Multi-Stage Change Point Detection with Copula Conditional Distribution with PCA and Functional PCA

Jong-Min Kim 1, Ning Wang 2,* and Yumin Liu 2

1 Division of Science and Mathematics, University of Minnesota-Morris, Morris, MN 56267, USA; jongmink@morris.umn.edu
2 Business School, Zhengzhou University, Zhengzhou 450001, China; yuminliu@zzu.edu.cn
* Correspondence: zz-wn@zzu.edu.cn

Received: 28 September 2020; Accepted: 9 October 2020; Published: 14 October 2020

Abstract: A global uncertainty environment, such as the COVID-19 pandemic, has affected the manufacturing industry severely in terms of supply and demand balancing. So, it is common that one stage statistical process control (SPC) chart affects the next-stage SPC chart. It is our research objective to consider a conditional case for the multi-stage multivariate change point detection (CPD) model for highly correlated multivariate data via copula conditional distributions with principal component analysis (PCA) and functional PCA (FPCA). First of all, we review the current available multivariate CPD models, which are the energy test-based control chart (ETCC) and the nonparametric multivariate change point model (NPMVCP). We extend the current available CPD models to the conditional multi-stage multivariate CPD model via copula conditional distributions with PCA for linear normal multivariate data and FPCA for nonlinear non-normal multivariate data.

Keywords: multivariate change point detection; copula; principal component analysis; function principal component analysis

1. Introduction

Since Hotelling (1949) proposed Hotelling $T^2$ statistics for the multivariate statistical process control (SPC), Crosier (1988), Lowry, Woodall, Champ and Rigdon (1992), and Zou and Tsung (2011) have proposed the multivariate versions of the cumulative sum (CUSUM) time-weighted control chart and the exponentially weighted moving average (EWMA) time-weighted SPCs. However, the manufacturing industry is still requiring a modern statistical technique dealing with non-normal high dimensional correlated multivariate data. In order to solve this difficult problem in lights of quality control, [1] proposed SPC charts as a tool for analyzing big data. Furthermore, Reference [2] discussed and compared the conventional SPCs with nonparametric SPCs in terms of the strengths and limitations.

The purpose of our paper is that, under a situation, such as high dimensional correlated variables over the several stage process in a manufacturing industry business, we consider modeling conditional multi-stage manufacturing processes for detecting faults in the several stages for the complex production system. Multi-stage CPD has emerged as a cutting-edge research area at the interface of the engineering and statistical sciences. Over the last two decades, Reference [3–5] developed the change point detection (CPD) models with needed pre-knowledge for in-control distribution and nonparametric CPD charts to detect mean, variance, and other distributional shifts. Reference [6] proposed online nonparametric multivariate CPD models. Recently, Reference [7] reviewed previous works focusing on energy divergence...
test theory and its applications in the CPD. Reference [8] proposed nonparametric multiple change point analysis of multivariate data which used the energy test for applications, such as a sliding window scheme with fixed window size, to detect change points in image data or a change point retrospective analysis. Reference [9,10] developed the ‘ecp’ R package for nonparametric multiple change point analysis of multivariate data. Reference [7] also proposed a nonparametric control chart for detecting multiple change points from multivariate time series, which is energy test-based control chart (ETCC), and compared their method with another nonparametric control chart, called nonparametric multivariate change point (NPMVCP), which was developed by [6]. The advantage of using ETCC by Reference [7] is that ETCC detects the changepoints of the mean and covariance together. Reference [11] also reviewed multi-stage manufacturing processes, and Reference [12] reviewed recent research in a dynamic screening system for sequential process monitoring. In this paper, we want to extend the current available multivariate CPD models into the conditional multi-stage multivariate CPD model via copula conditional distribution. Recently, copula modeling has been popular in biostatistics, economics, and finance because copula functions do not need normal, linear, and independent assumptions. Furthermore, copula approaches to quality control for monitoring the bivariate auto-correlated binary observations have been discussed by Reference [13,14] because copulas do not require any assumptions, such as independence, linearity, and normality, for the residual analysis, and it is possible to look at both the marginal distributions and the joint dependence structure [15].

The layout of the paper is as follows. Section 2 describes principal component analysis (PCA), functional PCA (FPCA), copula definitions, ETCC, and NPMVCP, and Section 3 describes our conditional multi-stage multivariate CPD procedure. Section 4 illustrates our proposed method with a simulated multivariate data and the real exchange currency data in America, Asia, and Europe. Finally, conclusions and future research studies are presented in Section 5.

2. Statistical Methods

We consider a situation, such as high dimensional correlated variables over conditional multi-stage processes. Among several statistical methods for the dimensional reduction of multivariate highly correlated variables, we employ the traditional linear PCA and a nonlinear FPCA in this paper. With the traditional PCA method, we consider normal multivariate data for the conditional multi-stage CPD. With a nonlinear FPCA method, we consider non-normal multivariate data for the conditional multi-stage CPD. For extending single-stage CPD to the conditional multi-stage CPD, we employ the conditional distribution by copula and the nonparametric multivariate control chart for the conditional multi-stage CPD.

2.1. Principal Component Analysis

The traditional linear PCA is one of the popular statistical methods to reduce the dimensionality of multivariate data into a smaller number of uncorrelated variables called principal components (PCs), while keeping variation in the original data.

The SPCs with primary principal components by PCA have been proposed to monitor a class of multivariate quality processes for handling multivariate data with multicollinearity between variables (see Reference [16] for details). We consider the PCA-based multivariate CPD method to demonstrate the model’s flexibility and performance by both a simulation study and a real data illustration based on Reference [7]. If the data follows the normality assumption, then we can use our proposed conditional multi-stage CPD with a linear PCA method.
2.2. Functional Principal Component Analysis

For the dimension reduction of the multivariate highly correlated and non-normal data, we also employ nonlinear FPCA to determine the factors (i.e., principal components). By using non-liner eigenfunctions to explain the variation of the time series and examine the sample covariance structure, FPCA is a better statistical dimension reduction method than the PCA proposed by Reference [17]. In addition, FPCA is the more appropriate statistical method to know the clustering pattern of the time-course data rather than the clustering pattern of the whole data at a certain time. We divide density variations into a set of orthogonal principal component functions that maximize the variance along each component estimating density functions by employing a nonparametric method and extracting common structures from the estimated functions.

The functional form of \( y_i(t) \) is given by the sum of the weighted basis functions, \( \phi_k(t) \), across the set of times \( T \).

\[
y_i(t) = \sum_{k=1}^{K} c_k \phi_k(t),
\]

where \( K \) is a number of basis functions. In this study, a Fourier basis is used to represent smooth functions as a basis function due to its flexibility and computational advantages. Here, our goal is to obtain a smooth function which fits well into the observed time series, \( y_i(t) \). To perform FPCA, we use the ‘fdapace’ R package (Reference [18]). This package is FPCA for sparsely or densely sampled random trajectories and time courses, via the principal analysis by conditional estimation (PACE) algorithm which produces covariance and mean functions, eigenfunctions, and principal component (scores), for both functional data and derivatives for both dense (functional) and sparse (longitudinal) sampling designs. For sparse designs, PACE gives fitted continuous trajectories with confidence bands, even for subjects with few longitudinal observations. Reference [19,20] developed the basic procedure behind the PACE approach for sparse functional data as follows: First, compute the cross-sectional mean \( \bar{\mu} \). Second, compute the cross-sectional covariance surface which is guaranteed to be positive semi-definite. Third, do eigenanalysis on the covariance to estimate the eigenfunctions \( \hat{\phi} \) and eigenvalues \( \hat{\lambda} \). Fourth, employ numerical integration to estimate the corresponding scores \( \hat{\eta} \), i.e.,

\[
\hat{\eta}_{ik} = \int_0^T [y(t) - \bar{\mu}(t)] \phi_k(t) dt.
\]

2.3. Copula

A copula is defined as a multivariate distribution function defined on the unit \([0,1]^p\), with \( p \) the number of marginal distributions. Copula is a flexible function to construct the dependence structure of random variables. In this paper, we consider a bivariate (two-dimensional) copula, where \( p = 2 \). Reference [21] proposed copula function such that any bivariate distribution function, \( F_{XY}(x,y) \), can be represented as a function of its marginal distribution of \( X \) and \( Y \), \( F_X(x) \) and \( F_Y(y) \), as

\[
F_{XY}(x,y) = P(X \leq x, Y \leq y) = C(F_X(x), F_Y(y), \theta) = C(U, V, \theta),
\]

where we denote \( U = F_X(x) \) and \( V = F_Y(y) \), which are the continuous cumulative distribution functions of \( X \) and \( Y \), and we denote as \( \theta \) an association parameter of the copula function. Therefore, the copula function describes the dependent mechanism between two random variables by eliminating the influence of the marginal distributions or any monotone transformation of the marginal distributions.

