THREE ESSAYS ON NONPARAMETRIC ECONOMETRICS WITH APPLICATIONS TO FINANCIAL ECONOMICS AND INSURANCE

A Dissertation

by

KUANGYU WEN

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	Ximing Wu
Committee Members,	David Bessler
	David Leatham
	Qi Li
Head of Department,	C. Parr Rosson III

May 2015

Major Subject: Agricultural Economics

Copyright 2015 Kuangyu Wen

ABSTRACT

This dissertation includes three essays. The first essay concerns nonparametric kernel density estimation on the unit interval. The Kernel Density Estimator (KDE) suffers boundary biases when applied to densities on bounded supports, which are assumed to be the unit interval. Transformations mapping the unit interval to the real line can be used to remove boundary biases. However, this approach may induce erratic tail behaviors when the estimated density of transformed data is transformed back to its original scale. I propose a modified transformation based KDE that employs a tapered and tilted back-transformation. I derive the theoretical properties of the new estimator and show that it asymptotically dominates the naive transformation based estimator while maintains its simplicity. I then propose three automatic methods of smoothing parameter selection. Monte Carlo simulations demonstrate the good finite sample performance of the proposed estimator, especially for densities with poles near the boundaries. An example with real data is provided.

The second essay proposes a new kernel estimator of copula densities. The standard kernel estimator suffers boundary biases since copula densities are defined on a bounded support and often tend to infinity on the boundaries. A transformation-based estimator aptly remedies both boundary biases and inconsistencies due to unbounded densities. This method, however, might entail undesirable boundary behaviors due to an unbounded multiplicative factor associated with the transformation. I propose a modified transformation-based estimator that employs an infinitesimal tapering device to mitigate the influence of the unbounded multiplier. I establish the asymptotic properties of our estimator and show that it dominates the original transformation estimator in terms of mean squared error due to bias correction. I present

two practically simple methods of smoothing parameter selection. I further show that the proposed estimator admits higher order bias reduction for Gaussian copulas and provides outstanding performance for Gaussian and near Gaussian copulas. This appealing feature makes our estimator particularly suitable for financial data analyses. Extensive simulations corroborate our theoretical analysis and demonstrate outstanding performance of the proposed method relative to competing estimators. Three empirical applications are provided.

The third essay studies nonparametric estimation of crop yield distributions and crop insurance premium rates. Since U.S. crop yield data are typically available at county level for only a few decades, nonparametric estimation of yield distribution for individual counties suffers from small sample sizes. The fact that nearby counties share similarities in their yield distributions suggests possible efficiency gains through information pooling. I propose a weighted kernel density estimator subject to selected spatial moment restrictions. The weights are calculated using the method of empirical likelihood and the spatial moments are specified based on the consideration of flexibility and robustness. I further extend the proposed method to the adaptive kernel density estimation. My simulations demonstrate the outstanding performance of the proposed methods in the estimation of crop yield distributions and that of crop insurance premium rates. I apply these methods to estimate corn yield distributions and crop insurance premium rates for the ninety-nine counties in Iowa.

DEDICATION

First and most importantly, I want to thank my committee chair, Dr. Ximing Wu. He made it possible for me to write this dissertation and stood by my side while I was entering the world of econometrics. I want to thank Dr. Wu for all the hours we spent in his office discussing all the questions I had, correcting my mistakes, making improvements and sharing his ideas with me. I also want to thank him for all the opportunities he gave me during the past five years. Without his supervision, this dissertation would not have been completed.

I also want to express my gratitude to Dr. David Bessler, Dr. David Leatham and Dr. Qi Li for their continuous support, enlightenment and encouragement. I thank them for reading previous drafts of this dissertation and providing valuable comments that improved the contents of this dissertation.

NOMENCLATURE

iid	identically and independently distributed
cdf	cumulative distribution function
pdf	probability density function
KDE	standard Kernel Density Estimator
TKDE/TKE	Transformation-based Kernel Density Estimator
MTK	Modified Transformation-based Kernel Density Estimator
WMISE	Weighted Mean Integrated Squared Error
CV	Cross Validation
GRP	Group Risk Plan
NASS	National Agricultural Statistics Services
EL	Empirical Likelihood
ELK	Empirical Likelihood Kernel Density Estimator
AKDE	Adaptive Kernel Density Estimator
ELAK	Empirical Likelihood Adaptive Kernel Density Estimator

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

Econometric problems are described using probability models. The observed data are viewed as a realization of a random vector associated with some true distribution. To understand the economic data, one often needs to identify this true distribution that governs the underlying data generating process. Two approaches have been commonly adopted for estimating econometric models. One is called parametric method. That is, we assume the true distribution comes from some parametric family with a finite dimensional vector of unknown parameters. Then the problem becomes seeking the value of these parameters. Standard routines such as Maximum Likelihood Estimation and Generalized Method of Moments can be used for estimation. However, this parametric method is inconsistent if the parametric assumption is misspecified. Motivated primarily by the problem of misspecification error, the other approach called nonparametric method serves as a complement. Nonparametric method is consistent under some minimal assumptions such as smoothness. Since nonparametric method does not require a parametric assumption, it is flexible, though it takes the risk of slower convergence rate. Nowadays, nonparametric method has gained its popularity not only in econometrics itself but also in finance and insurance applications. Therefore, in this dissertation, I propose some nonparametric estimators for general purpose as well as for some specific cases in finance and insurance.

The three essays in this dissertation are motivated by some interesting finance and insurance applications. The first example is about recovery rate on defaulted bonds or loans. In credit risk portfolio models, there are two key elements: one is probability of default and the other is recovery probability given default. Therefore, estimating the density function of recovery rate is practically important. Recovery

rate is expressed in the form of a percentage ratio. Thus, its density function is defined on the support $[0, 1]$. This example motivates the first essay on the topic of nonparametric kernel density estimation on a bounded support. The second example is the role of copula in risk management. Since the last decade, copula has become a powerful tool in modeling losses of a portfolio of multiple securities. The parametric Gaussian copula was once dominant in pricing Collateralized Debt Obligations (CDOs). However, after the 2007 credit crisis, both researchers and practitioners realized that Gaussian copula might not be adequate to describe the dependence among financial returns. Nonparametric copula estimation arises naturally as a more flexible and reliable alternative. The third example concerns estimating crop yield distribution and crop insurance premium rate. Crop insurance program is one of the most expensive federal policies. Actuarially fair premium rate is key to the welfare of farmers, vitality of the crop insurance industry and efficiency of the crop insurance program. The literature had a long time debate on the most suitable parametric family for yield distribution, yet, reached no agreement. Some studies opted to using nonparametric method for yield density estimation. However, a practical problem is that yield data is usually limited. This motivates the third essay to aim at improving the accuracy of nonparametric yield density estimation when sample size is small.

This dissertation tries to propose some nonparametric kernel estimators to address the above motivating applications. In the first essay, I propose a new kernel estimator to estimate densities with bounded support $[0, 1]$. The second essay proposes a new kernel estimator of copula densities. The second essay can be viewed as an extension of the ideas in the first essay. Extensive simulation experiments confirm both estimators have better overall performance than the existing competing estimators. The third essay proposes a weighted kernel density estimator subject to selected spatial moment restrictions. This method is shown to improve the accuracy

of estimated premium rate.

In the rest of text, each essay will form a separate chapter. Some technical proofs are gathered in the appendices.

2. AN IMPROVED TRANSFORMATION-BASED KERNEL ESTIMATOR OF DENSITIES ON THE UNIT INTERVAL*

2.1 Introduction

This paper concerns with kernel type estimation of densities on bounded supports, which without loss of generality are assumed to be $[0, 1]$. Given a sample $\{X_1, \ldots, X_n\}$ of observations from a univariate distribution F_X with a density f_X, the standard kernel density estimator (KDE) is given by

$$\hat{f}_X(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) = \frac{1}{n} \sum_{i=1}^{n} K_h\left(x - X_i\right),$$

where K is the kernel function usually taken to be a symmetric and unimodal density defined on $\mathbb{R}$ or some finite interval, $K_h(x) = K(x/h)/h$ and h is the bandwidth. When applied to densities with bounded supports, the standard KDE is known to suffer boundary biases since it cannot detect the boundaries of the support and thus places positive weight outside the support. For general treatments of the KDE and the boundary bias problem, see e.g. Wand and Jones (1995) and Simonoff (1996).

Various boundary bias correction methods have been proposed. Schuster (1999) and Cline and Hart (1991) considered the reflection method, which is most suitable for densities with zero derivatives near the boundaries. Boundary kernel method and local polynomial method are more general without restrictions on the shape of densities. Local polynomial method can be seen as a special case of boundary kernel method and draws much attention due to its good theoretical properties. Though early versions of these methods might produce negative estimates or inflate

*This dissertation is derived in part from an article published in the Journal of the American Statistical Association, available online:
http://www.tandfonline.com/doi/abs/10.1080/01621459.2014.969426#.VSZ_IPD_OCg.

variance near the boundaries, remedies and refinements have been proposed; see e.g. Müller (1991), Jones (1993), Jones and Foster (1996), Cheng (1997), Zhang and Karunamuni (1998, 2000) and Karunamuni and Alberts (2005). Cowling and Hall (1996) proposed a pseudo-data method that estimates density functions based on the original data plus pseudo-data generated by linear interpolation of order statistics. Zhang, Karunamuni, and Jones (1999) combined the pseudo-data, transformation and reflection methods.

A second strand of the literature eschews explicit boundary correction and utilizes instead globally modified kernel estimators. Marron and Ruppert (1994) considered transformations via the empirical distribution function. Hall and Park (2002) presented an alternative "empirical translation correction" method. Chen (1999) proposed beta kernel estimators that use the beta density as the kernel function. Jones and Henderson (2007) presented a Gaussian copula based estimator. Geenens (2014) combined transformation and local likelihood estimation. These estimators share a commonality of employing effectively varying local bandwidths induced by the transformation of data or flexible kernel functions.

In this paper I adopt the transformation approach and propose a new estimator that is shown to improve upon the conventional transformation estimator. Transformation-based kernel density estimation (TKDE) was originally proposed by Wand, Marron, and Ruppert (1991). Let g be some smooth and monotonically increasing function and define $Y_i = g(X_i)$. Denote the density of Y_i by f_Y, which can be estimated by the standard KDE $\hat{f}_Y(y) = \frac{1}{n}\sum_{i=1}^{n} K_h\left(y - Y_i\right)$. The transformation-based estimator of f_X is then obtained, via a back-transformation, as

$$\hat{f}_{X,T}(x) = \frac{1}{n}\sum_{i=1}^{n} g'(x) K_h\left(g(x) - g(X_i)\right), \tag{2.1}$$

where $g'(x) = \partial g(x)/\partial x$. With a proper transformation function, f_Y may be satis-

factorily estimated by the standard KDE and this benefit is likely to be retained in the subsequent estimator $\hat{f}_{X,T}$ in the X-domain.

