	0.0183	0.0197	0.0190
		(0.0052)	(0.0011)	(0.0039)	(0.0013)
	$\tau = 0.5$	0.0187	0.0180	0.0186	0.0181
		(0.0055)	(0.0013)	(0.0042)	(0.0013)
	$\tau = 0.8$	0.0496	0.0269	0.0479	0.0253
		(0.0098)	(0.0029)	(0.0099)	(0.0035)
Frank copula	τ = -0.833	0.0680	0.0331	0.0661	0.0316
		(0.0129)	(0.0052)	(0.0092)	(0.0022)
	τ = -0.5	0.0192	0.0170	0.0167	0.0170
		(0.0082)	(0.0024)	(0.0036)	(0.0013)
	$\tau = 0.5$	0.0165	0.0160	0.0165	0.0174
		(0.0085)	(0.0009)	(0.0048)	(0.0011)
	$\tau = 0.833$	0.0659	0.0415	0.0662	0.0353
		(0.0165)	(0.0081)	(0.0091)	(0.0035)
Clayton copula	$\tau = 0.5$	0.0190	0.0184	0.0199	0.0187
		(0.0055)	(0.0012)	((0.0041)	(0.0019)
	$\tau = 0.714$	0.0570	0.0253	0.0470	0.0237
		(0.0107)	(0.0055)	(0.0180)	(0.0031)
	$\tau = 0.833$	0.0693	0.0472	0.0714	0.0398
		(0.0205)	(0.0108)	(0.0148)	(0.0050)
	$\tau = 0.857$	0.0710	0.0579	0.0763	0.0467
		(0.0172)	(0.0147)	(0.0095)	(0.0065)
Gumbel copula	$\tau = 0.5$	0.0176	0.0175	0.0204	0.0186
		(0.0058)	(0.0009)	(0.0052)	(0.0018)
	$\tau = 0.714$	0.0336	0.0229	0.0383	0.0227
		(0.0079)	(0.0032)	(0.0090)	(0.0029)
	$\tau = 0.833$	0.0565	0.0368	0.0607	0.0325
		(0.0156)	(0.0058)	(0.0144)	(0.0041)
	$\tau = 0.857$	0.0634	0.0430	0.0608	0.0369
		(0.0256)	(0.0078)	(0.0150)	(0.0044)

For Gaussian copula-based Markov models with 100 observations shown in the first panel of Table IV, the ESE performs better than the KDE in terms of the ACISE over different Kendall's τ. The same advantage of the ESE over the KDE remains the same for Frank, Clayton and Gumbel copulas when the sample size is 100. The average ratio of the ACISE between the ESE and the KDE across all Kendall's τ and all four copulas is 26%. When the sample size is 300, the ESE dominates the KDE over different Kendall's τ for Gaussian, Clayton and Gumbel copulas. However, for Frank copula-based Markov models, the performance of the ESE is close but slightly worse than the KDE when the Kendall's $\tau = \pm 0.5$. The corresponding average ACISE ratio between the ESE and the KDE improves to 29%. The performance of both estimators improve with the sample size except for the cases of Kendall's $\tau = \pm 0.5$ for Gaussian and Frank copulas. Regarding the sample variance, the ESE dominates the KDE in terms of small sample variance.

The second example concerns the out-of-sample forecasting performance of the ESE and the KDE. The measure of out-of-sample performance relies on the integrated square difference (ISD) between the forecasting conditional density and the true conditional density. The ISD and the corresponding standard error are reported in Table V. In the experiments, I use the last 10 observations for the use of forecast and use the remaining samples for estimation. For example, suppose the sample size is 100, I use the first 90 observations to estimate the conditional density and the remaining 10 observations to forecast the conditional density. The underlying copula densities, the sample size and the number of experiments remain the same as before. The procedures and criteria used in the first example are also employed in the step of estimation.