**Definition 1.** A \( p \)-dimensional copula is a function \( C : [0,1]^p \rightarrow [0,1] \) with the following properties:

1. For all \((U_1, \ldots, U_p) \in [0,1]^p\), then \( C(U_1, \ldots, U_p, \theta) = 0 \) if at least one coordinate of \((U_1, \ldots, U_p)\) is 0;
2. \( C(1, \ldots, 1, U_i, 1, \ldots, 1, \theta) = U_i \) for all \( U_i \in [0,1], (i = 1, \ldots, p) \);
3. $C$ is r-increasing, (see Reference [22]).

**Definition 2.** A Gaussian copula is a distribution over $[0,1]^p$. It is constructed from a multivariate normal distribution over $\mathbb{R}^p$ by using the probability integral transform. For a given correlation matrix $R \in [-1,1]^{p \times p}$, the Gaussian copula with parameter matrix $R$ can be written as

$$C(U_1, \ldots, U_p, \theta) = \Phi_R(\Phi^{-1}(U_1), \ldots, \Phi^{-1}(U_p)),$$

where $\theta$ is an association parameter of the Gaussian copula function, $\Phi^{-1}$ is the inverse cumulative distribution function of a standard normal, and $\Phi_R$ is the joint cumulative distribution function of a multivariate normal distribution with mean vector zero and a covariance matrix equal to the correlation matrix $R$.

**Definition 3.** The $p$-dimensional random vector $X = (X_1, \ldots, X_p)$ is said to have a (non-singular) multivariate Student-$t$ distribution with $v$ degrees of freedom, mean vector $\mu$ and positive-definite dispersion or scatter matrix $\Sigma$, denoted $X \sim t_p(\nu, \mu, \Sigma)$, if its density is given by

$$f(x) = \frac{\Gamma\left(\frac{v+p}{2}\right)}{\Gamma\left(\frac{v}{2}\right)\sqrt{(\pi v)^p|\Sigma|}} \left(1 + \frac{(x-\mu)^T \Sigma^{-1}(x-\mu)}{v}\right)^{-\frac{v+p}{2}}.$$

Note that, in this standard parameterization, $\text{cov}(X) = \frac{v}{v-2}\Sigma$ so that the covariance matrix is not equal to $\Sigma$ and is in fact only defined if $v > 2$. Useful reference for the multivariate t-copula is Reference [23].

**Definition 4.** (Archimedean Copula). Let $C$ be an associative, Archimedeed copula. Then, there exists a strictly decreasing and convex (hence continuous) function (called the generator) $\varphi : [0,1] \to [0,\infty)$ with $\varphi(1) = 0$ such that for every pair $(U, V)$ in $[0,1] \times [0,1]$,

$$C(U, V, \theta) = \varphi^{-1}(\varphi(U) + \varphi(V)),$$

where $\varphi^{-1}$ is the "pseudo-inverse" of $\varphi$, given by

$$\varphi^{-1}(x) = \begin{cases} \varphi^{-1}(x), & \text{if } 0 \leq x \leq \varphi(0) \\ 0, & \text{if } \varphi(0) < x \leq +\infty \end{cases}.$$

Table 1 shows the most commonly used Archimedean copula functions, such as Clayton copula, Farlie-Gumbel-Morgenstern (FGM) copula, Frank copula, and Gumbel copula with an association parameter $\theta$ of each copula function.

| Copula | Copula Function |
|--------|-----------------|
| FGM    | $C_{FGM}(U, V, \theta) = UV + \theta UV(1-U)(1-V), \ \theta \in (-1,1]$ |
| Clayton| $C_U(U, V, \theta) = (U^{-\theta} + V^{-\theta} - 1)^{-\frac{1}{\theta}}, \ \theta \in (0, \infty)$ |
| Frank  | $C^\ell(U, V, \theta) = -\frac{1}{\theta} \log \left[1 + \frac{(e^{-\theta u-1})(e^{-\theta v-1})}{e^{-\theta u v-1}}\right], \ \theta \in \mathbb{R} \setminus \{0\}$ |
| Gumbel | $C^G(U, V, \theta) = \exp \left[-\left(-\log U\right)^\theta + \left(-\log V\right)^\theta\right]^{1/\theta}, \ \theta \geq 1$ |
Because of the limited range of the association parameter, \( \theta \), in the Clayton copula, FGM copula, and Gumbel copula functions in Table 1, we have difficulty applying the Clayton copula, FGM copula, and Gumbel copula functions to SPC, except for the Frank copula. We employed the Gaussian copula in Definition 1, the t-copula in Definition 2, and one of the Archimedean copula, the Frank copula introduced in Definition 3, for our proposed conditional multi-stage CPD.

2.4. Energy Test-Based Control Chart (ETCC)

Reference [7] proposed a nonparametric CPD model which can simultaneously detect any change of mean, variance, or dependence structure all together in the multivariate distribution. Furthermore, Reference [7] used the maximum energy divergence-based permutation test to screen out the multiple change points for multivariate time series and employs the discrepancy of empirical characteristic functions of two random vectors. The empirical distribution of the test statistic can be obtained by permutation samples. Then, the sequential detection of change points can be performed under the algorithm introduced by the change point model (see Reference [24]) to form an online detection. For a change point detection problem, it is set that the change occurs at \( \tau \) when the two random vectors \( \{ X_i \in \mathbb{R}^p : X_i \sim F, i = 1, \ldots, \tau \} \) and \( \{ Y_j \in \mathbb{R}^p : Y_j \sim G, j = \tau + 1, \ldots, T \} \) have a distribution shift. In a multiple change point case, \( \tau_i, i = 1, 2, \ldots \), the changes’ detection can be formulated as

\[
X_t \sim \begin{cases} 
F_0, & t \leq \tau_1 \\
F_1, & \tau_1 < t \leq \tau_2 \\
F_2, & \tau_2 < t \leq \tau_3 \\
\vdots & \\
F_j, & \tau_j < t \leq \tau_{j+1} \\
\vdots & 
\end{cases}
\]

Because the corresponding characteristic functions of \( X_i \) and \( Y_j \), i.e., \( f_x \) and \( f_y \), are uniquely determined, using the divergence between characteristic functions of the two random vectors to monitor the change is an applicable routine. Reference [25] employed an integrated weighted distance between two characteristic functions, and proved that the larger the distance, the more likely a change may occur between the two random vectors. Reference [7] named a nonparametric CPD model which is a nonparametric control chart as an energy test-based control chart (ETCC). Based on the ETCC, Reference [26] made an R package ‘EnergyOnlineCPM’ which centers on the Phase II nonparametric CPD model to online detect the multiple change points.

2.5. NPMVCP by Holland and Hawkins (2014)

Reference [6] proposed a nonparametric SPC by employing multivariate rank-based test by Reference [27]. The multivariate CPD model by Reference [24] defines changes in a sequence, \( X_1, \ldots, X_t \), as follows:

\[
X_t \sim \begin{cases} 
F(\mu), & i \leq \tau \\
F(\mu + \sigma), & i > \tau 
\end{cases}
\]
and $H_0 : \sigma = 0$, versus $H_1 : \sigma \neq 0$. The test statistics and their asymptotic distribution are given for $k \in \{1, \ldots, t - 1\}$ as:

$$
\frac{t_k}{t - k} \bar{r}_t^{(k)} \sum_{k,t}^{-1} r_t^{(k)} \frac{d}{d\theta} \chi^2 \text{ if } t \to \infty,
$$

where $\sum_{k,t}$ is the pooled sample covariance matrix for the centered rank vector $r_t^{(k)}$ computed by using a kernel function. Reference [6] developed the test statistic

$$
r_{k,t} = \bar{r}_t^{(k)} \sum_{k,t}^{-1} r_t^{(k)},
$$

where $\sum_{k,t} = \frac{t-k}{tk} \sum_{l}$ is the unpooled estimator of the covariance matrix of the centered ranks.

3. Multi-Stage CPD with Copula Conditional Distribution

In this research, we consider the conditional multi-stage multivariate CPD by performing the conditional transformed data by copula functions (Gaussian, t, Frank).