When estimating a density defined on $[0,1]$, if we use a transformation mapping the unit interval to the real line, the density of the transformed data Y_i's can be estimated by the standard KDE free of boundary biases. The original density f_X, obtained via a back-transformation of f_Y, can then be estimated without boundary biases as well. A commonly used family of transformations from the unit interval to the real line is the quantile functions of distributions defined on the real line. Similar to Geenens (2014), I consider the Probit transformation $g(x) = \Phi^{-1}(x)$, where Φ is the standard normal distribution function. I have experimented with transformation via the logistic distribution and obtained results essentially identical to those from the Probit transformation. In this paper, I focus on the Probit transformation because when used in conjunction with the Gaussian kernel function, it permits a simple analytical form for the proposed estimator, which I shall show below (see Chaudhuri and Marron (2000) for a discussion on many appealing properties of the Gaussian kernel).

Under the Probit transformation, the TKDE (2.1) for a density defined on $[0,1]$ takes the form

$$\hat{f}_{X,T}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{K_h\left(\Phi^{-1}(x) - \Phi^{-1}(X_i)\right)}{\phi(\Phi^{-1}(x))}, x \in [0,1], \tag{2.2}$$

where ϕ is the standard normal density function. As x approaches 0 or 1, the multiplication factor $\{\phi(\Phi^{-1}(x))\}^{-1}$ tends to infinity, resulting in possibly erratic tail behaviors in $\hat{f}_{X,T}(x)$. My new estimator is inspired by the observation that the drawback associated with the multiplication factor can be alleviated by tapering the multiplication factor when x is close to the boundaries. Denote the Gaussian density with mean μ and variance σ^2 by $\phi_{\mu,\sigma}$. I shall suppress the subscripts μ

and/or σ if $\mu = 0$ and/or $\sigma = 1$ for simplicity. I note that one natural way to deflate $\{\phi(\Phi^{-1}(x))\}^{-1}$ for x near 0 or 1 is to replace it with $\{\phi_\sigma(\Phi^{-1}(x))\}^{-1}$, where $\sigma > 1$. In addition, it is desirable that this tapering device is allowed to differ across the two tails if the underlying densities have asymmetric tails. This is made possible by further introducing a location parameter μ, resulting in a tapered and tilted back-transformation $\{\phi_{\mu,\sigma}(\Phi^{-1}(x))\}^{-1}$ with two tuning parameters μ and σ.

The modified transformation-based kernel density estimation, which is called "MTK" below for simplicity, is formally defined as

$$\hat{f}_{X,M}(x) = \frac{1}{n\hat{c}} \sum_{i=1}^{n} \frac{K_h\left(\Phi^{-1}(x) - \Phi^{-1}(X_i)\right)}{\phi_{\mu,\sigma}(\Phi^{-1}(x))}, x \in [0,1], \tag{2.3}$$

where $\hat{c}$ is a normalization constant, which admits a simple closed form under the Probit transformation. I derive the theoretical properties of the MTK and propose three automatic methods of smoothing parameter selection. I demonstrate the good finite sample performance of the MTK via numerical simulations and present an application to real world data.

The proposed MTK for densities with bounded supports possesses several desirable properties. First, like the TKDE and several other boundary bias corrected estimators, it is free of boundary biases. At the same time, it is shown to dominate the TKDE in terms of the asymptotic mean integrated squared error. Second, some methods of boundary bias correction require complicated data-driven transformations or local bandwidths. In contrast, the MTK uses a fixed transformation and maintains the simplicity of the TKDE with a single global bandwidth. Third, the MTK produces a bona fide density that is non-negative and integrates to unity. Fourth, with the proposed automatic smoothing parameter selection methods, it is easy to implement and computationally inexpensive. Lastly, the MTK is shown to significantly outperform the TKDE and several other competing estimators for

densities with poles near the boundaries.

The rest of the text is organized as follows. Section 2.2 presents the modified transformation-based kernel density estimator, followed by its theoretical properties in Section 2.3. Section 2.4 proposes three methods of automatic smoothing parameter selection. Sections 2.5 and 2.6 report Monte Carlo simulations and an example on real data. Some technical details are gathered in Appendix A.

2.2 Modified transformation-based kernel estimator

Using transformation in kernel type density estimations has a long history in the literature. The standard KDE is marked by a single global bandwidth, which may not be suitable for densities with varying degree of roughness or complicated features. Wand, Marron, and Ruppert (1991) proposed the transformation based kernel density estimator given in (2.1). With a proper transformation function $Y = g(X)$, f_Y may be satisfactorily estimated by the standard KDE $\hat{f}_Y$ and this benefit can be retained by the subsequent estimator $\hat{f}_{X,T}$ in the X-domain. For instance, Wand, Marron, and Ruppert (1991) applied the shifted-power transformation to skewed data, which is shown to improve the subsequent kernel density estimation of the original data. This transformation approach in kernel density estimation has been further developed by, among others, Marron and Ruppert (1994), Ruppert and Cline (1994), Yang and Marron (1999), Karunamuni and Alberts (2006), Koekemoer and Swanepoel (2008), and Geenens (2014).

Under a transformation that maps the unit interval to the real line, the TKDE (2.2) provides a viable way to alleviate boundary biases in kernel density estimation as discussed above. To understand the essence of this approach, consider a given $x \in (0,1)$. A Taylor expansion of $\Phi^{-1}(X_i)$ around x in (2.2) yields $\hat{f}_{X,T} \approx n^{-1}\sum_{i=1}^{n}\{h\phi(\Phi^{-1}(x))\}^{-1}K\{(x-X_i)/(h\phi(\Phi^{-1}(x)))\}$. Thus the TKDE behaves sim-

ilarly to a KDE with a local bandwidth $h(x) = h\phi(\Phi^{-1}(x))$. Since $\phi(\Phi^{-1}(x)) \to 0$ as $x \to 0$ or 1, the effective bandwidth near the boundaries becomes increasingly small and no smoothing goes beyond the boundaries.

The transformation approach for kernel estimation of densities with bounded supports, however, suffers from one particular drawback. Since the multiplication factor $\{\phi(\Phi^{-1}(x))\}^{-1}$ in the TKDE $\hat{f}_{X,T}(x)$ becomes increasingly large as x approaches the boundaries, small bumps at the tails of $\hat{f}_Y$ may be magnified through the back-transformation, resulting in erratic fluctuations of $\hat{f}_{X,T}$ near the boundaries. Geenens (2014) proposed a local log-polynomial estimator of the TKDE for densities with bounded supports. In this study, I propose an alternative estimator that entails simple modification of the back-transformation in the TKDE. The Modified Transformation-based Kernel Density Estimator (MTK), given in (2.3), introduces a multiplicative bias reduction of the TKDE while maintaining the simplicity of the TKDE with a single bandwidth.

To ensure that the MTK integrates to one, the normalization constant $\hat{c}$ is given by

$$\hat{c} = \frac{1}{n} \int_0^1 \sum_{i=1}^{n} \frac{K_h\left(\Phi^{-1}(x) - \Phi^{-1}(X_i)\right)}{\phi_{\mu,\sigma}(\Phi^{-1}(x))} dx.$$

For small h, $\hat{c}$ can be closely approximated by

$$\tilde{c} = \frac{1}{n} \sum_{i=1}^{n} \frac{\phi(\Phi^{-1}(X_i))}{\phi_{\mu,\sigma}(\Phi^{-1}(X_i))}.$$

Furthermore if I use the Gaussian kernel, i.e. setting $K(\cdot) = \phi(\cdot)$, the normalization constant admits a simple analytical form

$$\hat{c} = \frac{\sigma^2}{n\eta} \sum_{i=1}^{n} \exp\left(-\frac{\sigma^2 - 1}{2\eta^2} \{\Phi^{-1}(X_i)\}^2 - \frac{\mu}{\eta^2} \Phi^{-1}(X_i) + \frac{\mu^2(h^2+1)}{2\eta^2} \right),$$

where $\eta = \sqrt{\sigma^2 + h^2\sigma^2 - h^2}$. For the rest of the text, I focus on the case of Gaussian

kernel estimator.

Next define

$$J(x;h,\mu,\sigma)=\frac{\phi(\Phi^{-1}(x))}{\hat{c}\phi_{\mu,\sigma}(\Phi^{-1}(x))}.$$

The MTK can be rewritten as

$$\hat{f}_{X,M}(x)=J(x;h,\mu,\sigma)\hat{f}_{X,T}(x). \tag{2.4}$$

Thus it is seen that the MTK introduces a multiplicative adjustment to the TKDE. The adjustment factor is controlled smoothly by two tuning parameters μ and σ. In particular, $\sigma(>1)$ tapers the multiplication factor of the TKDE and this tapering is further 'tilted' by μ. When $\mu=0$ and $\sigma=1$, $J(x;h,\mu,\sigma)=1$ and the MTK reduces to the TKDE.

2.3 Asymptotic properties

In this section I investigate the theoretical properties of the MTK. In addition to the usual conditions that $h\to 0$ and $nh\to\infty$ as $n\to\infty$, I also assume that $\mu\to 0$ and $\sigma\to 1$ as $n\to\infty$. The construction (2.4) facilitates our analysis: the multiplication factor $J(x;h,\mu,\sigma)$ is relatively simple and the properties of $\hat{f}_{X,T}$ have been well established in the literature. To ease notation, I shall denote $\Phi^{-1}(X_i)$ and $\Phi^{-1}(x)$ by Y_i and y_x respectively in this section.

2.3.1 Asymptotic bias and variance

Consider first a fixed $x\in(0,1)$. A Taylor expansion of $J(x;h,\mu,\sigma)$ with respect to μ at zero and σ at one yields

$$J(x;h,\mu,\sigma)=1+(E[Y_1]-y_x)\mu+(E[Y_1^2]-y_x^2)(\sigma-1)+o(\mu)+o(\sigma-1). \tag{2.5}$$

The asymptotic bias of the TKDE is given by Wand, Marron, and Ruppert (1991):

$$E[\hat{f}_{X,T}(x)] = f_X(x) + \frac{h^2}{2}\frac{f_Y''(y_x)}{\phi(y_x)} + o(h^2). \tag{2.6}$$

It follows that

$$\text{abias}\{\hat{f}_{X,M}(x)\} = \{(E[Y_1] - y_x)\mu + (E[Y_1^2] - y_x^2)(\sigma - 1)\}f_X(x) + \frac{h^2}{2}\frac{f_Y''(y_x)}{\phi(y_x)}. \tag{2.7}$$

Below I shall show that the optimal μ and $\sigma - 1$ are both of order h^2. Thus the optimal MTK has the usual h^2 order interior bias.

