TECHNICAL REPORT

CTR/70/07 October 2007

Nonparametric multivariate conditional distribution and quantile regression

K. Yu, X. Sun and G. Mitra

CARISMA, School of Information Systems, Computing and Mathematics, Brunel University

Brunel UNIVERSITY WEST LONDON

School of Information Systems, Computing and Mathematics
Brunel University, Uxbridge, Middlesex, UB8 3PH

CARISMA

Nonparametric multivariate conditional distribution and quantile regression

Keming Yu, Xiaochen Sun and Mitra

Brunel University, West London, UK

Summary. In nonparametric multivariate regression analysis, one usually seeks methods to reduce the dimensionality of the regression function to bypass the difficulty caused by the curse of dimensionality. We study nonparametric estimation of multivariate conditional distribution and quantile regression via local univariate quadratic estimation of partial derivatives of bivariate copulas. Without restricting the form of underlying regression function or using dimensional reduction, we show that a d-dimensional multivariate conditional distribution and quantile regression could be estimated by $\frac{d(d+1)}{2}$ times of univariate smoothers. The asymptotic bias and variance as well as smoothing parameter selection method are derived. Simulations show that the method works quite well. The techniques are illustrated by application to exchange rate data.

Keywords: Conditional distribution; Conditional quantiles; Copula; High-dimension; Local quadratic regression; nonparametric estimation; Partial derivative; Semiparametric estimation.

1 Multivariate nonparametric regression

The ever-growing number of high-dimensional, super-large databases requires effective analysis techniques to deal with a large number of independent variables in the model. These models include, but are not limited to, the following regression quantities:

1) Conditional mean $M(x) = E(Y|X = x)$;

2) Conditional distribution function $F(y|x) = Pr[Y \leq y|X = x]$;

3) Conditional quantiles, or quantile regression $Q_p(x)$ with $Pr[Y \leq Q_p(x)|X = x] = p$, $0 < p < 1$.

Where both X and Y are random variables.

While estimating a conditional distribution is an important feature of many statistical problems, the problem of characterising the dependency between random variables at a given quantile is also an important issue (Koenker, 2005), especially if the distributions of the variables involved are fat tails or/and skew as is standard with many important economic variables such as financial returns as well as biological variables such as body weight. In addition, extreme quantiles of a conditional distribution are useful to be used for measuring value-at-risk (VaR) in finance, flooding quantiles and risk level in environment and public health.

One familiar estimation approach begins by assuming that, for example, $M(x)$ belongs to a known, finite dimensional, parametric family of functions. That is, $M(x) = G(x, q)$ for almost every x in the support of X, where G is a known function and q is a finite-dimensional vector. If $G(x, q)$ is the true conditional mean function for some q and certain regularity conditions are satisfied, then q and G can be estimated consistently (by least squares among other ways) with $n^{-1/2}$ rates of convergence in probability. The estimation results can be highly misleading, however, if $G(x, q)$ is misspecified, meaning that there is no q such that $M(x) = G(x, q)$ for almost every x.

The possibility of misspecification can be largely eliminated by estimating nonparametrically (Härdle and Linton, 1994). In recent years there has been increased interest in using nonparametric methods for high-dimensional data analysis. For example, Nadaraya-Watson type estimators for $M(x)$ and $F(y|x)$: $\hat{M}(x) = \frac{\sum_{i=1}^{n} K_h(x-X_i) Y_i}{\sum_{i=1}^{n} K_h(x-X_i)}$, and $\hat{F}(y|x) = \frac{\sum_{i=1}^{n} K_h(x-X_i) I(Y_i \leq y)}{\sum_{i=1}^{n} K_h(x-X_i)}$, where K is a kernel function with $h = (h_1, ..., h_d)$ and $K_h(x - X_i) = K_h(x_1 - X_{i1}, ..., x_d - X_{id})$. When X is multidimensional, however, many nonparametric methods do not perform well in this case. The problem of rapidly increasing variance for increasing dimensionality is sometimes referred to as the "curse of dimensionality", a phenomenon that the data points are sparse in the high dimensional sample space even if the sample size is large (Silverman, 1986).

Hence avoiding the curse of dimensionality with multivariate regression is expected. One well-known model is the semiparametric single-index modelling $M(x) = G(\boldsymbol{\beta}' x)$, where $\boldsymbol{\beta}$ is an unknown $d \times 1$ vector, G is an unknown function and the quantity $\boldsymbol{\beta}' x$ is called an index. The index aggregates the dimensions of x, thereby achieving dimension reduction. The other class of models for multivariate regression are additive models (Buja, Hastie and Tibshirani, 1989; Hastie and Tibshirani, 1990), which have been popular among statisticians and data analysts in multivariate nonparametric regression. In such a model, $M(x) = \sum_{k=1}^{d} f_k(x^k)$, where x^k is the k'th component of x and the functions f_k are unknown. But the restriction to additivity does not allow every conceivable type of response surface to be accurately described. Instead, few theoretical results but not practical implementation (Yu, Park and Mammen, 2008; Horowitz and Mammen, 2007; Horowitz, 2001) considered $M(x) = G(\sum_{k=1}^{d} f_k(x^k))$, with G known or unknown. In terms of estimation of $Q_p(x)$ and $F(y|x)$, Horowitz and Lee (2005, Yu and Lu (2004) and De Gooijer and Zerom (2003) considered additive models for multivariate quantile regression, and Hall and Yao (2005) proposed a dimension reduction method to estimate not the distribution of $Y|X$, but that of $Y|\theta X$, where the unit vector θ is selected so that the approximation is optimal under a least-squares criterion.

The proposed method in this paper seeks direct estimation multivariate distribution of $Y|X$ and quantile regression by simple univariate smoothers. The paper is organised as follows. Section 2 describes copula based representation of multivariate conditional distribution and quantile regression. Section 3 carries out the first simulation using semiparametric models. Section 4 proposes local univariate quadratic regression for multivariate analysis. Section 5 outlines some theoretic results and bandwidth selection as well as further simulation. Section 6 applies the method for a real data analysis. Section 7 presents a brief discussion of the method.

2 Copula for multivariate conditional distribution

2.1 Copula

The concept of copulas was introduced by Sklar (1959), and has for a long time been recognized as a powerful tool for modeling dependence between random variables.