**Corollary 1.** For two random variables $X_1$ and $X_2$, we can derive the conditional distribution of $X_1$ given $X_2$, $F_{1|2}(X_1|X_2)$, as follows:

$$
F_{1|2}(X_1|X_2) = \frac{\partial C(U_1, U_2, \theta_{12})}{\partial U_2},
$$

where $\theta_{12}$ is an association parameter of the copula function, $U_1 = F(X_1)$, and $U_2 = F(X_2)$. Similarly, for two random variables $X_2$ and $X_3$, we can derive the conditional distribution of $X_3$ given $X_2$, $F_{3|2}(X_3|X_2)$, as follows:

$$
F_{3|2}(X_3|X_2) = \frac{\partial C(U_2, U_3, \theta_{23})}{\partial U_2},
$$

where $\theta_{23}$ is an association parameter of the copula function, $U_2 = F(X_2)$, and $U_3 = F(X_3)$.

**Corollary 2.** Assume we have three random variables $X_1$, $X_2$, $X_3$. We can derive the conditional cumulative distribution function as follows:

$$
F_{3|12}(X_3|X_1, X_2) = \frac{\partial C(U_1, U_2, U_3, \theta_{3|12})}{\partial U_1 \partial U_2},
$$

where $\theta_{3|12}$ is an association parameter of the copula function, $U_1 = F(X_1)$, $U_2 = F(X_2)$, and $U_3 = F(X_3)$.

The procedures for estimating the parameter of the copula for the conditional distribution function can be defined as follows. The first step is that we employ the empirical CDF approach to transform the observations to uniform distributed data in $[0, 1]$. (see Reference [28] for details). Because the empirical marginal distributions of $U$ and $V$ are uniform on $[0, 1]$ such that they are parameter-free, the rank-based approach allows us to compute joint probabilities without knowing marginal distributions. In this paper, the association parameter estimation for bivariate copulas is computed by using a maximum likelihood estimation method which can be used in the ‘BiCopEst’ function from the ‘CD Vine’ R package [29]. The second step is that after the parameters $\theta_{ij}$ and $\theta_{jk}$ in $C(U_i, U_j, \theta_{ij})$ and $C(U_j, U_k, \theta_{jk})$ are estimated, the conditional CDFs $\frac{\partial C(U_i, U_j, \hat{\theta}_{ij})}{\partial U_i} = F(X_i | X_j) = U_{ij}$ and $\frac{\partial C(U_j, U_k, \hat{\theta}_{jk})}{\partial U_j} = F(X_k | X_j) = U_{k|j}$ are computed with
the estimates $\hat{\theta}_{ij}$ and $\hat{\theta}_{jk}$ by partial derivatives of $C(U_i, U_j, \hat{\theta}_{ij})$ and $C(U_j, U_k, \hat{\theta}_{jk})$. The third step is that the association parameter $\hat{\theta}_{kij}$ of $C(F(X_i|X_j), F(X_k|X_j), \theta_{kij}) = C(U_{ij}, U_{kij}, \hat{\theta}_{kij})$ is estimated by the maximum likelihood estimation method. The last step is that with the estimated parameter $\hat{\theta}_{kij}$ the conditional CDF $\frac{aC(U_{ij}, U_{kij}, \hat{\theta}_{kij})}{aU_{kij}} = F(X_k|X_i, X_j) = U_{kij}$ is computed. By following these procedures, we can make the conditional transformed data with a copula function for conditional multi-stage CPD.

For the dimensional reduction to the smaller number of principal components compared to the number of variables in the whole dataset, we apply PCA or FPCA to each stage dataset $(X_i, i = 1, 2, 3)$, and, iteratively, we perform PCA or FPCA on each stage dataset$(X_i, i = 1, 2, 3)$ to do a dimensional reduction. Each stage size is $n_i$ for $i = 1, 2, 3$, and we set an equal sample size for each stage $(n_1 = n_2 = n_3)$ for the computation convenience of the copula method such that $n = n_1 + n_2 + n_3$ for the simulated multivariate data $X = (X_1, X_2, X_3)$. After performing PCA or FPCA with $X = (X_1, X_2, X_3)$, we transform the PCA scores generated from $X = (X_1, X_2, X_3)$ to the uniform distribution transformed data $Y = (Y_1, Y_2, Y_3)$ by the empirical CDF approach, and then we apply the copula conditional distribution to these $Y = (Y_1, Y_2, Y_3)$, so that the conditional transformed data are generated, such as the $Y_1$ vector, $F(Y_2|Y_1)$ vector, and $F(Y_3|Y_1, Y_2)$ vector.

Finally, we employ the energy test-based control chart (ETCC, Reference [7]) and nonparametric multivariate change point model (NPMVCP, Reference [6]) of each stage for detecting change points to the $Y_1$ vector, $F(Y_2|Y_1)$ vector, and $F(Y_3|Y_1, Y_2)$ vector. We propose the conditional multi-stage CPD scheme based on the copula conditional distributions and PCA or FPCA for multivariate correlated data as follows:

1. Apply PCA or FPCA to the simulated multivariate data $X = (X_1, X_2, X_3)$ for dimensional reduction to several principal components.
2. Transform the PCA or FPCA scores to the uniform distribution transformed data $Y = (Y_1, Y_2, Y_3)$ by the empirical CDF approach.
3. Apply the copula conditional distribution for transforming three-stage multivariate data $Y = (Y_1, Y_2, Y_3)$ to two conditional datasets $F(Y_2|Y_1)$ and $F(Y_3|Y_1, Y_2)$.
4. Apply multivariate CPD methods (ETCC and NPMVCP) to both conditional stages $F(Y_2|Y_1)$ and $F(Y_3|Y_1, Y_2)$.
5. Detect change points by ETCC and NPMVCP from each $F(Y_2|Y_1)$ and $F(Y_3|Y_1, Y_2)$.

4. Illustrated Example

In order to illustrate our proposed conditional multi-stage multivariate CPD, we compare our method with recent multivariate CPD models by using simulated multivariate data and real data in Section 4.

4.1. Simulation Study

We want to generate multi-stage simulated multivariate dataset so that a current stage process is affected by the previous stage process and multivariate data have high correlations among variables, we employed the copula dependence method which can express the multi-stage dependence and can make a high correlation structure among variables in each stage. With this simulated dataset, we want to verify our conditional multi-stage CPD scheme by the copula conditional distribution. We generate three stage simulated datasets $(X_1, X_2, X_3)$. We name $X_1$ as stage 1, $X_2$ as stage 2, and $X_3$ as stage 3.

For the dataset $X_1$, we simulate the highly correlated multivariate data by using the ‘copula’ R package with the ‘normalCopula’ function for the five variables, and each variable has sample size 400 with the correlation parameters $(0.9, 0.8, 0.8, 0.8, 0.7, 0.7, 0.7, 0.6, 0.5, 0.4)$, specifying that the type of the symmetric positive definite matrix characterizing the elliptical copula is unstructured. For each marginal distribution
for the five variables in \( X_1 \), we use three gamma distributions (\( X_{11} \) follows gamma distribution with shape parameter (set to 5) and scale parameter (set to 1), \( X_{12} \) follows gamma distribution with shape parameter (set to 5) and scale parameter (set to 2), and \( X_{13} \) follows gamma distribution with shape parameter (set to 5) and scale parameter (set to 3)) and two exponential distributions (\( X_{14} \) follows exponential distribution with parameter (set to 5) and \( X_{15} \) follows exponential distribution with parameter (set to 2)).

For the dataset \( X_2 \), we simulate the highly correlated multivariate data by using ‘copula’ R package with the ‘normalCopula’ function for the five variables, and each variable has sample size 400 with the correlation parameters (0.4, 0.5, 0.6, 0.7, 0.7, 0.7, 0.6, 0.5, 0.4), specifying that the type of the symmetric positive definite matrix characterizing the elliptical copula is unstructured. For each marginal distribution for the five variables in \( X_2 \), we use three gamma distributions (\( X_{21} \) follows gamma distribution with shape parameter (set to 2) and scale parameter (set to 1), \( X_{22} \) follows gamma distribution with shape parameter (set to 2) and scale parameter (set to 2), and \( X_{23} \) follows gamma distribution with shape parameter (set to 2) and scale parameter (set to 3)) and two exponential distributions (\( X_{24} \) follows exponential distribution with parameter (set to 2), and \( X_{25} \) follows exponential distribution with parameter (set to 5)).