Here I offer a heuristic argument on how the MTK improves the boundary bias of the TKDE. I rewrite the asymptotic bias of the MTK as follows:

$$\begin{aligned}\text{abias}\{\hat{f}_{X,M}(x)\} =& \{(E[Y_1] - y_x)\mu + (E[Y_1^2] - y_x^2)(\sigma - 1)\}f_X(x) \\ &+ \frac{h^2}{2}\left\{f_X(x)(y_x^2 - 1) - 3f_X'(x)y_x\phi(y_x) + f_X''(x)\phi^2(y_x)\right\},\end{aligned} \tag{2.8}$$

where the second term is the asymptotic bias of the TKDE expressed in the X-domain density. Consider for example a U-shape density on $[0, 1]$ whose tails go to infinity. Since $y_x^2 \to \infty$ and $\phi(y_x) \to 0$ as $x \to 0$ or 1, the second term of (2.8) is dominated by $\frac{h^2}{2}f_X(x)(y_x^2 - 1)$, resulting in a positive bias of the TKDE for x near the boundaries. For $\sigma > 1$, the dominant term in the first term of (2.8) is $-y_x^2(\sigma - 1)f_X(x)$, which is negative and counters the positive boundary bias of the TKDE. Interestingly, the KDE is known to suffer negative boundary biases if the densities go to infinity near the boundaries. The TKDE, in contrast, overcompensates the boundary biases of the KDE because the multiplication factor $\{\phi(y_x)\}^{-1} \to \infty$ as $x \to 0$ or 1. The MTK fine-tunes the TKDE to correct for the overcompensation at the boundaries.

It is tempting to set $\mu^* = 0$ and $\sigma^* = 1 + \frac{h^2}{2}(y_x^2 - 1)/(y_x^2 - E[Y_1^2])$ so that the asymptotic bias is reduced to $\frac{h^2}{2}\left\{-3f_X'(x)y_x\phi(y_x) + f_X''(x)\phi^2(y_x)\right\}$. This strategy, motivated by pointwise bias reduction, does not correspond to any optimality

consideration. Moreover, it is not defined on the entire support of y (e.g., when $y_x^2 = E[Y_1^2]$, σ^* is not defined). My numerical experiments show that this approach does not provide satisfactory performance compared with competing methods. It is also noted that $\sigma^* \approx 1 + h^2/2$ for large y_x^2. This alternative simplification has been investigated in my simulations and was found to be dominated by other methods as well. I note that a similar observation regarding this 'infeasible' asymptotic refinement is made by Geenens (2014), who proceeded to advocate a local log-polynomial estimator.

Let us now consider the asymptotic variance. The asymptotic variance of the TKDE, given by Wand, Marron, and Ruppert (1991), is

$$\text{avar}\{\hat{f}_{X,T}(x)\} = \frac{f_Y(y_x)}{2\sqrt{\pi}nh\phi^2(y_x)} = \frac{f_X(x)}{2\sqrt{\pi}nh\phi(y_x)}.$$

It follows that

$$\begin{aligned}\text{avar}\{\hat{f}_{X,M}(x)\} =& \text{avar}\{J(x;\mu,\sigma,h)\hat{f}_{X,T}(x)\} \\ \approx& \{1 + (E[Y_1] - y_x)\mu + (EY_1^2 - y_x^2)(\sigma - 1)\}^2 \text{avar}\{\hat{f}_{X,T}(x)\} \\ =& \text{avar}\{\hat{f}_{X,T}(x)\}, \text{ for } x \in (0,1),\end{aligned} \tag{2.9}$$

since $\mu \to 0$ and $\sigma \to 1$ as $n \to \infty$.

Next I consider the boundary scenarios. Suppose that $x/h^m \to \delta$ or $(1-x)/h^m \to \delta$ for some $\delta, m > 0$. Using that $\Phi^{-1}(x) \sim -\sqrt{-2\log x}$ for $x \to 0$ and $\Phi^{-1}(1-x) \sim -\sqrt{-2\log(1-x)}$ for $x \to 1$, I can show that

$$\text{abias}\{\hat{f}_{X,M}(x)\} \sim Cmh^2 \log h^{-1} f_X(x), \quad \text{avar}\{\hat{f}_{X,M}(x)\} \sim \frac{f_X(x)}{\sqrt{2}nh^{1+m}\delta},$$

where C is a finite positive constant.

Like the TKDE, the MTK suffers an increase in the order of bias and variance

over a small region close to the boundaries. The inflation of boundary variance is a common phenomenon. Both the beta kernel estimator by Chen (1999) and the Gaussian-copula estimator by Jones and Henderson (2007) share the same property and the boundary variance of the latter is identical to that of the MTK. Nevertheless, as indicated by Chen (1999) and Jones and Henderson (2007), the influence of slightly-increased orders in the boundary region upon the global property is negligible. In practice it does not compromise the good performance of the MTK, which is demonstrated by my simulations below.

2.3.2 Mean integrated square error

I shall now explore the global properties of the MTK, focusing on the weighted mean integrated squared error (WMISE). Let $w(x), x \in [0,1]$, be some non-negative weight function. The WMISE is defined as

$$\text{WMISE}(\hat{f}_{X,M}) = E\left[\int_0^1 \{\hat{f}_{X,M}(x) - f_X(x)\}^2 w(x)dx\right]. \tag{2.10}$$

Following Jones and Henderson (2007), I set $w(x) = \phi(\Phi^{-1}(x))$ to insure the integrability of the MISE. In fact, the WMISE in the X-domain with this particular weight function is equivalent to the (unweighted) MISE in the Y-domain. Wand, Marron, and Ruppert (1991) noted that good performance in the Y-domain often translates into that in the X-domain, which is a main motivation of taking the transformation approach in the first place. This observation is confirmed by many numerical investigations, including my own simulations reported below.

To ease exposition, I define $\mathbf{\Theta} = [\mu, \sigma - 1]^T$. Plugging the asymptotic bias (2.7) and variance (2.9) of the MTK into the WMISE yields

$$\text{WMISE}(\hat{f}_{X,M}) \approx \int_0^1 [\{\text{abias}(\hat{f}_{X,M}(x))\}^2 + \text{avar}(\hat{f}_{X,M}(x))]w(x)dx$$

(2.11)

$$=\boldsymbol{\Theta}^T \boldsymbol{A_1} \boldsymbol{\Theta} + h^2 \boldsymbol{\Theta}^T \boldsymbol{A_2} + \frac{h^4 A_3}{4} + \frac{1}{2\sqrt{\pi} n h},$$

where

$$\begin{aligned} \boldsymbol{A}_1 &= \int_{-\infty}^{\infty} [EY_i - y, EY_i^2 - y^2]^T [EY_i - y, EY_i^2 - y^2] f_Y^2(y) dy, \\ \boldsymbol{A}_2 &= \int_{-\infty}^{\infty} [EY_i - y, EY_i^2 - y^2]^T f_Y(y) f_Y''(y) dy, \\ A_3 &= \int_{-\infty}^{\infty} \left\{ f_Y''(y) \right\}^2 dy. \end{aligned} \tag{2.12}$$

The optimal smoothing parameters, which minimize the asymptotic WMISE (2.11), are then given by

(2.13)

$$h_{0,M} = (2\sqrt{\pi})^{-1/5} \left(A_3 - \boldsymbol{A}_2^T \boldsymbol{A}_1^{-1} \boldsymbol{A}_2 \right)^{-1/5} n^{-1/5},$$

(2.14)

$$\boldsymbol{\Theta}_{0,M} = -\frac{h^2}{2} \boldsymbol{A}_1^{-1} \boldsymbol{A}_2.$$

It follows that the optimal asymptotic WMISE of the MTK is

$$\text{WMISE}_0(\hat{f}_{X,M}) = \frac{5}{4} (2\sqrt{\pi})^{-4/5} \left(A_3 - \boldsymbol{A}_2^T \boldsymbol{A}_1^{-1} \boldsymbol{A}_2 \right)^{1/5} n^{-4/5}. \tag{2.15}$$

I show in Appendix A.1 that $A_3 - \boldsymbol{A}_2^T \boldsymbol{A}_1^{-1} \boldsymbol{A}_2 \geq 0$ with equality when f_X is the uniform distribution.

Note that the optimal parameters for the TKDE, as a special case of MTK with $\boldsymbol{\Theta} = \boldsymbol{0}$, are readily obtained as

(2.16)

$$h_{0,T} = (2\sqrt{\pi})^{-1/5} A_3^{-1/5} n^{-1/5},$$

(2.17)

$$\text{WMISE}_0(\hat{f}_{X,T}) = \frac{5}{4}(2\sqrt{\pi})^{-4/5} A_3^{1/5} n^{-4/5}.$$

Since $\boldsymbol{A_1}$ is positive semi-definite, I have $A_3 \geq A_3 - \boldsymbol{A}_2^T \boldsymbol{A}_1^{-1} \boldsymbol{A}_2$. Thus the MTK has the usual convergence rate of kernel density estimators and dominates the TKDE in terms of the WMISE (2.10). Since the two estimators share the same asymptotic variance, it is understood that the reduction in the WMISE comes from the asymptotic bias reduction discussed above.

2.4 Selection of smoothing parameters

It is well known that the performance of kernel density estimations depends crucially on the bandwidths (see, e.g., Jones, Marron, and Sheather (1996a,b) for comprehensive reviews of the selection of bandwidth in kernel density estimations). The MTK requires the selection of smoothing parameters (h, μ, σ). In this section, I present three automatic methods of smoothing parameter selection.

2.4.1 Plug-in method

I have derived the optimal smoothing parameters (2.13) and (2.14) that minimize the asymptotic WMISE. One natural course to proceed is to use their sample analogs in the estimation. This requires estimating the quantities $\boldsymbol{A_1}, \boldsymbol{A_2}$ and A_3 specified in (2.12).