A copula function is a multivariate distribution function with unknown univariate marginal distributions. For example, a **bivariate copula** is a function $\mathcal{C} : [0,1] \times [0,1] \to [0,1]$ such that:

1. For $(u, v) \in [0,1]^2$,

$$C(u,0) = C(0,v) = 0, \;\; C(u,1) = u, \;\; and \;\; C(1,v) = v.$$

2. $(u_1, v_1, u_2, v_2) \in [0,1]^4$, $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0.$$

Sklar's theorem states that every multivariate distribution F with marginals F_1, F_2, ..., F_d can be written as

$$F(x_1, ..., x_d) = C(F_1(x_1), F_2(x_2), ..., F_d(x_d))$$

for some copula C. In particular, let X and Y be two random variables with joint distribution F and continuous marginal distributions. Then there exists a unique copula C satisfying

$$F(x, y) = C(F_X(x), F_Y(y)).$$

One copula that we will refer to frequently in this paper is the Gaussian copula. Let $\Phi_\alpha(.,.)$ be the distribution function of the bivariate normal distribution with mean zero, variances 1, and the correlation coefficient α. Then the Gaussian copula is given by

$$C(u, v; \alpha) = \Phi_\alpha(\Phi^{-1}(u), \Phi^{-1}(v)),$$

where $0 \leq u, v \leq 1$ and $\Phi(.)$ is the distribution function of a standard normal random variable. By Sklar's theorem, for any two marginal distribution functions $F_X(x)$ and $F_Y(.)$, the distribution defined as

$$H(x, y) = C(F_X(x), F_Y(y); \alpha) = \Phi_\alpha(\Phi^{-1}(F_X(x)), \Phi^{-1}(F_Y(y)))$$

is a bivariate distribution function whose marginals are $F_X(x)$ and $F_Y(.)$ respectively. Hence Sklar's theorem allows one to construct bivariate distributions with non-Normal marginal distributions and the Gaussian copula. Different copulas typically exhibit different dependence properties. Joe (1997), Nelsen (1999) and Bouyé and Salmon (2002) contain excellent discussions of various dependence measures and of dependence properties of numerous parametric copulas.

Another popular copula family is the Archimedean copula, which includes the Clayton copula

$$C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}},$$

and Frank copula

$$C(u, v; \delta) = -\frac{1}{\delta}\log\Big(1 + \frac{(\exp(-\delta u) - 1)(\exp(-\delta u) - 1)}{\exp(-\delta) - 1}\Big).$$

Where $\theta > 0$ and $\delta \neq 0$.

2.2 Copula for conditional distribution

Let $\frac{\partial C(u,v)}{\partial u}$ be the partial derivative of $C(u, v)$ over u, the conditional distribution $F(y|x)$ can be given by

$$F(y|x) = \frac{\partial C(F_X(x), F_Y(y))}{\partial F_X(x)}, \tag{1}$$

for some bivariate copula $C(u, v)$. For example, for the Gaussian copula we have

$$\frac{\partial C(u, v; \alpha)}{\partial u} = \Phi\Big(\frac{\Phi^{-1}(v) - \Phi^{-1}(u)}{\sqrt{1 - \alpha^2}}\Big),$$

and for the Frank copula, we have

$$\frac{\partial C(u, v; \delta)}{\partial u} = \exp(-\delta u)((1 - \exp(-\delta))(1 - \exp(-\delta v))^{-1} - (1 - \exp(-\delta u)))^{-1}.$$

When $X = x$ with $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $F(y|x)$ can be specified by three partial derivatives of bivariate copula and marginal distributions. In fact, there is a copula $C_{y,x_1|x_2}$ such that

$$F(y|x) = \frac{\partial C_{y,x_1|x_2}\Big(F(x_1|x_2),\ F(y|x_2)\Big)}{\partial F(x_1|x_2)},$$

and there are copulas C_{y,x_2} and C_{x_1,x_2} respectively such that $F(y|x_2) = \frac{\partial C_{y,x_2}(F(x_2), F(y))}{\partial F(x_2)}$ and $F(x_1|x_2) = \frac{\partial C_{x_1,x_2}(F(x_2), F(x_1))}{\partial F(x_2)}$.

Similarly, when dimension $d = 3$ with $x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, $F(y|x)$ can be specified as

$$F(y|x) = \frac{\partial C_{y,\, x_1|x_2,x_3}(F(y|x_2,x_3),\ F(x_1|x_2,x_3)}{\partial F(x_1|x_2,x_3)},$$

where

$$\begin{aligned} F(x_1|x_2,x_3) &= \frac{\partial C_{x_1,x_2|x_3}(F(x_1|x_3),\ F(x_2|x_3))}{\partial F(x_2|x_3)}, \\ F(y|x_2,x_3) &= \frac{\partial C_{y,x_2|x_3}(F(y|x_3),\ F(x_2|x_3))}{\partial F(x_2|x_3)}, \end{aligned}$$

and

$$\begin{aligned} F(x_1|x_3) &= \frac{\partial C_{x_1,x_3}(F(x_1),\ F(x_3))}{\partial F(x_3)}, \\ F(x_2|x_3) &= \frac{\partial C_{x_2,x_3}(F(x_2),\ F(x_3))}{\partial F(x_3)}, \\ F(y|x_3) &= \frac{\partial C_{y,x_3}(F(y),\ F(x_3))}{\partial F(x_3)}, \end{aligned}$$

which involves in total 6= 1+2=3 partial derivatives of bivariate copula.

In general, for a $d-$dimensional $x = \begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_d \end{pmatrix}$, the conditional distribution $F(y|x)$ may be estimated in terms of $\frac{d(d+1)}{2}$ times partial derivatives of bivariate copulae.

2.3 Copula for conditional quantiles

Given value of x, the pth $(0 < p < 1)$ conditional quantile denoted by $Q_p(x)$ of Y is the solution y of

$$F(y|x) = p.$$

Once we have $F(y|x)$ being specified by the partial derivative of a bivariate copula, we can obtain an estimate of $Q_p(x)$. For example, when $X = x$ with $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, $Q_p(x) = F_Y^{-1}\Big((p^{-\frac{\theta}{1+\theta}} - 1)u^{-\theta} + 1)^{-\frac{1}{\theta}}|x_2\Big)$, under the Clayton copula if we employ the bivariate Clayton copula $C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}$. Similarly, $Q_p(x) = F_Y^{-1}\Big(\Phi(\rho\Phi^{-1}(u) + \sqrt{1-\rho^2}\Phi^{-1}(p))|x_2\Big)$ under Gaussian copula and $Q_p(x) = F_Y^{-1}\Big(-\delta^{-1}\log(1 + (e^{-\delta} - 1)(1 + e^{-\delta u}(p^{-1} - 1)))|x_2\Big)$ under Frank copula. Where $u = F(x_1|x_2)$.

3 Simulation studies

Given variables (X, Y) and their observations, if those bivariate copulas belong to some known parametric families, then combining copula parameter estimation and nonparametric marginal distribution estimation results in a semiparametric version of multivariate conditional distribution. Before we introduce a fully nonparametric method in Section 4, we carry out some simulation studies for this semiparametric method. That is, we employ parametric copula estimation such as maximum likelihood estimation and empirical marginal distribution estimation to illustrate the efficiency of the proposed method.