When sample size is 100, as shown in Table V, the ESE considerably outperforms the KDE in all copula-based models. The average ratio of the ISD between the ESE

Table V. ISD of One-step-ahead Conditional Density Forecast

		KDE(N=100)	ESE(N=100)	KDE(N=300)	ESE(N=300)
Gaussian copula	τ = -0.8	0.1457	0.0982	0.0993	0.0784
		(0.0814)	(0.0706)	(0.0784)	(0.0973)
	τ = -0.5	0.0180	0.0121	0.0113	0.0024
		(0.0191)	(0.0107)	(0.0056)	(0.0021)
	τ = 0.5	0.0200	0.0159	0.0165	0.0045
		(0.0109)	(0.0145)	(0.0100)	(0.0071)
	τ = 0.8	0.3314	0.2615	0.2636	0.2326
		(0.6004)	(0.2279)	(0.3926)	(0.2517)
Frank copula	τ = -0.833	0.4411	0.2670	0.4360	0.2237
		(0.2929)	(0.2083)	(0.3257)	(0.1974)
	τ = -0.5	0.0210	0.0038	0.0146	0.0031
		(0.0151)	(0.0037)	(0.0074)	(0.0024)
	τ = 0.5	0.0218	0.0055	0.0197	0.0036
		(0.0128)	(0.0063)	(0.0083)	(0.0026)
	τ = 0.833	0.4693	0.3463	0.4393	0.2901
		(0.6261)	(0.5542)	(0.4992)	(0.2959)
Clayton copula	τ = 0.5	0.0474	0.0264	0.0393	0.0242
		(0.1029)	(0.0250)	(0.0549)	(0.0235)
	τ = 0.714	0.2760	0.1115	0.2927	0.1087
		(0.4334)	(0.1331)	(0.4561)	(0.1847)
	τ = 0.833	0.7837	0.4340	0.8662	0.4863
		(0.7447)	(0.4231)	(0.7354)	(0.4310)
	τ = 0.857	1.0836	0.6779	1.1293	0.7759
		(1.0137)	(0.8724)	(0.9115)	(0.9140)
Gumbel copula	τ = 0.5	0.0272	0.0082	0.0219	0.0070
		(0.0189)	(0.0095)	(0.0218)	(0.0096)
	τ = 0.714	0.2014	0.1182	0.1572	0.0757
		(0.3513)	(0.1781)	(0.4077)	(0.2424)
	τ = 0.833	0.7231	0.4975	0.5830	0.2603
		(0.9712)	(0.7146)	(1.0631)	(0.4381)
	τ = 0.857	0.8058	0.5608	0.6259	0.3770
		(0.9612)	(0.7989)	(1.1809)	(0.5002)

and the KDE across all Kendall's τ and all four copulas is 43%. The general pattern of performance remains the same when the sample size goes to 300. The corresponding average ISD ratio between the ESE and the KDE improves to 51%. Besides, for Gaussian, Frank and Gumbel copulas, the forecasting performance improves with the sample size, shown in the first, second and the forth panels in Table V. However, the ESE does not dominate the KDE in all cases in terms of the sample variance.

3.4. Conclusion

Different from the semiparametric stationary Markov models by Chen and Fan [6], this paper studies a class of stationary Markov models of order 1 in the context of the two-stage ESE in which the copulas density function and the marginal distributions are estimated nonparametrically. Because of the β-mixing properties, I focus on the stationary Markov process of order 1 with continuous state space. The copula density functions in the second stage are approximated by the series estimate on sieve spaces. In this study, the sieve series estimator can be characterized as the Exponential Series Estimator under mild smoothness conditions. The ESE has information-theoretic interpretations and has no boundary bias. By using the β-mixing properties, I show that the copula density function approximated by the ESE for stationary first-order Markov models has the same convergent rate as the estimator of Wu [43]. I also establish the L_2 convergent rate of the proposed estimator.

I also examine the finite sample performance of the proposed estimator in two examples. In the first example, I discuss the in-sample estimation performance of the proposed estimator and then I discuss the one-step-ahead forecasting performance of the proposed estimator in the second example. The results show that our estimator outperform the kernel estimator in the conditional density estimation except for the

Frank copula-based Markov model. In addition, the proposed estimator considerably dominates the kernel estimator when used in the one-step-ahead forecast.