For the dataset \( X_3 \), we simulate the highly correlated multivariate data by using ‘copula’ R package with ‘normalCopula’ function for the 5 variables and each variable has sample size 400 with the correlation parameters (0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5) to specify the symmetric positive definite matrix characterizing that the elliptical copula is unstructured. For each marginal distribution for five variables in \( X_3 \), we use three gamma distributions (\( X_{31} \) follows gamma distribution with shape parameter (set to 3) and scale parameter (set to 1), \( X_{32} \) follows gamma distribution with shape parameter (set to 2) and scale parameter (set to 2), and \( X_{33} \) follows gamma distribution with shape parameter (set to 3) and scale parameter (set to 3)) and two exponential distributions (\( X_{34} \) follows exponential distribution with parameter (set to 4), and \( X_{35} \) follows exponential distribution with parameter (set to 3)).

Figure 1 shows the data plots of the three stages (\( X_1, X_2, X_3 \)). In Figure 1, stage 1 shows bigger spread than stage 2 and stage 3, stage 2 shows smaller spread than stage 1 and stage 3, and stage 3 has bigger spread than stage 2. The correlation matrix with the simulated multivariate data in Table 2 shows high correlations exist among the five variables in \( X_1, X_2, \) and \( X_3 \).

To perform the change point detection for the conditional multi-stage multivariate highly correlated simulated dataset, we apply PCA or FPCA to the simulated data \( X = (X_1, X_2, X_3) \) and then generate PCA or FPCA scores to the uniform distribution transformed data \( Y = (Y_1, Y_2, Y_3) \) by the empirical CDF approach. For FPCA, we employ Fourier basis functions for constructing functional eigenfunctions with \( K = 7 \) and \( T = 433 \) introduced in Section 2.2, and three eigenfunctions by the FPCA are transformed to the uniform distributed data \( Y = (Y_1, Y_2, Y_3) \) by the empirical CDF approach. We apply the copula conditional distribution for transforming three-stage multivariate data \( Y = (Y_1, Y_2, Y_3) \) to two conditional datasets \( F(Y_2|Y_1) \) and \( F(Y_3|Y_1, Y_2) \) and then apply multivariate CPD methods (ETCC and NPMVCP) to each \( Y_1, F(Y_2|Y_1) \) and \( F(Y_3|Y_1, Y_2) \) to detect change points for each stage \( Y_1, F(Y_2|Y_1) \) and \( F(Y_3|Y_1, Y_2) \).

Table 3 shows the PCA variance proportions for \( X_1, X_2 \) and \( X_3 \). Table 4 shows the PCA variance proportions with copula conditional distributions of t-copula, Gaussian copula and Frank copula, \( F(Y_2|Y_1) \) and \( F(Y_3|Y_1, Y_2) \). Table 5 shows the change points by both ETCC and NPMVCP with the whole simulated data case and PCA components of the simulated multivariate data. We compare our proposed method with recent methods on multi-stage change point detection of multivariate data. We chose a nonparametric multiple change point analysis of multivariate data developed by Reference [9,10] with ‘ecp’ R package. Table 6 shows change point detections with nonparametric multiple change point analysis of simulated multivariate data with using the command ‘ks.cp3o_delta’ for the change points estimation by pruned objective via the Kolmogorov–Smirnov statistic, and the window size between segments is 30 in the ‘ecp’ R package. Compared with the results of ETCC and NPMVCP in Table 5, James, Zhang, and Matteson (2019) detected more change points with simulated multivariate data for each stage (stage 1, stage 2, stage 3).
Table 7 shows the change point detections by both ETCC and NPMVCP with FPCA components of simulated multivariate data. The ETCC and NPMVCP with FPCA components of simulated multivariate data in Table 7 detected more change points with simulated multivariate data for each stage (stage 1, stage 2, stage 3) than the James, Zhang, and Matteson (2019) nonparametric multiple change point method.

For considering the conditional multivariate data, the change point detections by both ETCC and NPMVCP with a copula conditional distribution of the t-copula, Gaussian copula, and Frank copula with PCA components are proposed in this paper. We notice that the performance of our copula-based method depends on the choice of the copula function. But, as we mentioned in Section 2.3, it is difficult to apply many copula functions to SPC because the range of the association parameter, $\theta$, in the Clayton copula, FGM copula, and Gumbel copula functions in Table 1 is restricted so that we had computation difficulty to apply the Clayton copula, FGM copula, and Gumbel copula functions to SPC. Since the range of the association parameter, $\theta$, from the Gaussian copula and the t-copula is $\theta \in (-\infty, \infty)$, and the Frank copula is $\theta \in (-\infty, \infty) \setminus \{0\}$, we can compare these copula functions to simulated multivariate data to choose the

![Figure 1. Plots with simulated multivariate data (region means stage).](image-url)
copula function properly, which is the critical issue about a copula-based CPD method. Table 8 shows the change point detections by both ETCC and NPMVCP with copula conditional distribution of the t-copula, Gaussian copula, and Frank copula with PCA components. From Table 8, we can notice that the change point detections by the three copula functions (Gaussian copula, t-copula, and Frank copula) are slightly different.

Through the empirical trial and error learning based on the certain manufacturing circumstance, we recommend that industry practitioners compare these copula-based CPD methods and choose a copula function properly. Figures 2–4 show the eigenvalues and eigenfunctions of the FPCA plots of $X_1$, $X_2$, and $X_3$. Table 9 shows the change point detections by both ETCC and NPMVCP with a copula conditional distribution of the t-copula, Gaussian copula, and Frank copula with FPCA components. With the simulated multivariate data, we found that the FPCA-based conditional multi-stage multivariate CPD method detected more change points for each stage case rather than the PCA-based conditional multi-stage multivariate CPD method. From Table 9, the FPCA-based conditional multi-stage multivariate CPD method is a promising research area for detecting change points if we can implement the proper copula function empirically.

Table 2. Correlation matrix with the simulated multivariate data. Note: $X_1$, $X_2$ and $X_3$ are vectors.

|   | $X_1$ | $X_{11}$ | $X_{12}$ | $X_{13}$ | $X_{14}$ | $X_{15}$ |
|---|-------|---------|---------|---------|---------|---------|
| $X_{11}$ | 1     | 0.9005  | 0.7873  | 0.7762  | 0.7778  |
| $X_{12}$ | 0.9005 | 1       | 0.7081  | 0.6399  | 0.6988  |
| $X_{13}$ | 0.7873 | 0.7081  | 1       | 0.5858  | 0.4561  |
| $X_{14}$ | 0.7762 | 0.6399  | 0.5858  | 1       | 0.3780  |
| $X_{15}$ | 0.7778 | 0.6988  | 0.4561  | 0.3780  | 1       |

|   | $X_2$ | $X_{21}$ | $X_{22}$ | $X_{23}$ | $X_{24}$ | $X_{25}$ |
|---|-------|---------|---------|---------|---------|---------|
| $X_{21}$ | 1     | 0.3501  | 0.4843  | 0.5644  | 0.6425  |
| $X_{22}$ | 0.3501 | 1       | 0.6626  | 0.6010  | 0.6890  |
| $X_{23}$ | 0.4843 | 0.6626  | 1       | 0.5559  | 0.4614  |
| $X_{24}$ | 0.5644 | 0.6010  | 0.5559  | 1       | 0.3409  |
| $X_{25}$ | 0.6425 | 0.6890  | 0.4614  | 0.3409  | 1       |

|   | $X_3$ | $X_{31}$ | $X_{32}$ | $X_{33}$ | $X_{34}$ | $X_{35}$ |
|---|-------|---------|---------|---------|---------|---------|
| $X_{31}$ | 1     | 0.5409  | 0.5325  | 0.5097  | 0.5043  |
| $X_{32}$ | 0.5409 | 1       | 0.5007  | 0.4843  | 0.4783  |
| $X_{33}$ | 0.5325 | 0.5007  | 1       | 0.4811  | 0.5467  |
| $X_{34}$ | 0.5097 | 0.4843  | 0.4811  | 1       | 0.4540  |
| $X_{35}$ | 0.5043 | 0.4783  | 0.5467  | 0.4540  | 1       |

Table 3. Principal component analysis (PCA) with the simulated multivariate data.

|   | $X_1$ | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 |
|---|-------|--------|--------|--------|--------|--------|
| Proportion of Variance | 0.7462 | 0.1229 | 0.0841 | 0.0409 | 0.0059 |
| Cumulative Proportion  | 0.7462 | 0.8691 | 0.9532 | 0.9941 | 1.0000 |

|   | $X_2$ | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 |
|---|-------|--------|--------|--------|--------|--------|
| Proportion of Variance | 0.6487 | 0.1494 | 0.1202 | 0.0707 | 0.0111 |
| Cumulative Proportion  | 0.6487 | 0.7980 | 0.9182 | 0.9889 | 1.0000 |

|   | $X_3$ | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 |
|---|-------|--------|--------|--------|--------|--------|
| Proportion of Variance | 0.6099 | 0.1157 | 0.0981 | 0.0912 | 0.0851 |
| Cumulative Proportion  | 0.6099 | 0.7255 | 0.8237 | 0.9149 | 1.0000 |
Table 4. PCA variance proportions with simulated multivariate data.