Consider a model

$$y = \frac{2}{7} - \frac{11}{7}x_1^2 + \frac{1}{7}x_2 + \epsilon,$$

with $(x_1, x_2) \sim N_2(0, \Sigma)$,

$$\Sigma = \begin{pmatrix} 10 & 3 \\ 3 & 2 \end{pmatrix}$$

and $\epsilon \sim N(0, \sqrt{(32/7)})$. Figures 1 and 2 display the true conditional distributions $F(y|x_1, x_2 = 0)$ and $F(y|x_1 = 0, x_2)$. We compare the proposed estimate with the Nadaraya-Watson estimator $\hat{F}(y|x) = \frac{\sum_{i=1}^n K_h(x-X_i)\, I(Y_i \le y)}{\sum_{i=1}^n K_h(x-X_i)}$ and use least-squares cross-validation bandwidth selection.

put Figures 1 and 2 here.

Following Section 2.2, we use Clayton Copula to model the joint distribution of (x_1, x_2) then obtain an estimate of $F(x_1|x_2)$. The parameter α in Clayton copula is typically around $\hat{\alpha} = 18.05$ with each simulated sample of 200. Similarly, we use Gaussian copula for the estimation of $F(y|x_2)$ and Frank copula for the estimation of $F(y|x_1, x_2)$. For each simulated sample of $n = 200$, the performance of each estimator is evaluated in terms of Mean Absolute Derivation Error (MADE)

$$MADE = \frac{\sum_{i=1}^n |\hat{F}(y_i|x_i) - F(y_i|x_i)|}{n}.$$

Figure 3 compares the performance of the proposed estimator and the Nadaraya-Watson estimator.

put Figure 3.

Also based on the simulated model above and following Section 2.3, we continue the estimation of a 5% conditional quantile (see Figure 4) and plot the performance of the semiparametric method and Nadaraya-Watson based quantile estimator in Figure 5.

put Figures 4 and 5.

Clearly, the proposed semiparametric method for multivariate $F(y|x)$ and $Q_p(x)$ is simple and promising for the simulated model. We also carried out more simulations by increasing the dimensions of x to 3 and 4 and the conclusions are consistently almost same.

4 Local univariate quadratic regression

4.1 Nonparametric regression for the partial derivatives of copula

Section 3 has demonstrated that the copula based univariate approach for multivariate conditional distribution and quantile regression is practicable. However, the choice of copula family is crucial, for example, for pricing the contract in finance. There is an uncertainty in choosing the right copula family (Embrechts, Lindskog and McNeil, 2003).

If the true copula is assumed to belong to a parametric family, then maximum likelihood method would be applied for copula estimation and the implementation of this method can be found in copula package in R. Like parametric regression, however, parametric based copula estimation may be misspecifiction, so nonparametric copula estimation may be a good alternative.

A few copulas such as Gaussian copula, student's t copula and Archimedean copula family (Nelsen, 1999) have been explored for application. Aas (2004) made a survey of four copulas and his comparison shows that the choice of copula can affect the joint distribution very seriously.

Chen and Huang (2005) and Fermanian and Scaillet (2003) have developed nonparametric estimation of copula. Basically, the idea behind nonparametric estimation is to estimate the joint copula via kernel smoothing. Typically, let K be a symmetric kernel supported on [-1, 1], $G(t) = \int_{-\infty}^{t} K(x)dx$ be the derived distribution function from K, then for the observation $\{X_i, Y_i\}_{i=1}^{n}$ of (X, Y), the kernel estimation of copula $C(u,v)$ is given by

$$\frac{1}{n}\sum_{i=1}^{n} G\Big(\frac{u-\hat{F}_X(X_i)}{h}\Big) G\Big(\frac{v-\hat{F}_Y(Y_i)}{h}\Big).$$

Where $\hat{F}_X(.)$ and $\hat{F}_Y(.)$ are the kernel estimation of marginal distributions $F_X(.)$ and $F_Y(.)$ respectively, and h is the bandwidth. The estimated copula may be useful for estimating copula quantile regression. However, it is impractical to do multivariate kernel distribution (density) estimation in the real situations such as when there are multiple factors to affect the value of option. Multivariate kernel density estimation usually suffer from the curse of dimensionality.

Moreover, in terms of conditional distribution and quantile estimation via copula, we don't need the whole copula but the partial derivative of the copula.

Note that

$$C(u,v) = C(u)C(v|u),$$

and $C(u) = u$, we have

$$\frac{\partial}{\partial u} C(u,v) = C(v|u) + u\,\frac{\partial}{\partial u} C(v|u).$$

Also note that

$$C(v|u) = E[I(V \le v)|U = u],$$

for each $(u,v) \in [0,1]^2$, we discuss the estimation of conditional copula $C(v|u)$ and its derivative $\frac{\partial}{\partial u} C(v|u)$ in order to estimate copula quantile regression.

Rewrite

$$\frac{\partial}{\partial u} C(u,v) = (1,u,0)\left(\begin{array}{c} C(v|u) \\ \frac{\partial}{\partial u} C(v|u) \\ \frac{1}{2} C^{(2,0)}(v|u) \end{array}\right),$$

we advise the local quadratic kernel estimation based on the minimization of

$$\sum_{i=1}^{n} \Big(I(V_i \le v) - a - b(U_i - u) - c(U_i - u)^2 \Big)^2 K_h(u - U_i).$$

This minimization is the standard local polynomial kernel smoothing estimation of conditional expectation $E[I(V \le v)|U = u]$ (Wand and Jones, 1995; Fan and Gijbels, 1996). Let $\hat{a}$ and $\hat{b}$ be the estimators of a and b of the minimization, then according to local quadratic regression theory, $\hat{a}$ and $\hat{b}$ estimate $C(v|u)$ and $\frac{\partial}{\partial u} C(v|u)$ respectively. We need to obtain $\frac{d(d+1)}{2}$ times $\hat{a}$ and b repeatedly for a $d-$dimensional covariate X. When $\hat{a}$ and $\hat{b}$ estimate the final $C(v|u)$, we may be able to define the solution v of the following equation

$$p = \hat{C}(v|u) = \hat{a}\,+\,u\,\hat{b}$$

as the estimation of copula quantile and denote it by $Q_p(u)$. Then the relationship between $Q_p(x)$ and $Q_p(u)$ is given by $Q_p(x) = F^{-1}_{Y|X_2}(Q_p(u)$ with $u = F(x_1|x_2)$.