CHAPTER IV

SUMMARY

This dissertation applies exponential series methods to estimate the copula functions. I focus on the theoretical development of multivariate density estimator which includes the copula function that captures contemporary dependence among each variable. In addition, I study the estimation of a class of copula-based nonparametric stationary Markov models on the basis of the exponential series functions.

In the first essay I propose an alternative estimator for multivariate densities. This estimator can be characterized as a transformation based estimator. The first stage estimates each marginal density separately. In the second stage, the joint density of estimated marginal cumulative distribution functions (CDF) are approximated by the exponential series estimator. The final estimate is then obtained as the product of the marginal densities and the joint density estimated in the second stage. I derive the convergence rate in terms of the Kullback-Leibler Information Criterion (KLIC). Another contribution of this study is to incorporate a variable selection algorithm into a sequential updating process of moment selection to overcome the curse of dimensionality. The Monte Carlo studies show that the proposed estimator outperforms the kernel density estimator and the relative performance of our method with respect to the kernel method increases with the dimensionality of sample space. Besides, I also examine the performance of estimating tail distributions. My method dominates the empirical and the kernel density estimators except for the fat-tailed case. An empirical estimation of conditional copula density of stock returns is also provided.

In the second essay, the nonparametric estimation of copula-based stationary Markov Models is proposed. We extend the semiparametric model by Chen and Fan

[6] and study a class of time series models in the context of the two-stage ESE in which the copulas density function and the marginal distributions are estimated non-parametrically. In particular, we focus on the stationary Markov process of order 1 with continuous state space because it has the β-mixing property for the analysis of weakly dependent processes. The copula density functions for time series models are approximated by the series estimate on sieve spaces. In this study, a finite dimensional linear space spanned by a sequence of power functions is treated as the sieve space where the estimation space of the copula density function is based. This sieve series estimator can be characterized as the Exponential Series Estimator under mild smoothness conditions. To estimate the unknown copula density function, we propose a two stage estimator in which the first stage estimates each marginal density separately and in the second stage, the joint density of estimated marginal cumulative distribution functions (CDF) are approximated by the exponential series estimator. By using the β-mixing properties, we show that the copula density function approximated by the ESE for stationary first-order Markov models has the same convergence rate as the estimator of Wu [43]. It helps us establish the consistency of the proposed two-stage estimator. Extensive Monte Carlo studies show the proposed estimator outperforms kernel estimators in the one-step-ahead forecast of conditional density functions and in the estimation of conditional density functions except for the Frank copula-based Markov models.

As the application of exponential series approximation is relatively new, much work remains to be done. For example, Patton [32] applied parametric conditional copulas to model the time-varying dependence. Manner and Reznikova [25] provides an in-depth review of time-varying copula. However, most of studies model time-varying copulas in a parametric manner. Alternatively we could let the copula dependence parameter to be time-varying in a regime-switching manner and approximate the regime-specific copula by exponential series functions. We shall investigate this model in future work.

REFERENCES

[1] A.R. Barron, C.H. Sheu, Approximation of density functions by sequences of exponential families, Annals of Statistics 19 (1991) 1347–1369.

[2] J. Cadima, J.O. Cerdeira, M. Minhoto, Computational aspects of algorithms for variable selection in the context of principal components, Computational Statistics and Data Analysis 47 (2004) 225–236.

[3] J. Cadima, I.T. Jolliffe, Variable selection and the interpretation of principal subspaces, Journal of Agricultural, Biological, and Environmental Statistics 6 (2001) 62–79.

[4] Y. Chan, H. Li, Tail dependence for multivariate t-copula and its monotonicity, Insurance: Mathematics and Economics, 42 (2008) 763–770.

[5] X. Chen, Large sample sieve estimation of semi-nonparametric models, in: J.J. Heckman, E.E. Lleamer (Eds.), Handbook of Econometrics, North-Holland, Amsterdam, 2007, pp. 5549–5632.