Define $G_{s,t}(Y) = (E[Y^s] - Y^s)(E[Y^t] - Y^t)$ with $G_{s,0}(Y) = (E[Y^s] - Y^s)$, and

(2.18) $$A_{s,t}^{(r)} = E[G_{s,t}(Y) f_Y^{(r)}(Y)],$$

where $f_Y^{(r)}(y) = \partial^r f_Y(y)/\partial y^r$. For simplicity, I denote $A_{s,t}^{(0)}$ by $A_{s,t}$. Assuming that

f_Y is sufficiently smooth, I can establish via integration by parts that

$$(2.19)\quad \boldsymbol{A_1} = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix},\ \boldsymbol{A_2} = \begin{bmatrix} A_{1,0}^{(2)} & A_{2,0}^{(2)} \end{bmatrix}^T,\ A_3 = A_{0,0}^{(4)}.$$

I opt to estimate these quantities nonparametrically. In particular, they can be estimated as follows:

$$(2.20)\quad \hat{A}_{s,t}^{(r)} = \begin{cases} \frac{1}{n^2 b^{r+1}} \sum_{i=1}^n \sum_{j=1}^n G_{s,t}(Y_i) K^{(r)}\left(\frac{Y_i - Y_j}{b}\right), & \text{if } (s,t) = (0,0); \\ \frac{1}{n(n-1) b^{r+1}} \sum_{i=1}^n \sum_{j \neq i} G_{s,t}(Y_i) K^{(r)}\left(\frac{Y_i - Y_j}{b}\right), & \text{otherwise,} \end{cases}$$

where $K^{(r)}(x) = \partial^r K(x)/\partial x^r$ and b is the bandwidth. My numerical investigations suggest that $A_3 = A_{0,0}^{(4)}$ is the most difficult to estimate among all $A_{s,t}^{(r)}$'s. I therefore choose, according to the rule of thumb, a bandwidth that is optimal for the estimation of A_3, which is given by

$$b = \left(\frac{16\sqrt{2}}{5}\right)^{1/7} \times s \times n^{-1/7},$$

where $s = \sqrt{1/n \sum_{i=1}^N (Y_i - \bar{Y})^2}$ with $\bar{Y} = 1/n \sum_{i=1}^n Y_i$. Technical details on the derivation of the plug-in bandwidth calculation are given in Appendix A.2.

2.4.2 Cross validation

Cross validation is a commonly used method to select smoothing parameters of kernel estimators. I consider the least square cross validation, whose objective function is given by

$$(2.21)\quad CV(h, \mu, \sigma) = \int_0^1 \left(\hat{f}_{X,M}(x)\right)^2 w(x) dx - \frac{2}{n} \sum_{i=1}^n \hat{f}_{X,M}^{-i}(X_i) w(X_i),$$

where $w(x)$ is a weight function and $\hat{f}_{X,M}^{-i}(X_i)$ is the "leave-one-out" version of my proposed estimator $\hat{f}_{X,M}$ evaluated at data point X_i (see, e.g., Wand and Jones (1995)). As discussed above, setting $w(x) = 1$ leads to the usual (unweighted)

cross validation on the X-domain density, while the weighted cross validation, with $w(x) = \phi(\Phi^{-1}(x))$, amounts to conducting the cross validation on the Y-domain density.

My experiments indicate that in most cases, the weighted and unweighted cross validation methods tend to yield comparable results, confirming Wand et al's (1991) observation that there exists little practical difference in conducting bandwidth selection in the X or Y domain. One notable exception is that when the densities have poles at the boundaries, the weighted cross validation performs considerably better, demonstrating the merit of transformation and the benefit of conducting bandwidth selection in the Y-domain.

I henceforth focus on the weighted cross validation with $w(x) = \phi(\Phi^{-1}(x))$. The second term of (2.21) can be evaluated straightforwardly. Taking the kernel function K to be the standard Gaussian density, the first term, after some tedious algebra, is shown to admit the following analytical form:

$$\begin{aligned}
&\int_0^1 \left(\hat{f}_{X,M}(x)\right)^2 \phi(\Phi^{-1}(x))dx \\
=&\frac{\sqrt{h^2\sigma^2+\sigma^2-h^2}}{2\sqrt{\pi}h\sigma} \Bigg/ \left(\sum_{i=1}^n \exp\left\{-\frac{(\sigma^2-1)Y_i^2+2\mu Y_i-\mu^2(h^2+1)}{2(h^2\sigma^2+\sigma^2-h^2)}\right\}\right)^2 \\
&\times \sum_{i=1}^n\sum_{j=1}^n \exp\left\{-\frac{\sigma^2(Y_i^2+Y_j^2)-2h^2\mu^2}{2h^2\sigma^2}+\frac{[\sigma^2(Y_i+Y_j)-2\mu h^2]^2}{4h^2\sigma^2(\sigma^2+h^2\sigma^2-h^2)}\right\}.
\end{aligned}$$

The weighted cross validation is then undertaken by minimizing the objective function (2.21) with respect to h, μ and σ under the restrictions that $h > 0, \sigma > 0$ and $\sigma^2 + h^2\sigma^2 - h^2 > 0$. Note that the last condition is satisfied trivially if $\sigma > 1$, which holds in all my numerical experiments.

Compared with the plug-in method, the cross validation is completely data driven and does not require estimating unknown quantities. On the other hand, it is com-

putationally more expensive due to the numerical optimization.

2.4.3 Profile cross validation

I next present a hybrid method of smoothing parameter selection that combines the benefits of the plug-in and cross validation methods and at the same time is computationally less demanding.

Recall that among the quantities in (2.12), A_3 is the most difficult to estimate. Also note that A_3 is only present in the optimal bandwidth $h_{0,M}$ given by (2.13), but not in the optimal tuning parameters $\boldsymbol{\Theta}_{0,M}$ given by (2.14). This observation motivates a *profile* cross validation approach, in which I first solve for $\boldsymbol{\Theta_{0,M}}(h)$ as a function of h and then conduct the cross validation with respect to h alone. Below I provide a step-by-step description of this approach:

1. Estimate $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$, which do not involve A_3, according to (2.19) and (2.20);
2. For a given h, calculate $\mu(h)$ and $\sigma(h)$ according to (2.14);
3. Plug $\mu(h)$ and $\sigma(h)$ into the cross validation objective function (2.21); conduct cross validation with respect to h.

The advantage of this profile approach is that it avoids the difficulty of estimating A_3 and lowers the dimension of optimization from three in the full cross validation to one, reducing the computational burden considerably. Regarding the preliminary bandwidth required in the estimation of $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$, I found that the usual rule of thumb provides essentially identical results to those obtained under more complicated selection rules. Hence I used the rule of thumb bandwidth to estimate $\boldsymbol{A_1}$ and $\boldsymbol{A_2}$ in the first step.

2.5 Simulations

I investigate the finite sample performance of the MTK using Monte Carlo simulations and compare it with the TKDE. I also consider the second beta kernel estimator of Chen (1999), which is shown to perform better among the two he proposed, and the Gaussian copula based estimator of Jones and Henderson (2007). I focus my comparison on these two competing estimators since they are in spirit similar to the transformation-based estimators in the sense that they all involve locally varying kernel functions. For the MTK, I experiment with all three methods of smoothing parameter selection discussed above. The optimal plug-in bandwidth (2.16) is used for the TKDE. I use the rule of thumb bandwidths, derived by Jones and Henderson (2007), for the beta kernel estimator and the Gaussian-copula kernel estimator.

I consider eight distributions with bounded support $[0, 1]$. This set of distributions, illustrated in Figure 2.1, are designed to capture a wide range of shapes and features such as asymmetry, skewness, multi-modality, sharp peak and poles at the boundaries. One may refer to Appendix A.3 for their constructions. I categorize these distributions into two groups to facilitate the interpretation of simulation results. The first group includes four distributions whose densities and derivatives of densities are bounded such that there are no poles throughout their supports. Density 1 is a bell-shaped symmetric density that vanishes toward the boundaries. Density 2 is skewed and has more probability mass near the boundaries compared to density 1. Density 3 is symmetric and bi-modal. Density 4 has a sharp peak in the middle. The second group contains densities with poles at the boundaries. Density 5 is symmetric with both tails tending to infinity. Density 6 is asymmetric with an unbounded left tail. Density 7 is bounded with unbounded derivatives at both tails. It is also asymmetric and bi-modal. Density 8 is bounded, asymmetric and has an

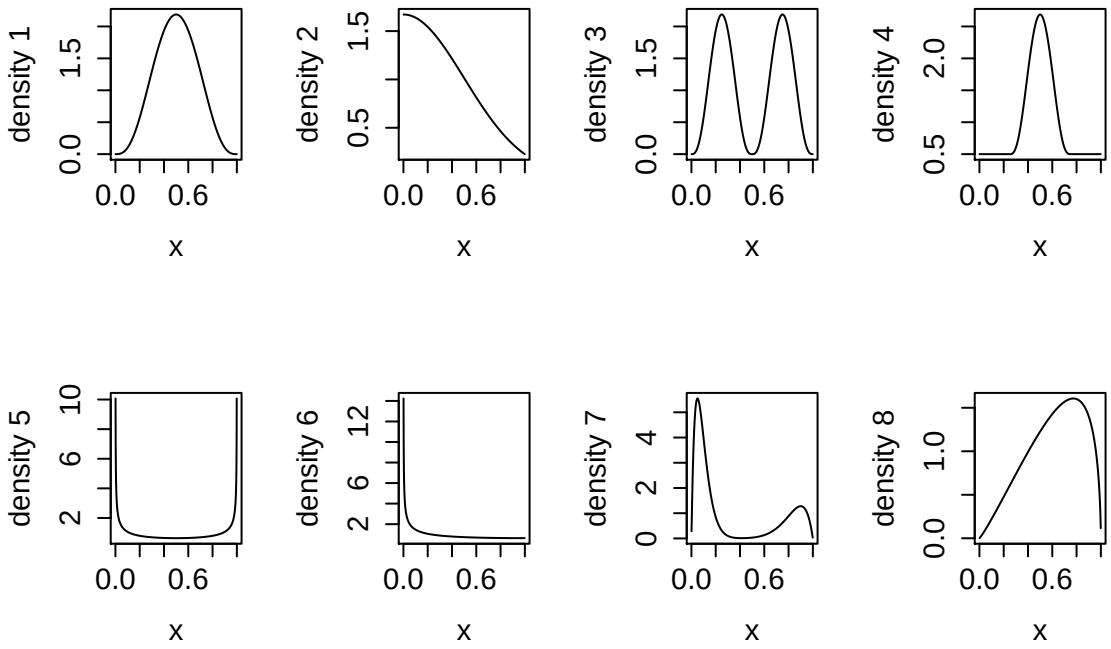


Figure 2.1: Density plot of distributions used in the simulations

unbounded derivative at the right boundary.

For each distribution, I conduct simulations with sample sizes $n = 100$ and $n = 500$. Each experiment is repeated 1,000 times. I evaluate the mean integrated square errors of the estimators on an equally spaced grid on $[0.001, 0.999]$ with an increment of 0.001. The MISE of each estimator, averaged across 1,000 repetitions, are reported in Table 2.1. The corresponding standard deviations are reported in parenthesis. For each experiment, the estimator with the minimum average MISE is highlighted in bold font.