4.2 Local univariate quadratic regression

We focus on the minimization of

$$\sum_{i=1}^{n}\Big(I(V_i \le v) - a - b - c(U_i - u)^2(U_i - u)\Big)^2 K_h(u - U_i).$$

Let $K_i = K_h(\frac{u-U_i}{h})$, $t_{n,j} = \sum_{i=1}^{n}(X_i - x)^j K_i\, I(V_i \le v)$, $j = 0, 1, 2$ and $s_{n,r} = \sum_{i=1}^{n}(X_i - x)^r K_i$, $r = 0, 1, 2, 3, 4$, from

$$\begin{aligned} t_{n,0} &= a\, s_{n,0} + b\, s_{n,1} + c\, s_{n,2} \\ t_{n,1} &= a\, s_{n,1} + b\, s_{n,2} + c\, s_{n,3} \\ t_{n,2} &= a\, s_{n,2} + b\, s_{n,3} + c\, s_{n,4} \end{aligned}$$

$\hat{a}(u, v)$ is given by

$$\hat{a}(u, v) = \frac{\sum_{i=1}^{n} w_i I(V_i \le v)}{\sum_{i=1}^{n} w_i},$$

with

$$w_i(u) = K_i\Big[1 - \hat{\theta}_1(U_i - u) + \hat{\theta}_2(U_i - u)^2\Big]$$

with

$$\hat{\theta}_1 = \frac{(s_{n,1}s_{n,4} - s_{n,2}s_{n,3})}{s_{n,2}s_{n,4} - s_{n,3}^2}, \quad \hat{\theta}_2 = \frac{(s_{n,1}s_{n,3} - s_{n,2}^2)}{s_{n,2}s_{n,4} - s_{n,3}^2}.$$

The weight function may be negative then result in no CDF for $C(v|u)$. A simple solution is to modify w_i to exclude negative values. (We then redefine the weights as follows

$$w_i(u) = \begin{cases} 0 & \text{if } \hat{\theta}_1(U_i - u) - \hat{\theta}_2(U_i - u)^2 > 1 \\ K_i\Big(1 - \hat{\theta}_1(U_i - u) + \hat{\theta}_2(U_i - u)^2\Big) & \text{if } \hat{\theta}_1(U_i - u) - \hat{\theta}_2(U_i - u)^2 \le 1 \end{cases}$$

This modification is asymptotically negligible, but constrains $\hat{a}(u, v)$ to be a valid CDF.

Similarly, we have

$$\hat{b}(u, v) = \frac{s_{n,3}\Big(t_{n,0} - \hat{a}s_{n,0}\Big)}{s_{n,2}\Big(t_{n,1} - \hat{a}s_{n,1}\Big)}.$$

Hence the estimator of $\frac{\partial}{\partial u}C(u, v)$ is given by

$$\widehat{\frac{\partial}{\partial u}C(u, v)} = \hat{a}(u, v) + u\, \hat{b}(u, v).$$

4.3 Steps of estimation

Take $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ as an example. Let $\hat{F}_{X_1}(x_1)$, $\hat{F}_{X_2}(x_2)$ and $\hat{F}_Y(y)$ be the marginal estimators. Then, the steps of estimation are given by

(1) Letting $U = \hat{F}_{X_1}(X_1)$, $V = \hat{F}_{X_2}(X_2)$;

(2) Obtain $\hat{a}(u, v)$ and $\hat{b}(u, v)$ by local quadratic regression;

(3) Estimate $F(x_1|x_2)$ by $\hat{F}(x_1|x_2) = \Big(\hat{a}(u, v) + u\,\hat{b}(u, v)\Big)_{U=\hat{F}_{X_1}(x_1), V=\hat{F}_{X_2}(x_2)}$;

(4) Repeat the steps (1)–(3) for $\hat{F}(y|x_2)$ and $\hat{F}(y|x)$;

(5) Obtain the value v from the last $\hat{a}(u, v) + u\,\hat{b}(u, v) = p$ and denoted it by $Q_p(u)$;

(5) Estimate $Q_p(x)$ by $\hat{Q}_p(x) = F^{-1}_{Y|X_2}(Q_p(u))$ with $u = F(x_1|x_2)$.

5 Theoretic results and bandwidth selection

It is clear from the discussion in Sections 2, 3 and 4 that the estimation of partial derivative of a copula plays a key role in the proposed method. While local polynomial has been suggested for our nonparametric estimation we aim at outlying some theoretic properties of the estimation and associated bandwidth selection.

Let vector $\boldsymbol{\beta} = (a,\, b,\, c)^T \equiv \begin{pmatrix} C(v|u) \\ C^{(1,0)}(v|u) \\ \frac{1}{2}\,C^{(2,0)}(v|u) \end{pmatrix}$. Under local quadratic fitting of Section 4, let $\hat{\boldsymbol{\beta}} = (\hat{a},\, \hat{b},\, \hat{c})^T$ be the estimator of $\boldsymbol{\beta}$. To derive the theoretic property of $\hat{\boldsymbol{\beta}}$, we write

$$\hat{\boldsymbol{\beta}} = (M^T W M)^{-1} M^T I(V),$$

where the matrix

$$M = \begin{pmatrix} 1 & U_1 - u & (U_1 - u)^2 \\ 1 & U_2 - u & (U_2 - u)^2 \\ \cdot & \cdot & \\ \cdot & \cdot & \\ \cdot & \cdot & \\ 1 & U_n - u & (U_n - u)^2 \end{pmatrix},$$

$W = \text{diag}\Big(K_h(U_1 - u), ..., K_h(U_n - u)\Big)$, vector $I(V) = (I(V_1 \le v), ..., I(V_n \le v))^T$, and vector $C = \begin{pmatrix} C(v|U_1) \\ . \\ . \\ . \\ C(v|U_n) \end{pmatrix}$.

It can be easily seen that conditional on $(U_1, ..., U_n)$ we have

$$E\{\hat{\boldsymbol{\beta}}\} = (M^T W M)^{-1} M^T W\, EI(V) = (M^T W M)^{-1} M^T W C,$$

and

$$Var\{\hat{\boldsymbol{\beta}}\} = (M^T W M)^{-1} M^T W (M^T W M)^{-1}.$$

Following Section 4.2, let $\mathbf{S_n} = \begin{pmatrix} s_{n,0} & s_{n,1} & s_{n,2} \\ s_{n,1} & s_{n,2} & s_{n,3} \\ s_{n,2} & s_{n,3} & s_{n,4} \end{pmatrix}$, $\mathbf{t_n} = (t_{n,0}, t_{n,1}, t_{n,2})^T$, then (Fan and Gijbels, 1995)

$$\hat{\boldsymbol{\beta}} = \mathrm{diag}(1, h^{-1}, h^{-2}) \mathbf{S_n}^{-1} \mathbf{t_n}.$$

For the convenience of notation, we denote

$$\mu_j = \int u^j K(u) du, \;\; \nu_j = \int u^j K^2(u) du,$$

and

$$\mathbf{S} = \begin{pmatrix} \mu_0 & \mu_1 & \mu_2 \\ \mu_1 & \mu_2 & \mu_3 \\ \mu_2 & \mu_3 & \mu_4 \end{pmatrix}, \;\; \tilde{\mathbf{S}} = \begin{pmatrix} \nu_0 & \nu_1 & \nu_2 \\ \nu_1 & \nu_2 & \nu_3 \\ \nu_2 & \nu_3 & \nu_4 \end{pmatrix}, \;\; \mu = \begin{pmatrix} \mu_3 \\ \mu_4 \\ \mu_5 \end{pmatrix}.$$

Assumptions:

(i) The kernel K is a continuous density function having bounded support;
(ii) $C^{(3,0)}(v|.)$ exists and is continuous in a neighborhood of u;
(iii) $C^{(3,0)}(v|.) \neq 0$ and the kernel K is symmetric.