| Copula     | Comp.1 | Comp.2 | Comp.3 |
|------------|--------|--------|--------|
| \( F(Y_2|Y_1) \) | 0.6487 | 0.1494 | 0.1202 |
| \( F(Y_3|Y_1,Y_2) \) | 0.6099 | 0.1157 | 0.0981 |

| Copula     | Comp.1 | Comp.2 | Comp.3 |
|------------|--------|--------|--------|
| \( F(Y_2|Y_1) \) | 0.6485 | 0.1493 | 0.1204 |
| \( F(Y_3|Y_1,Y_2) \) | 0.6099 | 0.1155 | 0.0982 |

| Copula     | Comp.1 | Comp.2 | Comp.3 |
|------------|--------|--------|--------|
| \( F(Y_2|Y_1) \) | 0.6485 | 0.1492 | 0.1207 |
| \( F(Y_3|Y_1,Y_2) \) | 0.6095 | 0.1153 | 0.0985 |

Table 5. Change point detection by energy test-based control chart (ETCC) and nonparametric multivariate change point (NPMVCP) of PCA with simulated multivariate data.

| Method | ETCC | NPMVCP | ETCC | NPMVCP | ETCC | NPMVCP |
|--------|------|--------|------|--------|------|--------|
| Whole data | 170 191 | 150 186 | 208 264 243 323 368 382 | 193 219 240 304 361 382 | 42 238 | 42 238 |
| PCA Components | 174 174 | 206 218 | 193 218 | 238 296 321 | 238 294 321 |

Table 6. Change point detection with the James, Zhang, and Matteson (2019) NPMVCP analysis of simulated multivariate data.

| Data | \( X_1 \) | \( X_2 \) | \( X_3 \) |
|------|--------|--------|--------|
| Change Points | 85 140 174 241 326 365 | 73 118 160 197 239 318 361 | 50 81 120 155 187 241 279 328 |

Table 7. Change point detection by ETCC and NPMVCP of functional PCA (FPCA) with simulated multivariate data.

| Functional PCA |
|----------------|
| \( X_1 \) | 24 45 68 89 110 131 151 171 191 212 232 252 272 292 313 333 354 375 396 |
| \( X_2 \) | 30 51 71 91 112 133 154 174 194 215 236 257 279 300 320 340 361 382 |
| \( X_3 \) | 21 42 62 82 103 123 144 164 184 208 229 249 269 290 311 332 353 374 397 |

| NPMVCP |
|--------|
| \( X_1 \) | 24 45 68 89 110 130 150 170 191 211 231 251 272 292 312 333 354 375 395 |
| \( X_2 \) | 30 50 70 91 112 133 153 173 194 215 236 257 279 299 319 339 360 381 |
| \( X_3 \) | 21 41 61 82 102 123 143 163 183 208 228 248 269 290 311 332 353 374 397 |

Table 8. Change point detection by ETCC and NPMVCP of PCA and copulas with simulated multivariate data.

| Method         | ETCC  | NPMVCP | ETCC  | NPMVCP |
|----------------|-------|--------|-------|--------|
| Copula         |       |        |       |        |
| t-Copula       | 212 193 217 369 186 192 213 364 | 6 325 327 331 | 5 318 324 331 |
| Gaussian Copula| 196 209 214 369 186 195 213 364 | 6 325 326 | 5 318 324 |
| Frank Copula   | 194 211 215 368 186 193 213 364 | 68 296 322 328 | 35 290 316 326 |
Figure 2. FPCA Plots of $X_1$.

Figure 3. FPCA Plots of $X_2$.

Figure 4. FPCA plots of $X_3$. 
We can find that there are high correlations among currencies in Asia and America but not in Europe. Our data set contains daily foreign exchange rates in each continental region which each continental region is financially and economically influenced by another continental region by the time zone difference. The data set was retrieved from the currency database retrieval system provided by Professor Werner Antweiler’s website at UBC (University of British Columbia)’s Sauder School of Business, http://fx.sauder.ubc.ca/data.html. We denote \( S_t \) to be an observed daily foreign exchange rate process in discrete time, \( t = 1, \ldots, n \), and \( r_t = \log(S_t / S_{t-1}) \) to be the rates of return of the exchange rates at time \( t \). In particular, we select highest Gross Domestic Product (GDP) to lowest GDP order in each continent so that, in Asia, we select Japan, South Korea, China, Taiwan, Philippines, Thailand, India, and Vietnam; in Europe, we select Norway, Switzerland, Denmark, Sweden, United Kingdom, Poland, Hungary, and Russia; and, in America, we select USA, Canada, Chile, Uruguay, Brazil, Mexico, Columbia, and Peru in Table 10.

Figure 5 shows the time plots of the twenty-four currencies in America, Europe, and Asia.

Table 11 shows the correlation matrix with the twenty-four currencies in the period (1/3/2013 to 10/6/2014). We can find that there are high correlations among currencies in Asia and America but not high correlations among currencies in Europe. Table 12 shows the results of PCA variance proportions...
with real exchange currency data in the period. The result in Table 12 shows that America and Asia have similar PCA component variance proportions but Europe is different from America and Asia in terms of PCA component variance proportions. Table 13 shows PCA variance proportions with copula conditional distributions with t-copula, Gaussian copula, and Frank copula for \( F(\text{Europe} \mid \text{Asia}) \) and \( F(\text{America} \mid (\text{Europe}, \text{Asia})) \) of real data. Table 14 shows the change point detection by ETCC and NPMVCP with real exchange currency data and PCA components. Figures 6–8 show the eigenvalues and eigenfunctions of FPCA plots of Asia, Europe, and America. Table 15 shows the change points with real exchange currency data by the \([9,10]\) nonparametric multiple change point analysis with the ‘ecp’ R package. Compared with the results of ETCC and NPMVCP in Table 14, Reference \([10]\) detected more change points with real exchange currency data with America, Asia, and Europe (1/3/2013 to 10/6/2014). Table 16 shows change point detections of FPCA components of real exchange currency data with America, Asia, and Europe (1/3/2013 to 10/6/2014). The ETCC and NPMVCP with FPCA components of real data in Table 16 detected more change points with real exchange currency data with America, Asia, and Europe (1/3/2013 to 10/6/2014) than Reference \([10]\) nonparametric multiple change point method. To consider the conditional multivariate real data, the change point detections by ETCC and NPMVCP with copula function of PCA components are shown in Table 17. Table 18 shows change point detection by ETCC and NPMVCP with copula function of FPCA components for real exchange currency data with America, Asia, and Europe (1/3/2013 to 10/6/2014).

For the second real data application, we consider real exchange currency data with America (1/3/2013 to 10/3/2014) and Asia (1/4/2013 to 10/6/2014) because of the time zone difference between America and Asia. Table 19 shows change point detection by ETCC and NPMVCP with real exchange currency data with America (1/3/2013 to 10/3/2014) and Asia (1/4/2013 to 10/6/2014) and the PCA components of the data. Table 20 shows the change point detection with real exchange currency data with America (1/3/2013 to 10/3/2014) and Asia (1/4/2013 to 10/6/2014) by Reference \([9,10]\) nonparametric multiple change point analysis. Compared with the results of ETCC and NPMVCP in Table 19, Reference \([10]\) detected more change points with real exchange currency data with America (1/3/2013 to 10/3/2014) and Asia (1/4/2013 to 10/6/2014). Table 21 shows change point detection by ETCC and NPMVCP with FPCA components of real exchange currency data with America (1/3/2013 to 10/3/2014) and Asia (1/4/2013 to 10/6/2014). The ETCC and NPMVCP with FPCA components of real data in Table 21 detected more change points with real exchange currency data with America (1/3/2013 to 10/3/2014) and Asia (1/4/2013 to 10/6/2014) than \([10]\) nonparametric multiple change point method. We also considered the conditional real data, Asia given America, in terms of time zone difference with America (1/3/2013 to 10/3/2014) and Asia (1/4/2013 to 10/6/2014). Table 22 shows change point detection by ETCC and NPMVCP of the copula conditional distribution and PCA with real exchange currency data of America (1/3/2013 to 10/3/2014) and Asia (1/4/2013 to 10/6/2014). Figures 9 and 10 show the eigenvalues and eigenfunctions of FPCA plots of America (1/3/2013 to 10/3/2014) and Asia (1/4/2013 to 10/6/2014). Table 23 shows change point detection by ETCC and NPMVCP of the copula conditional distribution and FPCA with real exchange currency data of America (1/3/2013 to 10/3/2014) and Asia (1/4/2013 to 10/6/2014). From these two real data examples, we can conclude that the FPCA-based conditional multi-stage multivariate CPD method detected more change points for each stage case rather than the PCA-based conditional multi-stage multivariate CPD method because FPCA is nonlinear PCA, which can be flexible to the real data.
Figure 5. Twenty-four countries in America, Europe, and Asia (1/3/2013 to 10/6/2014).
Table 10. Eight countries in Asia, 8 countries in Europe, and 8 countries in America. Order by 2017 GDP.