[6] X. Chen, Y. Fan, Estimation of copula-based semiparametric time series models, Journal of Econometrics 130 (2006) 307–335.

[7] X. Chen, X. Shen, Sieve extremum estimates for weakly dependent data, Econometrica 66 (1998) 289–314.

[8] X. Chen, W.B. Wu, Y. Yi, Efficient estimation of copula-based semiparametric Markov models, Annals of Statistics 37 (2009) 4214–4253.

[9] C. Chui, X. Wu, Exponential series estimation of empirical copulas with application to financial returns, Advances in Econometrics 25 (2009) 263–290.

[10] B.R. Crain, Estimation of distributions using orthogonal expansion, Annals of Statistics 2 (1974) 454–463.

[11] W.F. Darsow, B. Nguyen, E.T. Olsen, Copulas and Markov processes, Illinois Journal of Mathematics 36 (1992) 600–642.

[12] R.A. DeVore, G.G. Lorentz, Constructive Approximation, Springer-Verlag, Berlin, 1993.

[13] P. Embrechts, A. McNeil, D. Straumann, Correlation and dependence properties in risk management: properties and pitfalls, in: M. Dempster (Eds.), Risk Management: Value at Risk and Beyond, Cambridge University Press, Cambridge, UK, 2002, pp. 176–223.

[14] J.D. Fermanian, O. Scaillet, Nonparametric estimation of copulas for time series, Journal of Risk 5 (2003) 25–54.

[15] I.J. Good, Maximum entropy for hypothesis formulation, especially for multidimensional contingency tables, Annals of Mathematical Statistics 34 (1963) 911–934.

[16] U. Grenander, Abstract Inference, Wiley Series, New York, 1981.

[17] P. Hall, N. Neumeyer, Estimating a bivariate density when there are extra data on one or both components, Biometrika 93 (2006) 439–450.

[18] J.Z. Huang, Projection estimation in multiple regression with application to functional ANOVA models, Annals of Statistics 26 (1998) 242–272.

[19] J.Z. Huang, Concave extended linear modeling: a theoretical synthesis, Statistica Sinica 11 (2001) 173–197.

[20] S. Jarner, G.O. Roberts, Polynomial convergence rates of Markov chains, Annals of Applied Probability 12 (2002) 224–247.

[21] E.E. Jaynes, Information theory and statistical mechanics, Physical Review 106 (1957) 620–630.

[22] H. Joe, Multivariate Models and Dependence Concepts, Chapman and Hall/CRC, London, UK, 1997.

[23] C. Kooperberg, C.J. Stone, A study of logspline density estimation, Computational Statistics and Data Analysis 12 (1991) 327–347.

[24] Q. Li, J. Racine, Nonparametric Econometrics: Theory and Practice, Princeton University Press, Princeton, New Jersey, 2007.

[25] H. Manner, O. Reznikova, A survey on time-varying copulas: specification, simulations and application, manuscript, Maastricht University, 2010.

[26] P. Marsh, Goodness of fit tests via exponential series density estimation, Computational Statistics and Data Analysis 51 (2007) 2428–2441.

[27] G.P. McCabe, Principal variables, Technometrics 26 (1984) 137–144.

[28] R.B. Nelsen, An Introduction to Copulas, Springer-Verlag, New York, 2006.

[29] J. Neyman, Smooth test for goodness of fit, Scandinavian Aktuarial 20 (1937) 149–199.

[30] W.K. Newey, J.L. Powell, Instrumental variable estimation of nonparametric models, Econometrica 71 (2003) 1565–1578.

[31] A.J. Patton, Modelling asymmetric exchange rate dependence, International Economic Review 47 (2006a) 527–556.

[32] A.J. Patton, Estimation of multivariate models for time series of possibly different lengths, Journal of Applied Econometrics 21 (2006b) 147–173.

[33] P.M. Robinson, Nonparametric estimator for time series, Journal of Time Series Analysis 4 (1983) 185–207.