Under local quadratic fitting and symmetric kernel,

$$Bias(\hat{b}) = \frac{1}{6} \frac{\mu_4}{\mu_2}\, C^{(3,0)}(v|u)\, h^2 + o_P(h^2),$$

and

$$Bias(\hat{a}) = \frac{1}{6} \frac{\mu_4^2 - \mu_2 \mu_6}{\mu_4 - \mu_2^2}\, C^{(4,0)}(v|u)\, h^4 + o_P(h^4),$$

therefore,

$$Bias\Big(\widehat{\frac{\partial}{\partial u} C(v|u)}\Big) = \frac{1}{24} \frac{\mu_4^2 - \mu_2 \mu_6}{\mu_4 - \mu_2^2}\, C^{(4,0)}(v|u)\, h^4 + u\, \frac{1}{6} \frac{\mu_4}{\mu_2}\, C^{(3,0)}(v|u)\, h^2 + o_P(h^2) + o_P(h^4).$$

Note that

$$\mathbf{S}^{-1} = \begin{pmatrix} \frac{\mu_4}{\mu_4-\mu_2^2} & 0 & -\frac{\mu_2}{\mu_4-\mu_2^2} \\ 0 & \frac{1}{\mu_2} & 0 \\ -\frac{\mu_2}{\mu_4-\mu_2^2} & 0 & \frac{1}{\mu_4-\mu_2^2} \end{pmatrix},$$

so

$$E\widehat{\frac{\partial}{\partial u}}C(v|u) - \frac{\partial}{\partial u}C(v|u) = \frac{\mu_4}{\mu_2}\frac{C^{(3,0)}(v|u)\,u}{3!}h^2\Big\{+o_P(1)\Big\}.$$

Similarly, from (see Chapter 3 of Fan and Gijbels, 1996)

$$Var\,(\hat{\boldsymbol{\beta}}) = \text{diag}(1,h,h^2)^{-1}\,S^{-1}S^*S^{-1}\,\text{diag}(1,h,h^2)^{-1}\frac{C(v|u)(1-C(v|u))}{nh}\{1+o_P(1)\},$$

and

$$\mathbf{S^{-1}S^*S^{-1}} = \begin{pmatrix} \frac{\mu_4^2-2\mu_4\mu_2\nu_2+\mu_2^2\nu_4}{(\mu_4-\mu_2^2)^2} & \frac{\mu_4\nu_1-\mu_2\nu_3}{\mu_2(\mu_4-\mu_2^2)} & \frac{\mu_2^2\nu_2-\mu_2\nu_4}{(\mu_4-\mu_2^2)^2} \\ \frac{\mu_4\nu_1-\mu_2\nu_3}{\mu_4-\mu_2^2} & \frac{\nu_2}{\mu_2} & \frac{\nu_3-\mu_2\nu_1}{\mu_4-\mu_2^2} \\ \frac{\mu_4\nu_2-\mu_2\mu_4-\mu_2\nu_4+\mu_2^2\nu_2}{(\mu_4-\mu_2^2)^2} & \frac{\nu_3-\mu_2\nu_1}{\mu_2(\mu_4-\mu_2^2)} & \frac{\mu_2^2+\nu_4-2\mu_2\nu_2}{(\mu_4-\mu_2^2)^2} \end{pmatrix},$$

we have

$$Var\Big(\widehat{\frac{\partial}{\partial u}}C(v|u)\Big) = (1,u,0)^T\,Var\,(\hat{\boldsymbol{\beta}})\,(1,u,0)$$

$$= \frac{C(v|u)(1-C(v|u))}{nh}\{1+o_P(1)\}\,(\sigma_1+2\sigma-2\,u\,h+\sigma_3\,u^2\,h^2),$$

where $\sigma_1 = \frac{\mu_4^2-2\mu_4\mu_2\nu_2+\mu_2^2\nu_4}{(\mu_4-\mu_2^2)^2}$, $\sigma_2 = \frac{\mu_4\nu_1-\mu_2\nu_3}{\mu_2(\mu_4-\mu_2^2)}$ and $\sigma_3 = \frac{\nu_2}{\mu_2}$.

Finally, from

$$\hat{Q}_p(u) - Q_p(u) \simeq -\frac{\widehat{\frac{\partial}{\partial u}}C(v|u) - \frac{\partial}{\partial u}C(v|u)}{\frac{\partial C(v|u)}{\partial v}},$$

we have

$$Bias(\hat{Q}_p(u)) = -\frac{1}{\frac{\partial C(v|u)}{\partial v}}\Big(Bias(\hat{a}) + u\,Bias(\hat{b})\Big)|_{v=Q_p(u)},$$

and

$$Var(\hat{Q}_p(u)) = \frac{1}{\Big(\frac{\partial C(v|u)}{\partial v}\Big)^2}Var\Big(\frac{\partial}{\partial u}C(v|u)\Big)|_{v=Q_p(u)}.$$

5.1 Bandwidth Selection

We need to address the bandwidth selection for local quadratic estimation of $\boldsymbol{\beta}$, so $Q_p(u)$.

Usually, bandwidth selection is based on minimization of (asymptotic) mean square error. Here we should use the asymptotic mean square error of β for bandwidth selection. From the Chapter 3 of Yu (1997), and note that $f(u) = 1$

$$\int_{-1}^{1} E(\hat{a} - a)^2 dv \approx \frac{R(K)V}{nh} + \frac{V_1 h^4}{4} + O(n^{-6/5}),$$

where $V = \int_{-1}^{1} F(v|u)(1 - F(v|u))dv, V_1 = \int_{-1}^{1} F^{2,0}(v|u)^2 dv$. We can use plug-in method explored by Hansen (2004).

6 Further simulation studies and real data analysis

6.1 Simulation studies

We carry out some simulation studies for the estimation of the partial derivative of a copula by local quadratic regression explored in Section 5 and implemented in R (http://stat.ethz.ch/R-manual/R-patched/library/stats/html/loess.html).

We consider three bivariate copulas: Gaussian, Clayton and Frank copula. Figure 6, for example, displays the estimated partial derivative of an Clayton based on the model of Section 3.

We plot the bias and mean square errors in the box-chat to illustrate the accuracy.

put Figure 6 here.

6.2 Real data analysis

The data was obtained from PACIFIC Exchange Rate Service at Sauder School of Business, UBC.

We use weekly FX data from 1/1/1993 to 31/12/2005. The four different rate we choose are: European Euros to U.S Dollars, British Pounds to U.S.Dollars and Japanese Yen to U.S.Dollars. There are 676 observations for each FX rate. Figure 7 display the 3-pair exchange rates.

put Figures 7 and 8 here.

Let x_1, x_2 and x_3 be the returns of the exchange rates of $US\$/EURO$, $US\$/JPY$ and $US\$/POUND$ respectively. Figure 8 display the returns.