| Order | ASIA       | EUROPE       | AMERICA    |
|-------|------------|--------------|------------|
| 1     | JAPAN      | NORWAY       | USA        |
| 2     | SOUTH KOREA| SWITZERLAND  | CANADA     |
| 3     | TAIWAN     | DENMARK      | CHILE      |
| 4     | CHINA      | SWEDEN       | URUGUAY    |
| 5     | PHILIPPINES| UNITED KINGDOM| BRAZIL    |
| 6     | THAILAND   | POLAND       | MEXICO     |
| 7     | INDIA      | HUNGARY      | COLUMBIA   |
| 8     | VIETNAM    | RUSSIA       | PERU       |

Table 11. Correlation matrix with real exchange currency data (1/3/2013 to 10/6/2014).

|          | Japan    | South_Korea | Taiwan  | China    | Philippines | Thailand | India | Vietnam |
|----------|----------|-------------|---------|----------|-------------|----------|-------|---------|
| Japan    | 1.0000   |             |         |          |             |          |       |         |
| South_Korea | 0.3259  | 1.0000      | 0.7789  | 0.6845   | 0.7290      | 0.6015   | 0.4852| 0.2849  |
| Taiwan   | 0.4589   | 0.7789      | 1.0000  | 0.8747   | 0.7991      | 0.7211   | 0.4921| 0.3990  |
| China    | 0.3962   | 0.6845      | 0.8747  | 1.0000   | 0.7437      | 0.6980   | 0.4592| 0.4669  |
| Philippines | 0.3608 | 0.7290      | 0.7991  | 0.7437   | 1.0000      | 0.6963   | 0.5629| 0.3817  |
| Thailand | 0.3474   | 0.6015      | 0.7211  | 0.6980   | 0.6963      | 1.0000   | 0.5147| 0.3021  |
| India    | 0.1679   | 0.4852      | 0.4921  | 0.4592   | 0.5629      | 0.5147   | 1.0000| 0.2014  |
| Vietnam  | 0.2039   | 0.2849      | 0.3990  | 0.4669   | 0.3817      | 0.3021   | 0.2014| 1.0000  |

|          | Norway    | Switzerland | Denmark | Sweden    | UK         | Poland    | Hungary | Russia   |
|----------|-----------|-------------|---------|-----------|------------|----------|---------|----------|
| Norway   | 1.0000    |             | -0.0388 | 0.0163    | 0.5449     | 0.1687   | 0.0887  | 0.1385   |
| Switzerland | -0.0388 | 1.0000      | 0.0632  | 0.0163    | 0.1622     | -0.1342  | 0.1775  | 0.1616   |
| Denmark  | 0.0163    | 0.0632      | 1.0000  | 0.0291    | -0.0071    | 0.0471   | 0.0782  | 0.0460   |
| Sweden   | 0.5449    | 0.0163      | 0.0291  | 1.0000    | 0.0779     | 0.1473   | 0.1775  | 0.1616   |
| UK       | 0.1674    | 0.1622      | -0.0071 | 0.0779    | 1.0000     | 0.0067   | 0.0321  | 0.2779   |
| Poland   | 0.1887    | -0.1342     | 0.0471  | 0.1473    | 0.0067     | 1.0000   | 0.5621  | 0.2759   |
| Hungary  | 0.1385    | -0.1332     | 0.0782  | 0.1775    | 0.0321     | 0.5621   | 1.0000  | 0.2984   |
| Russia   | 0.1701    | -0.1780     | 0.0460  | 0.1616    | 0.2779     | 0.2759   | 0.2984  | 1.0000   |
| USA      | 1         | 0.6630      | 0.5746  | 0.5512    | 0.3608     | 0.4618   | 0.6070  | 0.8030   |
| Canada   | 0.6630    | 1.0000      | 0.5011  | 0.4230    | 0.4462     | 0.5000   | 0.5182  | 0.5910   |
| Chile    | 0.5746    | 0.5011      | 1.0000  | 0.2876    | 0.5417     | 0.6306   | 0.6687  | 0.6310   |
| Uruguay  | 0.5512    | 0.4230      | 0.2876  | 1.0000    | 0.1977     | 0.2683   | 0.2905  | 0.4387   |
| Brazil   | 0.3608    | 0.4462      | 0.5417  | 0.1977    | 1.0000     | 0.6419   | 0.5184  | 0.4440   |
| Mexico   | 0.4618    | 0.5000      | 0.6306  | 0.2863    | 0.6419     | 1.0000   | 0.6156  | 0.6061   |
| Colombia | 0.6069    | 0.5182      | 0.6687  | 0.2905    | 0.5184     | 0.6156   | 1.0000  | 0.6524   |
| Peru     | 0.8030    | 0.5910      | 0.6310  | 0.4387    | 0.4440     | 0.6061   | 0.6524  | 1.0000   |

Table 12. PCA variance proportions with real exchange currency data (1/3/2013 to 10/6/2014).

| Country | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 | Comp.6 | Comp.7 | Comp.8 |
|---------|--------|--------|--------|--------|--------|--------|--------|--------|
| America | 2.1624 | 1.0333 | 0.7731 | 0.6997 | 0.5973 | 0.5771 | 0.5700 | 0.3931 |
| Proportion of Variance | 0.5845 | 0.1335 | 0.0747 | 0.0612 | 0.0446 | 0.0416 | 0.0406 | 0.0193 |
| Cumulative Proportion | 0.5845 | 0.7179 | 0.7926 | 0.8538 | 0.8984 | 0.9401 | 0.9807 | 1.0000 |
| Asia    | 2.1742 | 0.9441 | 0.8996 | 0.7510 | 0.6311 | 0.5154 | 0.4917 | 0.3202 |
| Proportion of Variance | 0.5909 | 0.1114 | 0.1012 | 0.0705 | 0.0498 | 0.0332 | 0.0302 | 0.0128 |
| Cumulative Proportion | 0.5909 | 0.7023 | 0.8035 | 0.8740 | 0.9238 | 0.9570 | 0.9872 | 1.0000 |
| Europe  | 1.4692 | 1.1665 | 1.0544 | 1.0216 | 0.9422 | 0.7560 | 0.6869 | 0.6277 |
| Proportion of Variance | 0.2698 | 0.1701 | 0.1390 | 0.1305 | 0.1110 | 0.0714 | 0.0590 | 0.0492 |
| Cumulative Proportion | 0.2698 | 0.4399 | 0.5789 | 0.7094 | 0.8203 | 0.8918 | 0.9508 | 1.0000 |
Table 13. PCA variance proportions with copulas of real data (1/3/2013 to 10/6/2014).

| Copula                          | Comp.1 | Comp.2 | Comp.3 | Comp.4 | Comp.5 | Comp.6 | Comp.7 | Comp.8 |
|--------------------------------|--------|--------|--------|--------|--------|--------|--------|--------|
| t-Copula                       |        |        |        |        |        |        |        |        |
| $F(\text{Europe} | \text{Asia})$ | 0.2460 | 0.1690 | 0.1360 | 0.1280 | 0.1090 | 0.0860 | 0.0690 | 0.0570 |
| $F(\text{America} | (\text{Europe, Asia}))$ | 0.3080 | 0.1630 | 0.1300 | 0.1030 | 0.0900 | 0.0770 | 0.0700 | 0.0590 |
| Gaussian Copula                |        |        |        |        |        |        |        |        |
| $F(\text{Europe} | \text{Asia})$ | 0.2460 | 0.1710 | 0.1360 | 0.1280 | 0.1090 | 0.0850 | 0.0690 | 0.0550 |
| $F(\text{America} | (\text{Europe, Asia}))$ | 0.3190 | 0.1600 | 0.1310 | 0.1000 | 0.0890 | 0.0760 | 0.0680 | 0.0560 |
| Frank Copula                   |        |        |        |        |        |        |        |        |
| $F(\text{Europe} | \text{Asia})$ | 0.2470 | 0.1720 | 0.1360 | 0.1270 | 0.1090 | 0.0850 | 0.0690 | 0.0550 |
| $F(\text{America} | (\text{Europe, Asia}))$ | 0.3180 | 0.1590 | 0.1310 | 0.1000 | 0.0900 | 0.0760 | 0.0690 | 0.0570 |