[34] D. Ruppert, D.B.H. Cline, Bias reduction in kernel density estimation by smoothed empirical transformations, Annals of Statistics 22 (1994) 185–210.

[35] X. Shen, W. Wong, Convergence rate of sieve estimates, Annals of Statistics 22 (1994) 580–615.

[36] A. Sklar, Fonctions de repartition *à* n dimensions et leurs marges, Publicians Institute Statistics University Paris 8 (1959) 229–231.

[37] C.J. Stone, Optimal global rates of convergence for nonparametric regression, Annals of Statistics 10 (1982) 1040–1053.

[38] C.J. Stone, Large sample inference for log-spline models, Annals of Statistics 18 (1990) 717–741.

[39] M.P. Wand, J.S. Marron, D. Ruppert, Transformations in density estimation, Journal of the American Statistical Association 86 (1991) 343–361.

[40] M.P. Wand, M.C. Jones, Comparison of smoothing parameterizations in bivariate kernel density estimation, Journal of the American Statistical Association 88 (1993) 520–528.

[41] H. White, J. Wooldridge, Some results on sieve estimation with dependent observations, in: W.A. Barnett, J. Powell, G. Tauchen (Eds.), Nonparametric and Semiparametric Methods in Econometrics and Statistics, Cambridge University Press, Cambridge, UK, 1991, pp. 459–493.

[42] X. Wu, Calculation of maximum entropy densities with application to income distribution, Journal of Econometrics 115 (2003) 347–354.

[43] X. Wu, Exponential series estimator of multivariate density, Journal of Econometrics 156 (2010) 354–366.

[44] L. Yang, J.S. Marron, Iterated transformation-kernel density estimation, Journal of the American Statistical Association 94 (1999) 580–589.

[45] A. Zellner, R.A. Highfield, Calculation of maximum entropy distribution and approximation of marginal posterior distributions, Journal of Econometrics 37 (1988) 195–209.

APPENDIX A

TECHNICAL PROOFS

Proposition 4. Given Assumption 1, 2 and 3, the proof is straightforward by using Proposition 1 and 2 and Theorem 4 in Wu [43]. ■

Proposition 5. From Sklar's theorem, it's known that $p_{\hat{\lambda}}(x) = (\prod_{j=1}^{d} \hat{f}_j(X_j))\, c_{\hat{\lambda}}(F_1(X_1), \ldots, F_j(X_j))$. Using Theorem 1 in Barron and Sheu [1], we have $D(f_j||\hat{f}_j) = O_p(\tilde{m}_j^{-2s_j} + \tilde{m}_j/n)$ so that $\prod_{j=1}^{d} \hat{f}_j$ converges to $\prod_{j=1}^{d} f_j$ in the sense of KLIC at rate $O_p(\max_j(\tilde{m}_j^{-2s_j} + \tilde{m}_j/n))$. It follows, by using Proposition 4,

$$D(p_0||p_{\hat{\lambda}}) = O_p(\max_j(\tilde{m}_j^{-2s_j} + \tilde{m}_j/n) + \prod_{j=1}^{d} m_j^{-2r_j} + \prod_{j=1}^{d} m_j/n)$$

■

Theorem 6. The Kullback-Leibler distance between $p_{\hat{\hat{\lambda}}}$ and p_0 is denoted by

$$D(p_0||p_{\hat{\hat{\lambda}}}) = \int p_0(x) \log \frac{p_0(x)}{p_{\hat{\hat{\lambda}}}(x)}\, dx = \int p_0(x) \log \frac{p_0(x)}{p_{\hat{\lambda}}(x)} \frac{p_{\hat{\lambda}}(x)}{p_{\hat{\hat{\lambda}}}(x)}\, dx$$

$$= \int p_0(x) \log \frac{p_0(x)}{p_{\hat{\lambda}}(x)}\, dx + \int p_0(x) \log \frac{p_{\hat{\lambda}}(x)}{p_{\hat{\hat{\lambda}}}(x)}\, dx.$$