Figures 9 and 10 illustrate the two-dimensional (x_1 conditional on x_2 and x_3) and one-dimensional (x_3 given x_2) conditional quantiles of returns respectively.

put Figures 9 and 10 here.

In particular for $US\$/POUND$ against $US\$/JPY$, we see high dependence in the lower tail for the pairs of rates and reasonable low dependence in the upper tail.

7 Discussion

Development of new methodologies for multivariate data analysis has exploded over the last decade in an effort to mine interesting information from the data. The original additive models approximate a regression function by additive univariate functions, which overcome successfully the difficulty caused by the curse of dimensionality. But since they ignore any interactions between regressors they are not sufficiently flexible to cope with the case where interactions are of concern. Sklar theorem in Copula studies can help to describe the dependence between high-dimensional random variables via marginal distributions. We study nonparametric conditional distribution and quantile regression estimation by kernel weighted local quadratic fitting. We show that a d-dimensional multivariate $F(y|x)$ and $Q_p(x)$ can be estimated by $d(d+1)/2$ times of local univariate smoothers.

Additive models are usually fitted by backfitting algorithm whereas the proposed models employ directly mean regression computing without involving any iteration. When d is moderate big, the computing for the proposed method is not a problem though $d(d+1)/2$ is not a small number for very large d.

8 Reference

Aas, K. (2004) Modelling the dependence structure of financial assets: A survey of four copulas. *Research Report SAMBA/22/04*, Norwegian Computing Center.

Bouy*é* E. and Salmon, M. (2002) Dynamic copula quantile regressions and tail area dynamic dependence in forex markets. Preprint, Warwick Business School.

Buja, A. , Hastie, T. J. and Tibshirani, R. (1989) Linear smoothers and additive models. *Ann. Statist.*, **17**, 453–555.

Chen, S. X. and Huang, T. (2005) Nonparametric estimation of copula functions for dependence modeling. *Canadian Journal of Statistics*, **35**, 265–282.

De Gooijer, J. G. and Zerom, D. (2003) On additive conditional quantiles with high-dimensional covariates. *J. Am. Statist. Ass.*, **98**, 135–46.

Embrechts, P. , Lindskog, F. and McNeil, A. J. (2003) Modelling dependence with copulas and applications to risk management. In *Handbook of Heavy-tailed Distributions in Finance* (ed. S.T. Rachev), pp. 329-384. Amsterdam: Elsevier/North-Holland.

Fan, J. and Gijbels, I. (1996) *Local Polynomial Modelling and its Applications.* London: Chapman & Hall.

Fermanian, J. D. and Scaillet, O. (2003) Nonparametric estimation of copulas for time series. *J. Risk*, **5**, 25–54.

Hansen, B. E. (2004) Nonparametric estimation of smooth conditional distributions. *Technical Report*, University of Wisconsin.

Härdle W, Linton O (1994) Applied nonparametric methods. In *Handbook of econometrics*(eds R. F. Engle and D. L. McFadden), **4**, 2295–2339. Amsterdam: North Holland.

Hastie, T. J. and Tibshirani, R. J. (1990) *Generalized additive models.* London: Chapman and Hall.

Hall, P. and Yao, Q. (2005) Approximating conditional distribution functions using dimension reduction. *Ann.Statist.*, **33**, 1404-1421.

Horowitz, J. L. (2001) The Bootstrap in econometrics. In *Handbook of Econometrics* (eds J. Heckman and E. Leamer), **5**, 3159-3228 New York: Elsevier Science.

Horowitz, J. L. and Lee, S. (2005) Nonparametric estimation of an additive quantile regression model. *J. Am. Statist. Ass.*, **100**, 1238–1249.

Horowitz, J. L. and Mammen, E. (2007) Rate-optimal estimation for a general class of nonparametric regression models with known link functions. *Ann. Statist.*, **35**,

Joe, H. (1997) *Multivariate Models and Dependence Concept.* New York: Chapman & Hall.

Koenker, R. (2005) *Quantile Regression (Econometric Society Monograph Series).* New York: Cambridge University Press.

Nelsen, R. B. (1999) *An Introduction to Copulas*, New York: Springer.

Sklar, A. (1959) Fonction de répartition *á* n dimensions et leurs marges. *Publications de L'Institute de Statistique*, Université de Paris 8, 229-231.

Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis.*

London: Chapman and Hall.

Wand, M. P. and Jones, M. C. (1995)*Kernel Smoothing.* London: Chapman and Hall.

Yu, K. (1997). *Smoothing Regression Quantile*, Unpublished PhD dissertation, The Open University, UK.

Yu, K., Lu, Z. (2004) Local linear additive quantile regression. *Scand. J. Statist.*, **31**, 333–346.

Yu, K., Park, B. and Mammen, E. (2008) Smooth backfitting in generalized additive models, *Ann. Statist.*, **36**, 228–260.