Consistent with my theoretical analysis, the MTK generally dominates the TKDE in my experiments. Among all estimators, the MTK provides the best performance for six out of eight densities when $n = 100$, and in seven densities when $n = 500$. The

Table 2.1: Simulation results: Average MISE (all numbers are multiplied by 1,000 to improve readability)

	Densities							
	1	2	3	4	5	6	7	8
	Panel 1: n=100							
$\hat{f}_{X,T}$	37.22	58.08	107.33	105.48	106.60	91.06	207.75	40.53
	(25.29)	(45.83)	(43.1)	(52.73)	(84.44)	(82.66)	(117.36)	(29.62)
$\hat{f}_{X,C2}$	**29.80**	**26.86**	533.95	126.56	317.03	324.50	260.74	35.06
	(20.08)	(18.83)	(53.75)	(56.9)	(49.42)	(61.9)	(63.16)	(18.93)
$\hat{f}_{X,GC}$	32.19	27.14	394.08	246.38	110.16	93.03	552.04	**24.77**
	(23.41)	(18.56)	(40.52)	(81.07)	(70.15)	(78.2)	(107.45)	(17.39)
$\hat{f}_{X,M}$, plug-in	30.67	40.50	106.92	**93.42**	**75.07**	**75.22**	206.38	29.96
	(24.85)	(35.41)	(45.11)	(40.36)	(69.12)	(74.98)	(109.81)	(23.23)
$\hat{f}_{X,M}$, CV	38.67	36.33	**93.65**	101.62	81.81	98.03	**147.76**	31.69
	(49.46)	(43.33)	(76.74)	(90.74)	(87.23)	(121.13)	(106.48)	(41.66)
$\hat{f}_{X,M}$, profile CV	38.43	35.96	96.95	101.49	97.94	109.46	164.67	33.96
	(41.48)	(43.38)	(80.27)	(87.33)	(103.46)	(106.13)	(115.9)	(41.32)
	Panel 2: n=500							
$\hat{f}_{X,T}$	11.07	16.21	26.55	30.12	29.49	27.00	52.06	11.63
	(6.65)	(9.77)	(10.87)	(12.09)	(22.54)	(23.12)	(27.53)	(6.65)
$\hat{f}_{X,C2}$	8.82	14.26	315.87	53.83	213.39	223.30	125.34	13.24
	(5.52)	(6.28)	(40.00)	(18.47)	(21.92)	(25.72)	(22.4)	(5.53)
$\hat{f}_{X,GC}$	10.22	**8.92**	229.06	137.77	33.43	29.08	253.58	7.98
	(6.72)	(5.38)	(20.39)	(34.99)	(21.84)	(22.86)	(37.89)	(5.41)
$\hat{f}_{X,M}$, plug-in	8.16	10.16	26.13	**28.09**	18.80	**17.26**	50.99	7.67
	(5.76)	(6.62)	(10.88)	(11.92)	(16.73)	(15.45)	(24.65)	(5.05)
$\hat{f}_{X,M}$, CV	**7.79**	9.73	**24.25**	28.47	**17.71**	24.78	42.26	**5.51**
	(9.89)	(7.03)	(15.15)	(15.41)	(17.15)	(20.72)	(27.1)	(5.82)
$\hat{f}_{X,M}$, profile CV	9.69	9.37	24.45	28.57	20.34	24.63	**41.41**	7.42
	(9.9)	(9.39)	(13.07)	(17.05)	(27.29)	(20.88)	(24.41)	(8.04)

three methods of smoothing parameter selection provide comparable performance in many cases. When $n = 100$, the plug-in method stands out, while when $n = 500$, the cross validation methods provide better overall results.

Comparison across individual experiments provides some useful insights into the relative strength of the estimators. When $n = 100$, the beta estimator outperforms others for the first two densities, which are "well behaved" in the sense that they are smooth, uni-modal and have no poles in the density or its derivative. (This is consistent with an observation by Jones and Henderson (2007) that the beta kernel estimator is most suitable for smooth bounded densities.) The MTK estimators outperform the others substantially in densities 3 and 4, which are marked by abrupt features. For the densities in group two with unbounded tails or derivatives, the MTK estimators clearly outperform their competitors in densities 5-7. Not surprisingly, the Gaussian copula estimator provides the best performance in density 8, which is a conditional Gaussian copula density. Nonetheless, the Gaussian copula estimator outperforms the MTK estimators only slightly. These results confirm my analyses that the MTK is particularly suitable for densities with poles at the boundaries.

The overall pattern remains similar in experiments with $n = 500$, wherein the MTK estimators score the best performance in seven densities. The only exception is density 2, where the Gaussian copula estimator slightly outperforms the MTK estimators. Remarkably, the MTK estimators perform best in density 8, for which the Gaussian copula density is expected to be the most competitive estimator. For illustration, I plot in Appendix A.4 a random example of estimated densities with $n = 500$. Since the MTK estimates resulted from different methods of bandwidth selection are rather similar, I report only those obtained from the profile bandwidth selection.

In summary the MTK dominates the TKDE and provides comparable and often

superior performance compared with two competing estimators. The three proposed methods of smoothing parameter selection appear to perform satisfactorily. Since the plug-in method has the smallest variation and is computationally inexpensive, it is recommended for general purpose applications, especially when the same size is small.

2.6 Empirical example

In this section I apply the MTK and other estimators to a real data set. The data contain the ratio of white student enrollment of 56 public schools in Nassau County, New York for the 1992-1993 school year. Estimating the density of white student enrollment ratios has been of interest to assess the common perception in the US in the 90's that public schools were still strongly segregated by race, despite political effort to integrate them. Since ratios are restricted on the $[0, 1]$ interval, this data set was used as an example to illustrate boundary bias problem in Simonoff (1996) and investigated by, among others, Geenens (2014) to compare kernel estimators of densities with bounded supports.

The same set of kernel estimators and methods of smoothing parameter selection as used in the simulations are applied to the school ratio data. The estimated densities are plotted in Figures 2.2 and 2.3 below, superimposed on the histogram of the data with individual observations marked on the horizontal axis. Tabulation of the data shows that 50% of the data fall in the interval of $[0.85, 0.96]$, which is captured by the apparent peak of the histogram around 0.9. Only 7% of the data are larger than 0.96 with a maximum at 0.976, causing a sharp decline of the density toward the top end of the distribution. There are two schools with exceptionally low white student ratios (less than 1%), suggesting a minor peak near the low end of the distribution. The estimated densities by the MTK with three different bandwidth

selection methods, which are plotted in Figure 2.2, all capture the sharp peak and abrupt decline of the density near the right boundary. The two estimates obtained under the plugin and the profile-CV methods show a minor peak at the low end, while that under the CV method fails to do so. These results are consistent with my findings in the simulations that the plugin and profile CV methods tend to perform better than the full CV method when the sample size is small.

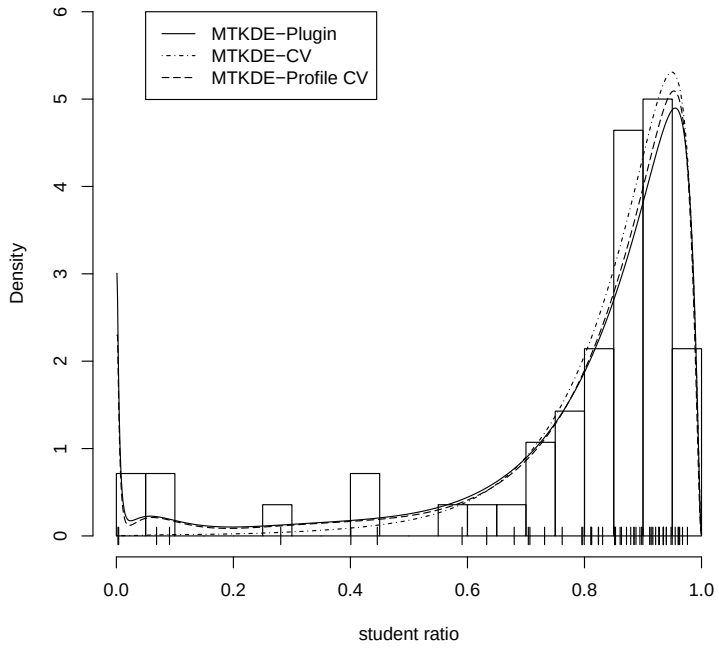


Figure 2.2: MTK estimates

Figure 2.3 reports the results from the TKDE and the beta and Gaussian copula estimators. The overall shape of the TKDE estimate is rather similar to those from the MTK, except for the explosion at the low end. As discussed above, this erratic tail behavior is caused by the multiplication factor of the TKDE, which tends to

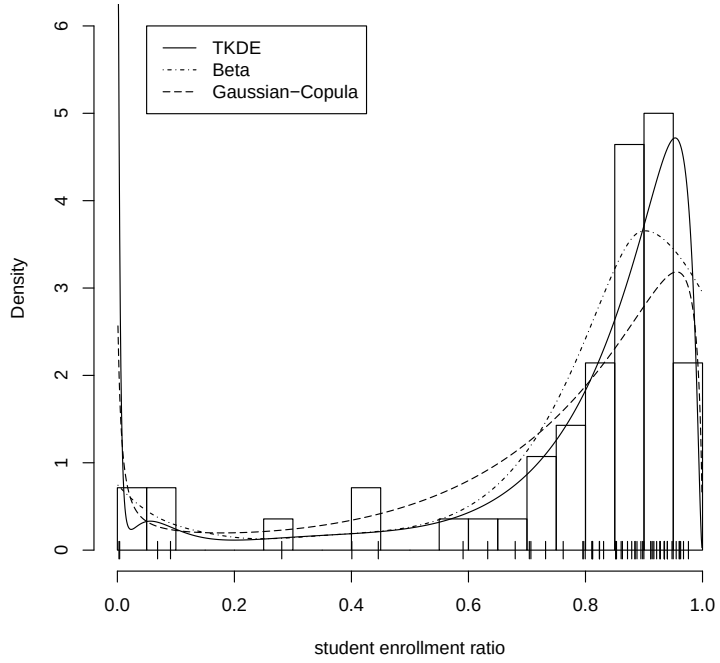


Figure 2.3: Other estimates

infinity towards the boundaries. The beta estimator appears to oversmooth the data, underestimating the mode around 0.9 and the minor peak near the low end while overestimating the density near the right boundary. The Gaussian copula based estimator captures both tails of the densities well, but appears to underestimate the mode of the density.