From Proposition 5, we have

$$D(p_0||p_{\hat{\lambda}}) = \int p_0(x) \log \frac{p_0(x)}{p_{\hat{\lambda}}(x)}\, dx = O_p(\max_j(\tilde{m}_j^{-2s_j} + \tilde{m}_j/n) + \prod_{j=1}^{d} m_j^{-2r_j} + \prod_{j=1}^{d} m_j/n)$$

Besides,

$$\int p_0(x) \log \frac{p_{\hat{\lambda}}(x)}{p_{\hat{\hat{\lambda}}}(x)}\, dx = \int [p_0(x) + p_{\hat{\lambda}}(x) - p_{\hat{\lambda}}(x)] \log \frac{p_{\hat{\lambda}}(x)}{p_{\hat{\hat{\lambda}}}(x)}\, dx$$

$$= \int p_{\hat{\lambda}}(x) \log \frac{p_{\hat{\lambda}}(x)}{p_{\hat{\hat{\lambda}}}(x)}\, dx + \int (p_0(x) - p_{\hat{\lambda}}(x)) \log \frac{p_{\hat{\lambda}}(x)}{p_{\hat{\hat{\lambda}}}(x)}\, dx$$

For the first part, $p_{\hat{\hat{\lambda}}}(x) - p_{\hat{\lambda}}(x)$ has the same convergence rate as the marginal distribution $\hat{F} - F$ in terms of KLIC. In addition, it's known that, as the number of parameters m_j goes to infinity, the convergence rate of marginal distribution is the optimal rate of convergence for nonparametric distribution estimation and is faster than that of density function, e.q. Stone [38]. Therefore, we have $\int p_{\hat{\lambda}}(x) \log \frac{p_{\hat{\lambda}}(x)}{p_{\hat{\hat{\lambda}}}(x)} \, dx = o_p(\max_j(\tilde{m}_j^{-2s_j} + \tilde{m}_j/n)$.

Wu [43] showed that the KLIC convergence implies convergence in terms of the integrated squared error. Therefore, we have

$$\int |p_0(x) - p_{\hat{\lambda}}(x)| \log \frac{p_{\hat{\lambda}}(x)}{p_{\hat{\hat{\lambda}}}(x)} \, dx \to 0$$

given $\log \frac{p_{\hat{\lambda}}(x)}{p_{\hat{\hat{\lambda}}}(x)}$ is bounded.

It implies

$$\int (p_0(x) - p_{\hat{\lambda}}(x)) \log \frac{p_{\hat{\lambda}}(x)}{p_{\hat{\hat{\lambda}}}(x)} \, dx \to 0$$

given $\log \frac{p_{\hat{\lambda}}(x)}{p_{\hat{\hat{\lambda}}}(x)}$ is bounded.

So $\int (p_0(x) - p_{\hat{\lambda}}(x)) \log \frac{p_{\hat{\lambda}}(x)}{p_{\hat{\hat{\lambda}}}(x)} \, dx$ is $o_p(1)$. ■

Proposition 8. Based on Assumption 7 and conditions in this proposition, the Markov process $\{U_t\}$ is ergodic with the polynomial decay rate by Theorem 3.6 in Jarner and Robert [20]. This and the definition of β-mixing for a stationary Markov process imply $\{U_t\}$ is β-mixing with $\beta_t \leq \beta_0 \cdot (1+t)^{\tilde{\lambda}/(\tilde{\lambda}-1)}$ for some $\beta_0 > 0$. Since F^* is continuous and by the definition of β-mixing, $\{Y_t\}$ is β-mixing with $\beta_t \leq \beta_0 \cdot (1+t)^{\tilde{\lambda}/(\tilde{\lambda}-1)}$. ■

Proposition 12. Given Assumption 9, 10 and 11, Wu [43] shows that $c_{\hat{\theta}}$ converges to c_θ in the sense of KLIC with the rate

$$D(c_0||c_{\hat{\theta}}) = O_p(\prod_{j=1}^{2} m_j^{-2r_j} + \prod_{j=1}^{2} m_j/n)$$