Table 14. Change point detection by ETCC and NPMVCP with real exchange currency data (1/3/2013 to 10/6/2014).

| Method | Asia | Europe | America |
|--------|------|--------|---------|
| ETCC   | 347  | 106 340| 103 353 |
| NPMVCP | 306  | 106 304| 103 302 |
| ETCC   | 344  | 109 341| 103 127 342|
| NPMVCP | 175  | 106 304| 103 126 302|

Table 15. Change point detection with the James, Zhang, and Matteson (2019) nonparametric multiple change point analysis of real exchange currency data (1/3/2013 to 10/6/2014).

| Change Points | Asia | Europe | America |
|---------------|------|--------|---------|
| 53 83 128 166 231 273 310 359 396 | 36 96 126 163 227 257 287 329 390 | 41 71 107 146 200 254 284 314 357 388 |

Figure 6. FPCA plots of Asia (1/3/2013 to 10/6/2014).
Table 16. Change point detection by ETCC and NPMVCP of FPCA with real exchange currency data (1/3/2013 to 10/6/2014).

|                | ETCC                  | NPMVCP               |
|----------------|-----------------------|----------------------|
| **Asia**       | 21 41 62 83 103 123 143 164 184 205 225 246 267 287 308 328 349 369 390 411 433 | 20 41 62 82 102 122 143 163 184 204 225 246 266 287 307 328 348 369 390 410 433 |
| **Europe**     | 21 41 61 82 103 124 145 165 185 205 226 246 267 286 307 327 347 367 388 409 430 | 20 40 61 82 103 124 144 164 184 205 225 245 265 286 306 326 346 367 388 409 429 |
| **America**    | 21 41 64 85 106 126 146 166 186 206 227 248 269 290 310 331 351 372 393 414 369 390 | 20 40 64 85 105 125 145 165 185 206 227 248 269 289 310 330 351 372 393 413 |

Table 17. Change point detection by ETCC and NPMVCP of PCA and copulas with real exchange currency data (1/3/2013 to 10/6/2014).

|                     | F(Europe/Asia) | F(America|Europe,Asia) |
|---------------------|---------------|------------|
| **Method**          | ETCC          | NPMVCP     | ETCC          | NPMVCP     |
| t-Copula            | 0             | 0          | 103           | 103         |
| Gaussian Copula     | 0             | 0          | 103           | 103         |
| Frank Copula        | 63            | 57         | 42 103        | 31 103      |
Table 18. Change point detection by ETCC and NPMVCP of FPCA and copulas with real exchange currency data (1/3/2013 to 10/6/2014).

| Copula Family | Copula Type | F(Europe/Asia) | ETCC | F(America/(Europe, Asia)) |
|---------------|-------------|----------------|------|--------------------------|
|               | Functional PCA by t-Copula |               |      |                          |
|               | ETCC        | 11 17 25 31 40 47 52 60 70 78 88 94 99 106 114 122 131 140 142 150 157 166 173 178 185 192 199 206 214 | 221 231 238 243 252 259 264 270 277 284 291 298 305 312 322 328 337 343 349 356 362 367 372 378 384 390 395 403 405 411 417 425 431 |
|               | NPMVCP      | 11 20 27 34 43 52 60 67 73 79 86 93 101 106 113 121 130 138 148 156 163 171 179 187 194 202 211 217 226 | 235 243 252 262 270 276 286 295 303 314 322 331 340 350 357 367 376 386 395 401 408 413 422 427 |
|               | Functional PCA by Gaussian Copula |               |      |                          |
|               | ETCC        | 11 17 26 32 39 50 60 70 79 87 95 100 107 114 122 131 137 143 150 158 165 172 178 186 192 198 206 213 221 | 228 236 242 249 253 262 268 275 282 288 295 302 310 319 327 336 341 348 353 359 364 370 375 381 387 393 402 409 416 424 429 |
|               | NPMVCP      | 9 16 21 30 39 45 50 58 69 77 85 92 98 104 112 120 130 135 141 148 156 163 169 176 183 190 197 204 212 219 | 222 236 242 249 253 262 268 275 282 288 295 302 310 319 327 336 341 348 353 359 364 370 375 381 387 393 402 409 416 424 429 |
|               | Functional PCA by Frank Copula |               |      |                          |
|               | ETCC        | 11 18 24 31 41 49 57 63 71 79 88 95 101 107 114 122 130 138 143 150 158 166 170 177 186 192 199 207 214 221 | 229 237 243 251 256 263 269 277 285 290 297 303 312 321 329 337 349 355 361 366 372 378 382 390 395 403 411 417 425 432 |
|               | NPMVCP      | 9 16 21 30 39 45 50 58 69 77 85 92 98 104 112 120 130 135 141 148 156 163 169 176 183 190 197 204 212 219 | 222 236 242 249 253 262 268 275 282 288 295 302 310 319 327 336 341 348 353 359 364 370 375 381 387 393 402 409 416 424 429 |
|               | Functional PCA by Frank Copula |               |      |                          |
|               | ETCC        | 11 18 26 33 42 50 58 65 72 77 85 91 99 105 111 119 128 136 146 155 160 169 178 184 192 201 208 216 224 | 233 241 249 256 266 274 285 292 302 310 322 332 340 350 357 367 376 386 395 401 408 413 422 427 |
|               | NPMVCP      | 9 16 21 30 39 45 50 58 69 77 85 92 98 104 112 120 130 135 141 148 156 163 169 176 183 190 197 204 212 219 | 222 236 242 249 253 262 268 275 282 288 295 302 310 319 327 336 341 348 353 359 364 370 375 381 387 393 402 409 416 424 429 |
Table 19. Change point detection by ETCC and NPMVCP with real exchange currency data of America (1/3/2013 to 10/3/2014) and Asia (1/4/2013 to 10/6/2014).

| Data          | America | Asia          |
|---------------|---------|---------------|
| Method        | ETCC    | NPMVCP        |
| Whole Data    | 103 127 345 | 103 126 302  |
| PCA Scores    | 106 125 147 354 | 104 124 144 302  |

Table 20. Change point detection with the James, Zhang, and Matteson (2019) nonparametric multiple change point analysis with real exchange currency data of America (1/3/2013 to 10/3/2014) and Asia (1/4/2013 to 10/6/2014).

| Data          | America | Asia          |
|---------------|---------|---------------|
| Change Points | 58 95 125 188 223 280 314 402 | 35 90 137 168 198 245 295 326 356 399 |

Figure 9. FPCA plots of America (1/3/2013 to 10/3/2014).

Figure 10. FPCA plots of Asia (1/4/2013 to 10/6/2014).
Table 21. Change point detection by ETCC and NPMVCP of FPCA with real exchange currency data of America (1/3/2013 to 10/3/2014) and Asia (1/4/2013 to 10/6/2014).

|            | FPCA         | ETCC         | NPMVCP       |
|------------|--------------|--------------|--------------|
| America    | 21 41 65 86 107 128 149 169 189 209 230 251 271 292 312 333 353 374 394 415 | 21 41 62 83 103 123 145 166 187 208 228 248 268 288 309 330 351 372 393 414 | 20 41 65 86 107 128 148 168 188 209 230 250 271 291 312 332 353 373 394 414 |
| Asia       | 21 41 65 86 107 128 149 169 189 209 230 251 271 292 312 333 353 374 394 415 | 21 41 62 83 103 123 145 166 187 208 228 248 268 288 309 330 351 372 393 414 | 20 41 65 86 107 128 148 168 188 209 230 250 271 291 312 332 353 373 394 414 |

Table 22. Change point detection by ETCC and NPMVCP of PCA and copulas with real exchange currency data of America (1/3/2013 to 10/3/2014) and Asia (1/4/2013 to 10/6/2014).

| Method        | PCA Scores | F(Asia|America) | ETCC      | NPMVCP |
|---------------|------------|-----------|-----------|----------|
| t-Copula      | 362        | 302       |           |          |
| Gaussian Copula| 377        | 302       |           |          |
| Frank Copula  | 377        | 302       |           |          |