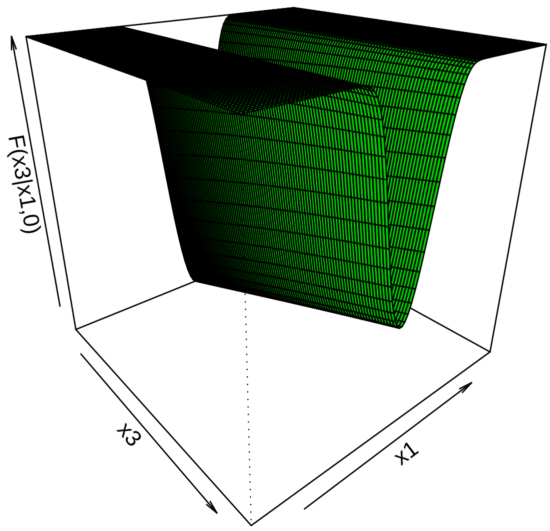

Fig. 1. True conditional distributions $F(y|x_1, x_2 = 0)$

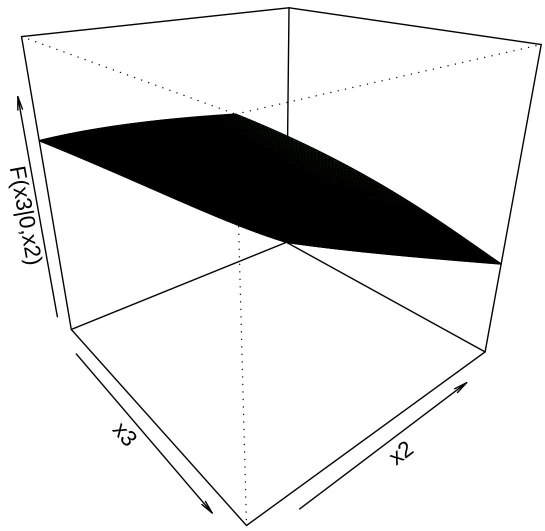

Fig. 2. True conditional distributions $F(y|x_1 = 0, x_2)$

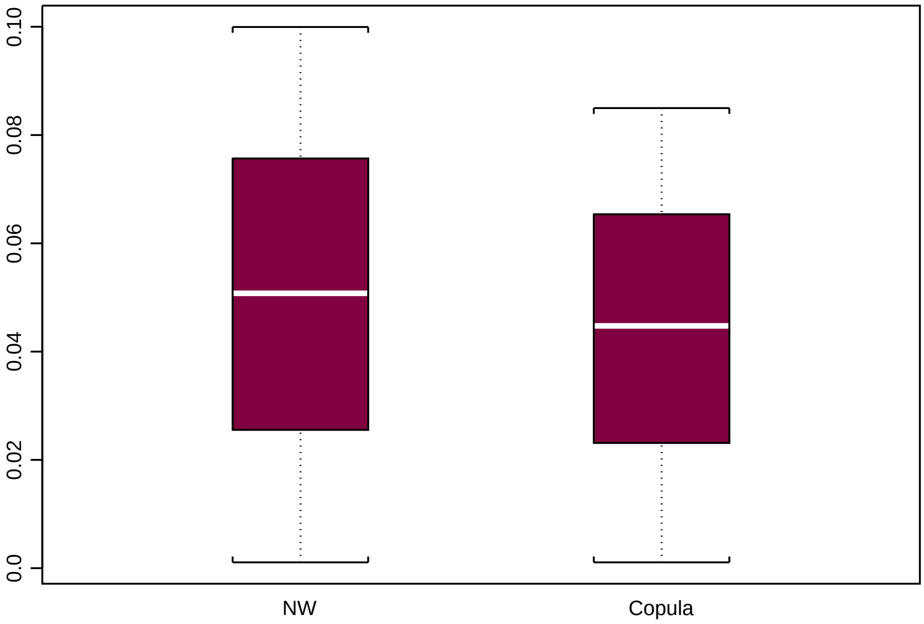

Fig. 3. boxplots of MADEs for estimating CDF

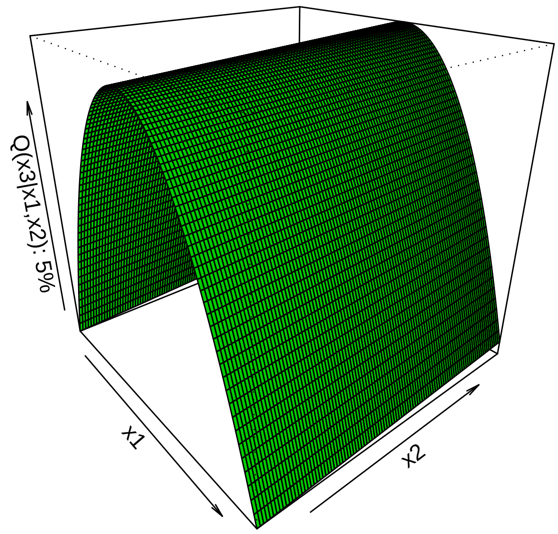

Fig. 4. True 5% conditional quantile

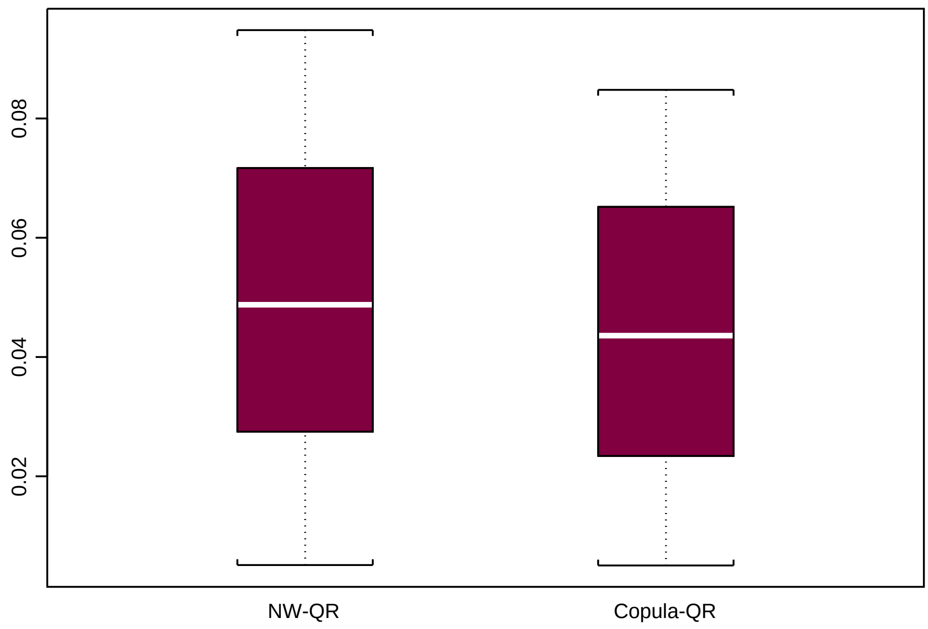

Fig. 5. boxplots of MADEs for estimating $Q_{0.05}(x)$

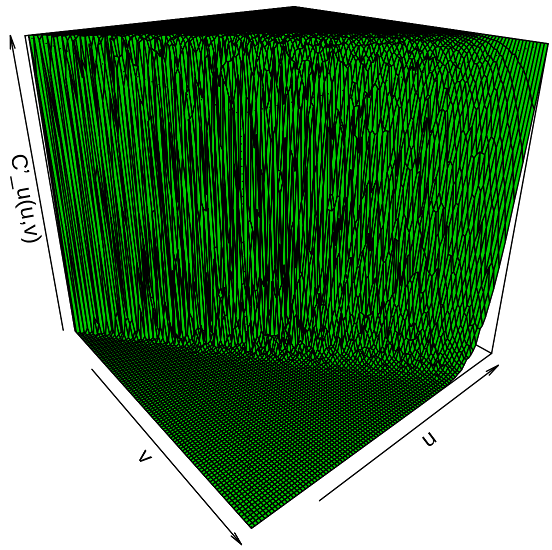