It follows that

$$||c_0 - c_{\hat{\theta}}||_2 = O_p(\prod_{j=1}^{2} m_j^{-r_j} + \sqrt{m_1 \cdot m_2/n})$$

by Corollary 5 of Wu [43]. ■

Proposition 13. From the triangle inequality, we know that

$$||c_0 - c_{\hat{\hat{\theta}}}||_2 \leq ||c_0 - c_{\hat{\theta}}||_2 + ||c_{\hat{\theta}} - c_{\hat{\hat{\theta}}}||_2$$

From Proposition 12, we have

$$||c_0 - c_{\hat{\theta}}||_2 = O_p(\prod_{j=1}^{2} m_j^{-r_j} + \sqrt{m_1 \cdot m_2/n})$$

Besides,

$$||c_{\hat{\theta}} - c_{\hat{\hat{\theta}}}||_2 = O(||F - \hat{F}||)$$

It's known that, as the number of parameters m_j goes to infinity, the convergence rate of marginal distribution is the optimal rate of convergence for nonparametric distribution estimation and is faster than that of density function, e.q. Stone [38]. Therefore, we have $||c_{\hat{\theta}} - c_{\hat{\hat{\theta}}}||_2 = o_p(\prod_{j=1}^{2} m_j^{-r_j} + \sqrt{m_1 \cdot m_2/n})$.
Finally, we complete the proof. ■

Proposition 15. Θ_n is a linear subspace of the space of finite many polynomial basis functions on the support $[0,1]^2$ and $E(l(h,Z))$ is strictly concave in h. These two conditions imply $h_0 \equiv argmax_{h \in W_2^r} E(l(h,Z))$ exists unequally and $||h_0||_\infty \leq k_0 < \infty$ can be supported by Lemmma $A.2$ of Wu [43]. Therefore, Condition $A.1$ in Huang [19] is satisfied.

Let $h_1, h_2 \in W_2^r$ be a pair of bounded functions, $E(l(h_1 + \tau(h_2 - h_1)), Z)$ is twice continuously differentiable with respect to $\tau \in [0,1]$ and $\frac{\partial^2}{\partial \tau^2} E(l(h_1 + \tau(h_2 - h_1)), Z)$

takes the form

$$E[\frac{\partial^2}{\partial \tau^2} l(h_1 + \tau(h_2 - h_1))] = E(-Var(h_2 - h_1))$$

Using the fact that p_0 is bounded away from zero and infinity, we know

$$Var(h_2 - h_1) \asymp \int (h_2 - h_1)^2 dZ$$

uniformly in $\tau \in [0,1]$. Therefore, Condition $A.2$ in Huang [19] holds.

Given the above conditions, we can define $\bar{h} \equiv argmax_{h_\theta \in \Theta_n} E(l(h_\theta, Z))$. For any pair of functions $g_1, g_2 \in \Theta_n$, $l(g_1 + \tau(g_2 - g_1))$ is twice continuously differentiable with respect to $\tau \in [0,1]$ and for $g \in \Theta_n$

$$\frac{\partial}{\partial \tau} l(\bar{h} + \tau g)|_{\tau=0} = E_n(\frac{\partial}{\partial \tau} l(\bar{h} + \tau g)|_{\tau=0})$$

and

$$E_n(\frac{\partial}{\partial \tau} l(\bar{h} + \tau g)|_{\tau=0}) = E_n(g) - E(g).$$

Therefore,

$$\frac{\frac{\partial}{\partial \tau} l(\bar{h} + \tau g)|_{\tau=0}}{||g||} = \frac{(E_n - E)g}{||g||}.$$

Condition $A.4(1)$ in Huang [19] follows fron Lemma 11 of Huang [18].