Table 23. Change point detection by ETCC and NPMVCP of FPCA and copulas with real exchange currency data of America (1/3/2013 to 10/3/2014) and Asia (1/4/2013 to 10/6/2014).

| FPCA by t-Copula for F(Asia|America) | ETCC      | NPMVCP |
|---------------------------|----------|----------|
|                           | 21 42 62 83 104 124 145 166 187 208 228 248 268 288 309 329 350 371 392 413 | 21 41 62 83 103 124 145 166 187 207 227 247 267 287 308 329 350 371 392 412 |
| FPCA by Gaussian Copula for F(Asia|America) | ETCC      | NPMVCP |
|                           | 20 40 61 82 102 123 145 166 187 208 228 248 268 288 309 330 351 372 393 414 | 20 40 61 82 102 123 145 166 187 207 227 247 267 287 308 330 351 372 393 414 |
| FPCA by Frank Copula for F(Asia|America) | ETCC      | NPMVCP |
|                           | 21 41 61 82 103 123 145 166 187 208 228 248 268 288 309 330 351 372 393 414 | 20 40 61 82 102 123 145 166 186 206 226 246 266 286 308 329 350 371 392 412 |

5. Conclusions

We proposed the conditional multi-stage multivariate CPD method by employing PCA or FPCA, copula conditional distribution, and the multivariate CPD models, which are energy test-based control chart (ETCC) and the nonparametric multivariate change point model (NPMVCP). With a simulation study and real data analysis, we showed that our proposed conditional multi-stage multivariate CPD method based PCA and FPCA is useful for detecting change points in the case of a multi-stage sequential process. Furthermore, we can conclude that the FPCA-based conditional multi-stage multivariate CPD method detects more change points compared to the PCA-based conditional multi-stage multivariate CPD method. Future study will employ FPCA with different types of bases to compare Fourier-based FPCA for multi-stage multivariate CPD and also develop a neural network-based multi-stage multivariate CPD method.

Author Contributions: J.-M.K. designed the model, analyzed the data and wrote the paper. N.W. proposed the idea of this paper, formulated the conceptual framework, designed the model, obtained inference and wrote the paper. Y.L. supervised this research, formulated the conceptual framework, designed the model, obtained inference and wrote the paper. All the authors cooperated to revise the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China Grant (No. 71672182, No. U1604262 and No. U1904211) and National Social Science Fund of China (No. 20BTJ059).

Conflicts of Interest: The authors declare no conflict of interest.
References

1. Qiu, P. Statistical process control charts as a tool for analyzing big data. In *Big and Complex Data Analysis: Statistical Methodologies and Applications*; Ahmem, E., Ed.; Springer: New York, NY, USA, 2017; pp. 123–138.

2. Qiu, P. Some perspectives on nonparametric statistical process control. *J. Q. Technol.* 2018, 50, 49–65. [CrossRef]

3. Qiu, P.; Hawkins, D. A rank-based multivariate cusum procedure. *Technometrics* 2001, 43, 120–132. [CrossRef]

4. Qiu, P.; Hawkins, D. A nonparametric multivariate cumulative sum procedure for detecting shifts in all directions. *J. R. Stat. Soc. Ser. D Stat.* 2003, 52, 151–164. [CrossRef]

5. Ross, G.J.; Tasoulis, D.K.; Adams, N.M. Nonparametric monitoring of data streams for changes in location and scale. *Technometrics* 2011, 53, 379–389. [CrossRef]

6. Holl, M.; Hawkins, D. A control chart based on a nonparametric multivariate change-point model. *J. Q. Technol.* 2014, 46, 1975–1987.

7. Okhrin, O.; Xu, Y.F. A Nonparametric Multivariate Control Chart for High-Dimensional Financial Surveillance. 2017, Submitted under review.

8. Matteson, D.S.; James, N.A. A nonparametric approach for multiple change point analysis of multivariate data. *J. Am. Stat. Assoc.* 2014, 109, 334–345. [CrossRef]

9. James, N.A.; Matteson, D.S. *ecp*: An R Package for Nonparametric Multiple Change Point Analysis of Multivariate Data. *J. Stat. Softw.* 2014, 62, 1–25. [CrossRef]

10. James, N.A.; Zhang, W.; Matteson, D.S. *ecp*: An R Package for Nonparametric Multiple Change Point Analysis of Multivariate Data. *R package*, version 3.1.2; 2019. Available online: https://cran.r-project.org/web/packages/ecp/index.html (accessed on 22 August 2019).

11. Hsu, H.-L.; Ing, C.-K.; Lai, T.L.; Yu, S.-H. Multistage Manufacturing Processes: Innovations in Statistical Modeling and Inference. In *Proceedings of the Pacific Rim Statistical Conference for Production Engineering*; ICSA Book Series in Statistics; Springer: Singapore, 2018; pp. 67–84.

12. Qiu, P.; You, L. Recent Research in Dynamic Screening System for Sequential Process Monitoring. In *Proceedings of the Pacific Rim Statistical Conference for Production Engineering*; ICSA Book Series in Statistics; Springer Publisher: Singapore, 2018; pp. 85–94.

13. Emura, T.; Long, T.-H.; Sun, L.-H. R routines for performing estimation and statistical process control under copula-based time series models. *Commun. Stat. Simul. Comput.* 2017, 46, 3067–3087. [CrossRef]

14. Kim, J.-M.; Baik, J.; Reller, M. Control charts of mean and variance using copula Markov SPC and conditional distribution by copula. *Commun. Stat. Simul. Comput.* 2020, In Press. [CrossRef]

15. Joe, H. *Multivariate Models and Multivariate Dependence Concepts*; CRC Press: Boca Raton, FL, USA, 1997.

16. Park, K.; Kim, J.-M.; Jung, D. GLM-based statistical control r-charts for dispersed count data with multicollinearity between input variables. *Q. Reliab. Eng. Int.* 2018, 34, 1103–1109. [CrossRef]

17. Pearson, K. On Lines and Planes of Closest Fit to System of Points in Space. *Philos. Mag.* 1901, 2, 559–572. [CrossRef]

18. Chen, Y.; Carroll, C.; Dai, X.; Fan, J.; Hadjipantelis, P.Z.; Han, K.; Ji, H.; Lin, S.-C.; Dubey, P.; Mueller, H.-G.; et al. *Fdapace: Functional Data Analysis and Empirical Dynamics*; R Package; 2019. Available online: https://cran.r-project.org/web/packages/fdapace/index.html (accessed on 17 August 2019).

19. Liu, B.; Müller, H.-G. Estimating Derivatives for Samples of Sparsely Observed Functions, with Application to Online Auction Dynamics. *J. Am. Stat. Assoc.* 2009, 104, 704–717. [CrossRef]

20. Yao, F.; Müller, H.-G.; Wang, J.-L. Functional Data Analysis for Sparse Longitudinal Data. *J. Am. Stat. Assoc.* 2005, 100, 577–590. [CrossRef]

21. Sklar, A. Fonctions de repartition á n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* 1959, 8, 229–231.

22. Nelsen, R.B. *An Introduction to Copulas*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2013.

23. Demarta, S.; McNeil, A.J. The t copula and related copulas. *Int. Stat. Rev.* 2005, 73, 111–129. [CrossRef]

24. Hawkins, D.M.; Qiu, P.; Kang, C.W. The changepoint model for statistical process control. *J. Q. Technol.* 2003, 35, 355–366. [CrossRef]

25. Székely, G.J.; Rizzo, M.L. Hierarchical clustering via joint between-within distances: Extending Ward’s minimum variance method. *J. Classif.* 2005, 22, 151–183.
26. Xu, Y.F. Reference Manual: An R package ‘EnergyOnlineCPM’. 2017. Available online: https://sites.google.com/site/EnergyOnlineCPM/ (accessed on 19 March 2020).
27. Choi, K.; Marden, J. An Approach to Multivariate Rank Tests in Multivariate Analysis of Variance. J. Am. Stat. Assoc. 1997, 92, 1581–1590. [CrossRef]
28. Kim, J.-M.; Hwang, S.Y. Directional Dependence via Gaussian Copula Beta Regression Model with Asymmetric GARCH Marginals. Commun. Stat. Simul. Comput. 2017, 46, 7639–7653. [CrossRef]
29. Brechmann, E.C.; Schepsmeier, U. Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine. J. Stat. Softw. 2013, 52, 1–27. [CrossRef]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.