3. CONSISTENT TRANSFORMATION KERNEL ESTIMATION OF COPULA DENSITIES

3.1 Introduction

Copula has been extensively used in statistics and econometrics nowadays. Given a bivariate random vector $(X, Y)^\top$, denote its associated joint cumulative distribution (cdf) by F and the corresponding marginal distributions by F_X and F_Y. According to *Sklar's theorem* (Sklar 1959), we have

$$F(x, y) = C\left(F_X(x), F_Y(y)\right),$$

where C is termed copula function. If F_X and F_Y are continuous, then C is unique. Copula eases multivariate modeling by two independent and convenient steps: first modeling each marginal distribution and second using a copula to couple them for the desired joint distribution. It is also known that copula characterizes the full dependence structure between the components of $(X, Y)^\top$. As a result of these advantages, copula has found widespread applications in modern quantitative finance and insurance. See Joe (1997) and Nelsen (2006) for comprehensive introductions of copula method.

A bivariate copula function C is defined on the support $\mathcal{I} = [0, 1]^2$ and is a cumulative distribution function of the random vector $(U, V)^\top$ where $U = F_X(X)$ and $V = F_Y(Y)$. This definition comes naturally from the above Sklar's theorem in conjunction with the *probability integral transformation* which implies both U and V are uniformly distributed, i.e. $\mathcal{U}_{[0,1]}$. If C is absolutely continuous, then its associated density function (pdf) exists and it admits the form

$$c(u, v) = \frac{\partial^2 C}{\partial u \partial v}(u, v),$$

where c is called copula density. Frequently, researchers choose to model c directly due to its desirable merits. For example, Geenens, Charpentier, and Paindaveine (2014) highlight that copula density is "more readily interpretable" in many aspects; moreover, in copula goodness-of-fit problems, Fermanian (2005, 2012) find it is easier to focus on copula density to design distribution-free tests. Based on an observed bivariate random sample $\left\{(X_i, Y_i)^\top, i = 1, \cdots, n\right\}$, estimating copula density c essentially amounts to fitting a bivariate distribution. In practice, the estimation procedure depends on how strong assumptions we are willing to make. The main body of the literature adopts parametric or semiparametric methods, where one assumes a parametric family to the copula density and choose to treat the marginal distributions either parametrically or nonparametrically, see for example, Nelsen (2006), Genest, Ghoudi, and Rivest (1995) and Chen, Fan, and Tsyrennikov (2006). However, they may lack flexibility and induce misspecification error. Therefore, the flexible nonparametric method, which is free of any distributional assumption, is needed as a complement. For these reasons, I attempt to address nonparametric estimation of copula density c in this article. To date, various nonparametric techniques have been adapted to this area, to name a few, splines methods (Shen, Zhu, and Song 2008; Kauermann, Schellhase, and Ruppert 2013), wavelets (Hall and Neumeyer 2006; Genest, Masiello, and Tribouley 2009; Autin, Pennec, and Tribouley 2010), Bernstein polynomials (Bouezmarni, Rombouts, and Taamouti 2010; Bouezmarni, El Ghouch, and Taamouti 2013; Janssen, Swanepoel, and Veraverbeke 2014), and maximum penalized likelihood (Qu and Yin 2012).

Kernel type copula density estimators seem to be less developed in the literature. Recall that c is the density function of the $\big(U = F_X(X), V = F_Y(Y)\big)^\top$. Typically, a genuine sample $\left\{(U_i, V_i)^\top, i = 1, \cdots, n\right\}$ is unavailable because F_X and F_Y are unknown. In the copula literature, it is a common practice to use the "pseudo-

sample" instead, namely

$$(3.1)\quad \hat{U}_i = \frac{n}{n+1}\hat{F}_{n,X}(X_i) \quad \text{and} \quad \hat{V}_i = \frac{n}{n+1}\hat{F}_{n,Y}(Y_i),$$

where $\hat{F}_{n,X}(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \le x\}}$ is the empirical distribution of F_X, and similarly is $\hat{F}_{n,Y}$ defined of F_Y. The rescaling factor $\frac{n}{n+1}$ is placed here to guarantee all the data points lie in the interior of $\mathcal{I}$. In other words, the pseudo-sample is formed by the ranks of $\left\{(X_i, Y_i)^\top, i = 1, \cdots, n\right\}$ divided by $n+1$, thus always takes values in $\left\{\frac{1}{n+1}, \frac{2}{n+1}, \cdots, \frac{n}{n+1}\right\}$ for each margin. The pseudo-sample $\left\{(\hat{U}_i, \hat{V}_i), i = 1, \cdots, n\right\}$ is then treated as if actually observed for estimation. From the pseudo-sample, the standard kernel estimator is given by

$$(3.2)\quad \hat{c}(u,v) = \frac{1}{n|\boldsymbol{H}|^{1/2}} \sum_{i=1}^{n} \boldsymbol{K}\left(\boldsymbol{H}^{-1/2}\begin{pmatrix} u - \hat{U}_i \\ v - \hat{V}_i \end{pmatrix}\right),$$

where $\boldsymbol{H}$ is a symmetric positive-definite bandwidth matrix and $\boldsymbol{K}$ is bivariate kernel function. See Wand and Jones (1995) for details. However, at least two reasons fail the standard kernel estimator for copula density estimation. First, copula density admits the bounded support $\mathcal{I}$. It is well known that the standard kernel estimator is inconsistent on the boundaries of $\mathcal{I}$, especially in the four corners; see Charpentier, Fermanian, and Scaillet (2006) for explicit asymptotic bias formulas. Second, many parametric copula densities are unbounded; for example, the commonly used Gaussian copula density is unbounded in two corners if certain correlation is present. This unboundedness feature violates the assumptions of standard kernel estimator and thus leads to inconsistent estimates.

To correct boundary biases, an early mirror reflection estimator is adopted in Gijbels and Mielniczuk (1990). It tries to obtain some artificial data points by mirror reflecting the sample with respect to all borders and corners. However, this method

only works well when the underlying copula density has null first derivatives on boundaries. Transformation-based kernel estimator (TKE) arises as a natural solution for boundary biases correction and it is considered in Charpentier, Fermanian, and Scaillet (2006) and is further improved in Geenens, Charpentier, and Paindaveine (2014). This transformation idea dates back to Wand, Marron, and Ruppert (1991) and Marron and Ruppert (1994) in density estimation. Specifically, it is a two-step procedure. First, it employs the Probit transformation that maps the bounded support $\mathcal{I}$ to the unbounded support $\mathcal{R}^2$, where the transformed density is finite and the standard kernel estimator can be applied free of boundary biases. Second, the desired copula density estimator is obtained after the back transformation. The TKE is shown to be asymptotically consistent, however in most cases, it suffers erratic tail behaviors caused by the unbounded multiplier on the boundaries associated with the back transformation. This is a serious concern in practice since copula analyses are mostly used in risk management, in which reliable tail estimates are of crucial importance.

In this article, I propose a modified transformation-based kernel estimator (MTK) to overcome this drawback of the TKE. My solution employs a smooth infinitesimal tapering device to mitigate the unpleasant influences of the unbounded multiplier. Moreover, it incorporates an interaction parameter to further allow directional tapering since copula densities are often stretched along the diagonals of $\mathcal{I}$. I derive the asymptotic properties of the MTK and demonstrate that it dominates the TKE in terms of asymptotic mean integrated square error. Based on my theoretical analyses, I propose two practical methods of selecting optimal smoothing parameters, which are computationally simple. I further conduct Monte Carlo simulations to demonstrate its superior finite sample performance. The MTK produces a *bona fide* density. One appealing property is that it retains the simplicity of the TKE with the

fixed Probit transformation and a single global bandwidth, in contrast to data driven transformation or locally varying bandwidths. The other particularly appealing property is that for Gaussian copulas, the MTK obtains a higher order convergence rate. Consequently, it yields outstanding performance when the underlying copulas are Gaussian or near Gaussian, which are practically plausible scenarios in the analyses of financial data. Therefore, my method provides a simple copula density estimator for the practitioners who seek both flexibility and excellent performance.

The rest text proceed as follows. In Section 3.2, I briefly describe the TKE and then formally introduce the MTK under diagonal bandwidth matrix. In Section 3.3, I derive the asymptotic properties of the MTK, followed by two practical methods to select the optimal smoothing parameters in Section 3.4. I show the properties of the MTK under Gaussian copulas in Section 3.5. Extensions of the MTK to non-diagonal bandwidth matrix are considered in Section 3.6. I report simulation results in Section 3.7 and apply the MTK to three real world datasets in Section 3.8. Some proofs are gathered in Appendix B.

3.2 Estimator

3.2.1 Preliminaries

Copula density c is the pdf of random vector $(U, V)^\top$, which admits the bounded support $\mathcal{I} = [0, 1]^2$, thus the standard kernel estimator suffers boundary biases problem. The transformation-based kernel estimator (TKE) is hereby proposed and its properties are discussed in Charpentier, Fermanian, and Scaillet (2006) and Geenens, Charpentier, and Paindaveine (2014).

Consider the Probit transformation, i.e.

$$S = \Phi^{-1}(U) \quad \text{and} \quad T = \Phi^{-1}(V),$$

where Φ is the cdf of standard Gaussian distribution and Φ^{-1} is the corresponding quantile function. Let g be the joint pdf of $(S,T)^\top$. The simple change of variable yields

$$(3.3)\quad g(s,t) = c(\Phi(s),\Phi(t))\phi(s)\phi(t), \quad \forall (s,t) \in \mathcal{R}^2$$

where ϕ is the pdf of standard Gaussian distribution. Since U and V are $\mathcal{U}_{[0,1]}$, both S and T follow the standard Gaussian distribution. Though this does not imply $(S,T)^\top$ is distributed to bivariate Gaussian distribution (only when C is Gaussian copula), we expect g well-behaved and easy for estimation. Based on the transformed genuine sample $\left\{\left(S_i = \Phi^{-1}(U_i), T_i = \Phi^{-1}(V_i)\right), i = 1, \cdots, n\right\}$, the density g in ST-domain can be estimated by standard kernel estimator, denoted by $\hat{g}$, free of boundary bias for two reasons. First, the support of the transformed random vector $(S,T)^\top$ becomes the unconstrained $\mathcal{R}^2$. Second, even though c may be unbounded, under Assumptions 1-3 provided in Appendix B.1, g together with its partial derivatives up to the second order are uniformly bounded on $\mathcal{R}^2$ (Geenens, Charpentier, and Paindaveine 2014, Lemma A.1). Then according to (3.3), the TKE is readily restored after the back transformation and I have

$$(3.4)\quad \hat{c}_t(u,v) = \frac{\hat{g}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)}{\phi\left(\Phi^{-1}(u)\right)\phi\left(\Phi^{-1}(v)\right)}, \quad \forall (u,v) \in (0,1)^2.$$

(Geenens, Charpentier, and Paindaveine 2014, Section 2.1) note that many desirable properties of $\hat{g}$, such as uniformly weak or strong asymptotic consistency, are retained to $\hat{c}_t$ after the back transformation. If $\hat{g}$ performs reasonably well, then $\hat{c}_t$ shall produce acceptable estimates.