Since, for $g_1, g_2 \in \Theta_n$,

$$\frac{\partial^2}{\partial \tau^2} l(g_1 + \tau(g_2 - g_1)) = \frac{\partial^2}{\partial \tau^2} E(l(g_1 + \tau(g_2 - g_1))).$$

Condition $A.4(2)$ in Huang [19] follows from Condition $A.2$. Finally, following Lemma 1 of Huang [18] and Theorem 4.2.6 of Devore and Lorentz [12], we get

$$A_n \rho_{2n} \le const.(m_1^{-r_1+1} \cdot m_2^{-r_2+1}) \to 0$$

$$\frac{A_n^2 (m_1 \cdot m_2)}{n} \le const. \frac{(m_1^3 \cdot m_2^3)}{n} \to 0$$

as $n \to \infty, m_1 \to \infty, m_2 \to \infty, r_1 \geq 1, r_2 \geq 1, r > 2$. This completes the proof. ■

Theorem 16. We have shown that the bivariate ESE for copula density function is the sieve estimator for the concave extended linear model in Proposition 15. So we can demonstrate that this ESE for copula density function is a special case of the series estimator in Theorem 3.2 of Chen [5]. Following Theorem 1 of Chen and Shen [7], it yields that the convergence rate of the ESE for copula density function remains the same for both β-mixing data and *i.i.d.* data. ■

APPENDIX B

COEFFICIENTS FOR NORMAL MIXTURES

The coefficients for the bivariate normal mixtures can be obtained in Table 1 of Wand and Jones []40. A trivariate normal random variable is given by $N(\mu, \sigma, \rho)$ where $\mu = (\mu_1, \mu_2, \mu_3), \sigma = (\sigma_1, \sigma_2, \sigma_3), \rho = (\rho_{12}, \rho_{13}, \rho_{23})$. The coefficients for the trivariate normal mixtures used in the simulation are as follows.

1. uncorrelated normal: $N((0,0,0),(\frac{1}{2},\frac{1}{\sqrt{2}},1),(0,0,0))$

2. correlated normal: $N((0,0,0),(1,1,1),(\frac{3}{10},\frac{5}{10},\frac{7}{10}))$

3. skewed: $\frac{1}{5}N((0,0,0),(1,1,1),(0,0,0)) + \frac{1}{5}N((\frac{1}{2},\frac{1}{2},\frac{1}{2}),(\frac{2}{3},\frac{2}{3},\frac{2}{3}),(0,0,0))$
   $+ \frac{3}{5}N((\frac{12}{13},\frac{12}{13},\frac{12}{13}),(\frac{5}{9},\frac{5}{9},\frac{5}{9}),(0,0,0))$

4. kurtotic: $\frac{2}{3}N((0,0,0),(1,\sqrt{2},2),(\frac{1}{2},\frac{1}{2},\frac{1}{2})) + \frac{1}{3}N((0,0,0),(\frac{2}{3},\frac{\sqrt{2}}{3},\frac{1}{3}),(-\frac{1}{2},-\frac{1}{2},\frac{1}{2}))$

5. bimodal I: $\frac{1}{2}N((-1,0,0),(\frac{2}{3},\frac{2}{3},\frac{2}{3}),(0,0,0)) + \frac{1}{2}N((1,0,0),(\frac{2}{3},\frac{2}{3},\frac{2}{3}),(0,0,0))$

6. bimodal II: $\frac{1}{2}N((-\frac{3}{2},0,0),(\frac{1}{4},1,1),(0,0,0)) + \frac{1}{2}N((\frac{3}{2},0,0)(\frac{1}{4},1,1),(0,0,0))$.

VITA

Meng-Shiuh Chang received his Bachelor of Science degree in Mechanical Engineering from National ChiaoTung University in Taiwan in 2001. He received the Master of Business Administration degree in Industrial Economics at Tamkang University in 2004 and the Master of Social Science in Econometrics and Economics at York University in 2005. He entered the Economics program at Texas A&M University in July 2006, transferred to the Agricultural Economics program in July 2007 and received his Doctor of Philosophy degree in August 2011. His research interests include Econometrics and Financial Economics.

Dr. Chang may be reached at Department of Public Finance and Taxation of Southwestern University of Finance and Economics, No.55 Guanghuacun Street, Chengdu 610074, China. His email is mslibretto@hotmail.com.