Fig. 6. Estimated partial derivative of Clayton copula

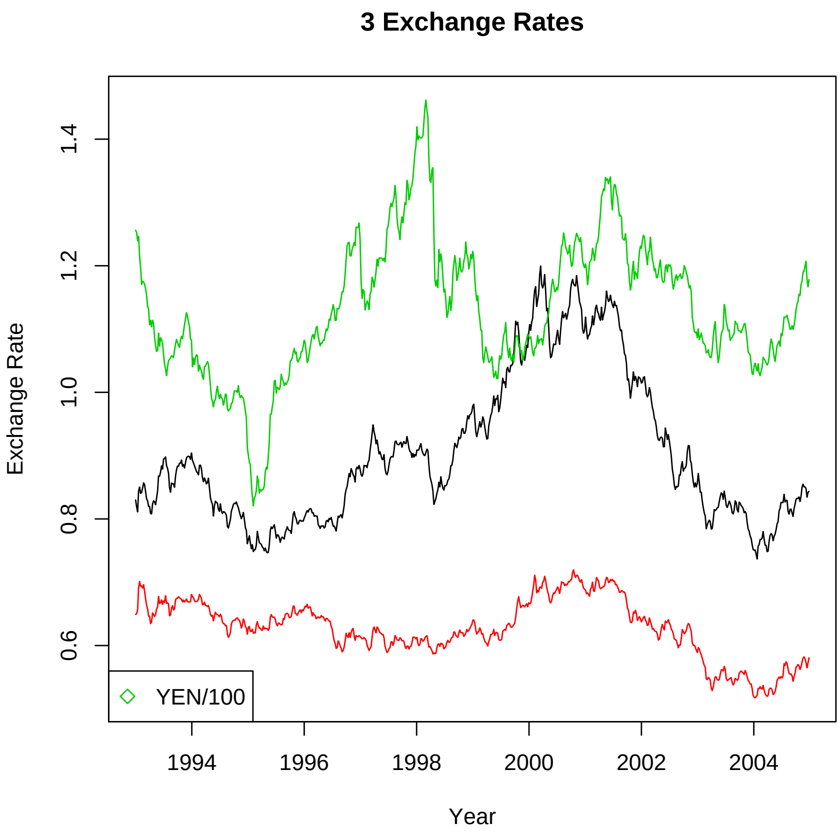

Fig. 7. 3-pair exchange rates with FX data

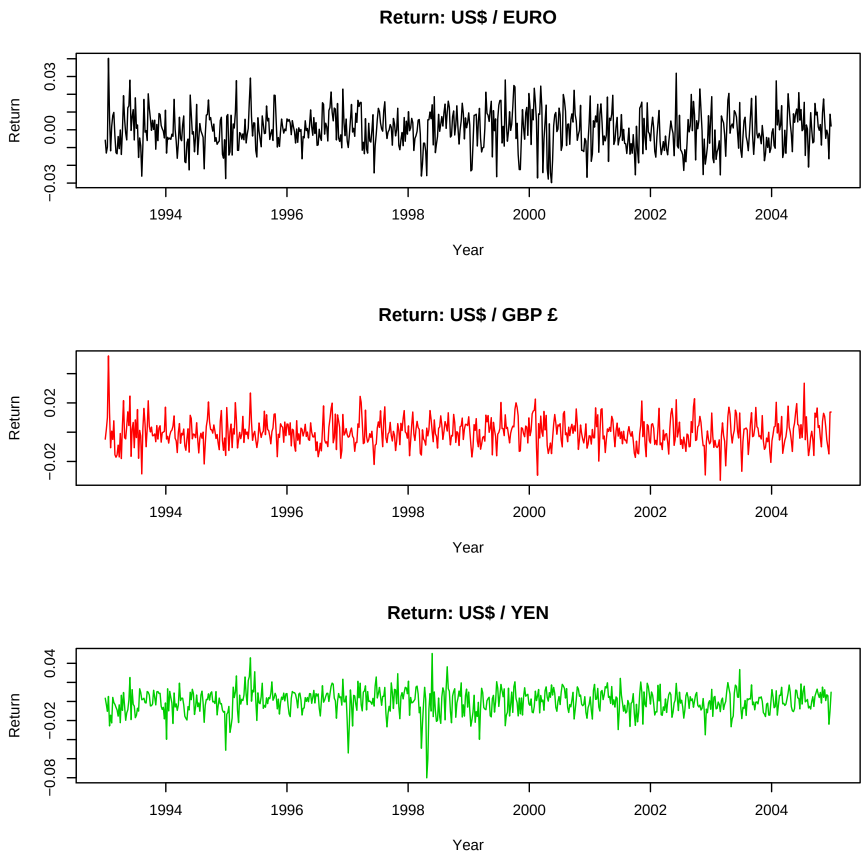

Fig. 8. 3-pair exchange rate returns with FX data

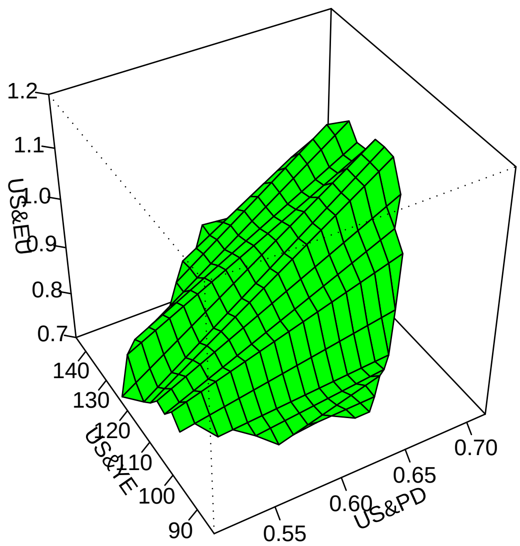

Fig. 9. 5% conditional quantile of $US\$/EURO$ conditional on ($US\$/JPY$, $US\$/POUND$)

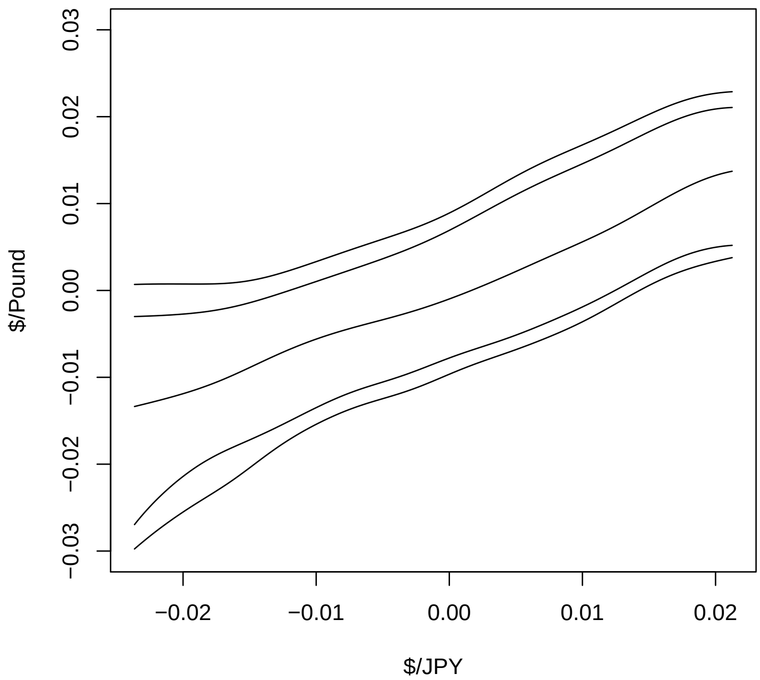

Fig. 10. conditional quantile of $US\$/POUND$ conditional on $US\$/JPY$.