Since the genuine sample $\left\{(U_i,V_i)^\top, i = 1, \cdots, n\right\}$ is unavailable, one has to use the transformed pseudo-sample $\left\{\left(\hat{S}_i = \Phi^{-1}(\hat{U}_i), \hat{T}_i = \Phi^{-1}(\hat{V}_i)\right), i = 1, \cdots, n\right\}$

instead. Formally, the TKE is defined as

$$(3.5)\quad \hat{c}_t(u,v) = \frac{1}{n|\boldsymbol{H}|^{1/2}\,\phi\left(\Phi^{-1}(u)\right)\phi\left(\Phi^{-1}(v)\right)} \sum_{i=1}^{n} \boldsymbol{K}\left(\boldsymbol{H}^{-1/2}\begin{pmatrix} \Phi^{-1}(u) - \hat{S}_i \\ \Phi^{-1}(v) - \hat{T}_i \end{pmatrix}\right),$$

where $\boldsymbol{H}$ is a symmetric positive-definite bandwidth matrix, $|\boldsymbol{H}|$ is the corresponding determinant, and $\boldsymbol{K}$ is a bivariate kernel function. The influence by working with the pseudo-sample instead of the genuine one is asymptotically negligible. This is intuitively understandable because the empirical distribution function is $\sqrt{n}$-consistent, thus converges faster than the subsequent kernel density estimator. In practice, some consequences should be expected, sometimes even in an advantageous manner. For example, obviously the pseudo-sample is more "uniform"; this usually leads to smaller variance of the estimate at a given point. One may refer to Charpentier, Fermanian, and Scaillet (2006) and Genest and Segers (2010) for detailed demonstrations.

3.2.2 Modified transformation-based kernel estimator

Despite that the TKE is designed to address the boundary biases issue induced by the bounded support, it seems inadequate for satisfactory estimates and needs further improvement. This can be noticed by looking at the multiplier associated with the back transformation; from (3.4), it has the form $\left(\phi\left(\Phi^{-1}(u)\right)\phi\left(\Phi^{-1}(v)\right)\right)^{-1}$. It is obvious that, when $u \to 0$ and/or $v \to 0$, the multiplier grows unboundedly, resulting in possibly erratic behaviors for the TKE. For instance, in practice, even slight biases of $\hat{g}$ at the tails will be magnified greatly by the multiplier; therefore large biases could be introduced and we may observe $\hat{c}_t$ exploding on the boundaries, especially in certain corners of $\mathcal{I}$.

This motives us to propose a modified transformation-based kernel estimator

(MTK) which implements further bias correction upon the TKE and at the same time maintains its simplicity. This new estimator is inspired by the idea that the undesirable consequences induced by the multiplier may be alleviated by the tapering method. Let $\phi_{1+\theta_1}(\cdot)$ be the pdf of the Gaussian distribution with mean 0 and standard deviation $1+\theta_1$. Specifically, I deflate the values of the original multiplier on the boundaries by replacing it with an infinitesimal tampered factor $\left(\phi_{1+\theta_1}\left(\Phi^{-1}(u)\right)\phi_{1+\theta_1}\left(\Phi^{-1}(v)\right)\right)^{-1}$ where $\theta_1 > 0$ and I assume $\theta_1 \to 0$ as the sample size $n \to \infty$. In order for a simplified form for analysis, I resort to its equivalent by noting that

$$\left(\phi_{1+\theta_1}\left(\Phi^{-1}(u)\right)\phi_{1+\theta_1}\left(\Phi^{-1}(v)\right)\right)^{-1} \sim \frac{\exp\left(-\theta_1\left(\{\Phi^{-1}(u)\}^2+\{\Phi^{-1}(v)\}^2\right)\right)}{\phi\left(\Phi^{-1}(u)\right)\phi\left(\Phi^{-1}(v)\right)}.$$

Since the copula density is often stretched along one of the diagonals of $\mathcal{I}$ if some dependence of $(X,Y)^\top$ is present, it is desirable that the degree of tapering adapts to the orientation of the copula density. This motivates us to further introduce an interaction term that allows directional tapering. Therefore, I finally have my tapered multiplier as

$$\frac{\exp\left(-\theta_1\left(\{\Phi^{-1}(u)\}^2+\{\Phi^{-1}(v)\}^2\right)-\theta_2\Phi^{-1}(u)\Phi^{-1}(v)\right)}{\phi\left(\Phi^{-1}(u)\right)\phi\left(\Phi^{-1}(v)\right)}.$$

Similarly, I require $\theta_2 \to 0$ as $n \to \infty$. Intuitively, this tapering effect is negligible in the interior of $\mathcal{I}$ and only affects the estimates near the boundaries. The two tuning parameters $\boldsymbol{\Theta} = (\theta_1, \theta_2)^\top$ smoothly control the amount of tapering thus need to be chosen carefully for good performance.

From this tapered multiplier, my proposed MTK is formally defined as, for any

$(u, v) \in (0, 1)^2$,

$$\hat{c}_m(u, v) = \frac{\exp\left(-\theta_1\left(\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2\right) - \theta_2\Phi^{-1}(u)\Phi^{-1}(v)\right)}{n\eta|\boldsymbol{H}|^{1/2}\,\phi\left(\Phi^{-1}(u)\right)\phi\left(\Phi^{-1}(v)\right)} \sum_{i=1}^{n} \boldsymbol{K}\left(\boldsymbol{H}^{-1/2}\begin{pmatrix} \Phi^{-1}(u) - \hat{S}_i \\ \Phi^{-1}(v) - \hat{T}_i \end{pmatrix}\right),$$

where $\boldsymbol{H}$ is the symmetric positive-definite bandwidth matrix and $\boldsymbol{K}$ is a bivariate kernel function. To ensure $\hat{c}_m$ actually integrates to one, the normalization term η is thus defined as

$$\eta = \int_{\mathcal{I}} \frac{\exp\left(-\theta_1\left(\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2\right) - \theta_2\Phi^{-1}(u)\Phi^{-1}(v)\right)}{n|\boldsymbol{H}|^{1/2}\,\phi\left(\Phi^{-1}(u)\right)\phi\left(\Phi^{-1}(v)\right)} \sum_{i=1}^{n} \boldsymbol{K}\left(\boldsymbol{H}^{-1/2}\begin{pmatrix} \Phi^{-1}(u) - \hat{S}_i \\ \Phi^{-1}(v) - \hat{T}_i \end{pmatrix}\right) \mathrm{d}u\mathrm{d}v.$$

A common choice for $\boldsymbol{K}$ is the Gaussian kernel function, i.e.

$$\boldsymbol{K}(\boldsymbol{x}) = (2\pi)^{-1}\exp\left(-\frac{1}{2}\boldsymbol{x}^\top\boldsymbol{x}\right). \tag{3.6}$$

Gaussian kernel function have many appealing properties; see, for example, Chaudhuri and Marron (1999) in the univariate setting. Specially in my case, it provides an analytical form for the normalization term η, as we will see in the following. Therefore, I stick to Gaussian kernel function in this article. This choice is also in line with Charpentier, Fermanian, and Scaillet (2006) and Geenens, Charpentier, and Paindaveine (2014).

For simplicity and tractability, I first consider $\boldsymbol{H} = h^2\boldsymbol{I}$ for some $h > 0$. This setup eases the subsequent theoretical analysis and at the same time is qualitatively no different from a more general non-diagonal $\boldsymbol{H}$, which will be revisited in Section

6. Then the MTK is simplified as

$$\hat{c}_{m1}(u,v) = \frac{\exp\left(-\theta_1\left(\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2\right) - \theta_2\Phi^{-1}(u)\Phi^{-1}(v)\right)}{n\eta h^2\phi\left(\Phi^{-1}(u)\right)\phi\left(\Phi^{-1}(v)\right)} \sum_{i=1}^{n}\phi\left(\frac{\Phi^{-1}(u)-\hat{S}_i}{h}\right)\phi\left(\frac{\Phi^{-1}(v)-\hat{T}_i}{h}\right), \tag{3.7}$$

and after some tedious algebra, the normalization factor is

$$\eta = \frac{1}{n\delta}\sum_{i=1}^{n}\exp\left\{-\frac{(4h^2\theta_1^2 - h^2\theta_2^2 + 2\theta_1)(\hat{S}_i^2 + \hat{T}_i^2) + 2\theta_2\hat{S}_i\hat{T}_i}{2\delta^2}\right\},$$

where

$$\delta = \sqrt{h^4(4\theta_1^2 - \theta_2^2) + 4h^2\theta_1 + 1}.$$

3.3 Asymptotic properties

The asymptotic properties of the $\hat{c}_{m1}$ are derived. To proceed, some notations are introduced. Let $\hat{c}_{t1}$ be the simplified TKE under $\boldsymbol{H} = h^2\boldsymbol{I}$. Define $\hat{c}^*_{m1}$ and $\hat{c}^*_{t1}$ analogously to $\hat{c}_{m1}$ and $\hat{c}_{t1}$ respectively but use the transformed genuine sample $\left\{\left(S_i = \Phi^{-1}(U_i), T_i = \Phi^{-1}(V_i)\right), i = 1, \cdots, n\right\}$. These two genuine versions, though infeasible, facilitate my theoretical analysis. Given a function $f(x,y)$, denote its partial derivative by $f^{(r_1,r_2)}(x,y) = \partial^{(r_1+r_2)}f(x,y)/\partial x^{r_1}\partial y^{r_2}$ if it exists.

Some known properties about the TKE in the literature are re-stated here since they are closely related to the MTK as I will show shortly. Charpentier, Fermanian, and Scaillet (2006) first provide analyses of the $\hat{c}^*_{t1}$ and then Geenens, Charpentier, and Paindaveine (2014) formally study the $\hat{c}_{t1}$. From their results, if the Gaussian kernel function (3.6) is used and under Assumptions 1-3 and 5, the following pointwise asymptotic normality results hold. First, in the ST-domain,

$$\sqrt{nh^2}\left(\hat{g}(s,t) - g(s,t) - b_g(s,t)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_g^2(s,t)\right), \quad \forall (s,t) \in \mathcal{R}^2, \tag{3.8}$$

where $b_g(s,t) = \frac{h^2\left(g^{(2,0)}+g^{(0,2)}\right)(s,t)}{2}$ and $\sigma_g^2(u,v) = \frac{g(s,t)}{4\pi}$. This immediately implies that, in the UV-domain,

$$\sqrt{nh^2}\left(\hat{c}_{t1}(u,v) - c(u,v) - b_{t1}(u,v)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{t1}^2(u,v)\right), \quad \forall (u,v) \in (0,1)^2, \tag{3.9}$$

where $b_{t1}(u,v) = \frac{h^2\left(g^{(2,0)}+g^{(0,2)}\right)\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)}{2\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}$ and $\sigma_{t1}^2(u,v) = \frac{c(u,v)}{4\pi\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}$. Note that $b_{t1}(u,v)$ and $\frac{\sigma_{t1}^2(u,v)}{nh^2}$ are exactly the asymptotic bias and variance for the $\hat{c}_{t1}^*$, as shown in Charpentier, Fermanian, and Scaillet (2006). These results formally prove the effect of resorting to the pseudo-sample is asymptotically negligible. For detailed proofs, one may refer to (Geenens, Charpentier, and Paindaveine 2014, Proposition 3.1 and Theorem 3.1).

Now define

$$J(u,v;h,\boldsymbol{\Theta}) = \frac{1}{\eta}\exp\left(-\theta_1\left(\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2\right) - \theta_2\Phi^{-1}(u)\Phi^{-1}(v)\right). \tag{3.10}$$

Then the MTK can be rewritten as

$$\hat{c}_{m1}(u,v) = J(u,v;h,\boldsymbol{\Theta})\hat{c}_{t1}(u,v). \tag{3.11}$$

Thus, it is seen that the MTK introduces a multiplicative adjustment to the TKE. The adjustment $J(u,v;h,\boldsymbol{\Theta})$ is controlled by the tuning parameters $\boldsymbol{\Theta}$. In particular, when $\boldsymbol{\Theta} = 0$, $J(u,v;h,\boldsymbol{\Theta}) = 1$ and the MTK reduces to the TKE.

Consider first the $\hat{c}_{m1}^*$. The construction (3.11) is readily adapted to $\hat{c}_{m1}^*(u,v) = J^*(u,v;h,\boldsymbol{\Theta})\hat{c}_{t1}^*(u,v)$ which eases the analysis. Given a fixed point $(u,v) \in (0,1)^2$, a Taylor expansion of $J^*(u,v;h,\boldsymbol{\Theta})$ with respect to $\boldsymbol{\Theta}$ at zero yields

$$J^*(u,v;h,\boldsymbol{\Theta}) = 1 + \boldsymbol{\Theta}^\top \boldsymbol{B}\left(\Phi^{-1}(u),\Phi^{-1}(v)\right) + o\left(\boldsymbol{\Theta}\right), \tag{3.12}$$

where

$$\boldsymbol{B}(s,t) = \begin{pmatrix} 2 - s^2 - t^2 \\ \mathbb{E}S_iT_i - st \end{pmatrix}.$$

Since the asymptotic properties of $\hat{c}^*_{t1}$ are known, then it follows that

$$\begin{aligned} \text{abias}\left\{\hat{c}^*_{m1}(u,v)\right\} &= b_{m1}(u,v) \\ &\equiv \frac{h^2\left(g^{(2,0)} + g^{(0,2)}\right)\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)}{2\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \\ &\quad + \boldsymbol{\Theta}^\top \boldsymbol{B}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) c(u,v), \end{aligned} \tag{3.13}$$

and

$$\text{avar}\left\{\hat{c}^*_{m1}(u,v)\right\} = \frac{\sigma^2_{m1}(u,v)}{nh^2} \equiv \frac{c(u,v)}{4\pi nh^2\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}. \tag{3.14}$$

Similarly, the effect of the pseudo-sample goes unnoticed asymptotically for $\hat{c}_{m1}$ as well, which is summarized in the following theorem.

Theorem 1. *Under Assumptions 1-5, the MTK estimator $\hat{c}_{m1}$ is such that for any $(u,v) \in (0,1)^2$,*

$$\sqrt{nh^2}\left(\hat{c}_{m1}(u,v) - c(u,v) - b_{m1}(u,v)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{m1}(u,v)\right),$$

where $b_{m1}(u,v)$ and $\sigma^2_{m1}(u,v)$ are defined as above.

Proof. See Appendix B.2. □

Compared to the $\hat{c}_{t1}$, it is seen that the $\hat{c}_{m1}$ introduces an additional bias correction term that involves the tuning parameters $\boldsymbol{\Theta}$ here. The asymptotic variance, however, remains the same. Interestingly, the variance formula is shared by may other copula density estimators, such as the kernel estimator $\tilde{c}^{(\tau,1)}$ in Geenens, Charpentier, and Paindaveine (2014) as well as the Beta kernel estimator and the Bernstein estimators derived in Janssen, Swanepoel, and Veraverbeke (2014).

I explore the global properties of the MTK, focusing on the weighted mean integrated squared error (weighted MISE). Let $w(u,v)$ be some non-negative weight function, then the weighted MISE is defined as

$$\text{wmise}\left\{\hat{c}\right\} = \mathbb{E}\left\{\int_{\mathcal{I}} \left(\hat{c}(u,v) - c(u,v)\right)^2 w(u,v)\mathrm{d}u\mathrm{d}v\right\}. \tag{3.15}$$

In this article, I set $w(u,v) = \phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))$ to guarantee the integrability of the weighted MISE. In fact, the weighted MISE in the UV-domain with this particular weight function amounts to the unweighted MISE in the ST-domain. Wand, Marron, and Ruppert (1991) note that the good performance in the transformed domain is usually translated into the original domain. This observation has been confirmed by many numerical experiments, including my simulations reported below. It is known that the asymptotic weighted MISE is equal to the sum of weighted integrated squared bias and weighted integrated variance. Thus, I have

$$\text{wmise}\left\{\hat{c}_{m1}\right\} \approx \frac{h^4}{4}\Gamma_3 + h^2\boldsymbol{\Theta}^\top\boldsymbol{\Gamma}_2 + \boldsymbol{\Theta}^\top\boldsymbol{\Gamma}_1\boldsymbol{\Theta} + \frac{1}{4\pi n h^2}, \tag{3.16}$$

where

$$\begin{aligned}
\boldsymbol{\Gamma}_1 &= \int_{\mathcal{R}^2} \boldsymbol{B}(s,t)\boldsymbol{B}(s,t)^\top g^2(s,t)\mathrm{d}s\mathrm{d}t \\
\boldsymbol{\Gamma}_2 &= \int_{\mathcal{R}^2} \boldsymbol{B}(s,t) g(s,t)\left(g^{(2,0)}(s,t) + g^{(0,2)}(s,t)\right)\mathrm{d}s\mathrm{d}t \\
\Gamma_3 &= \int_{\mathcal{R}^2} \left(g^{(2,0)}(s,t) + g^{(0,2)}(s,t)\right)^2 \mathrm{d}s\mathrm{d}t.
\end{aligned} \tag{3.17}$$

Then the optimal smoothing parameters, which minimize (3.16), are given by

$$h_{0,m1} = \left[\frac{1}{2\pi(\Gamma_3 - \boldsymbol{\Gamma}_2^\top\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2)}\right]^{1/6} n^{-1/6}, \tag{3.18}$$

and

$$\boldsymbol{\Theta}_{0,m1} = -\frac{h_{0,m1}^2}{2}\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2. \tag{3.19}$$

It follows that the optimal asymptotic weighted MISE is

$$\text{wmise}_0\left\{\hat{c}_{m1}\right\} \approx \frac{1}{4}\left(2\pi\right)^{-2/3}\left(\Gamma_3 - \boldsymbol{\Gamma}_2^\top \boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2\right)^{1/3} n^{-2/3}. \tag{3.20}$$

In Appendix B.2, I show that $\Gamma_3 - \boldsymbol{\Gamma}_2^\top \boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2 \geq 0$; therefore, $h_{0,m1}$ is generally well-defined. Moreover, $h_{0,m1}$ and $\boldsymbol{\Theta}_{0,m1}$ satisfy the conditions required for Theorem 1. The related results for the TKE, as a special case of the MTK with $\boldsymbol{\Theta} = 0$, are readily obtained as

$$h_{0,t} = \left[\frac{1}{2\pi\Gamma_3}\right]^{1/6} n^{-1/6}$$
$$\text{wmise}_0\left\{\hat{c}_t\right\} \approx \frac{1}{4}\left(2\pi\right)^{-2/3}\Gamma_3^{1/3} n^{-2/3}.$$

Since $\boldsymbol{\Gamma}_1$, by construction, is positive-semidefinite, it follows that $\boldsymbol{\Gamma}_2^\top \boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2 \geq 0$. Therefore, the MTK dominates the TKE in terms of the asymptotic weighted MISE (3.16), and has the usual convergence rate in bivariate kernel density estimation. Since the two estimators share the same asymptotic variance, it is understood that the reduction in the asymptotic weighted MISE comes from the bias correction term mentioned above.

3.4 Smoothing parameters selection

3.4.1 Plug in method

It is well known that smoothing parameters are crucial for kernel density estimators. The MTK, more precisely $\hat{c}_{m1}$ here, requires the selection of bandwidth h together with tuning parameters $\boldsymbol{\Theta}$. I have derived the optimal $h_{0,m1}$ and $\boldsymbol{\Theta}_{0,m1}$, see (3.18) and (3.19), which minimize the asymptotic weighted MISE (3.16). Thus, an immediate plug in method is replacing the unknown quantities there by the corresponding sample analogs. This requires estimations of $\boldsymbol{\Gamma}_1$, $\boldsymbol{\Gamma}_2$ and Γ_3 specified by (3.17) in ST-domain.

The estimation of Γ_3 has been studied in Wand and Jones (1995) and Duong and Hazelton (2003). Define $\psi_{r_1,r_2} = \int_{\mathcal{R}^2} g^{(r_1,r_2)}(s,t)g(s,t)\mathrm{d}s\mathrm{d}t$. Then Γ_3 can be decomposed as

$$\Gamma_3 = \psi_{4,0} + \psi_{0,4} + 2\psi_{2,2}.$$

This formulation is based on the fact that

$$\int_{\mathcal{R}^2} g^{(r_1,r_2)}(s,t)g^{(r_1',r_2')}(s,t)\mathrm{d}s\mathrm{d}t = \begin{cases} (-1)^{r_1+r_2}\psi_{r_1+r_1',r_2+r_2'} & \text{if } \sum_{i=1,2} r_i + r_i' \text{ is even} \\ 0 & \text{otherwise} \end{cases}$$

if the density g is sufficiently smooth; see Wand and Jones (1995) for the proof. Since ψ_{r_1,r_2} admits the expectation form $\mathbb{E}\left(g^{(r_1,r_2)}(S_i,T_i)\right)$, this motives us to estimate it nonparametrically as

$$\hat{\psi}_{r_1,r_2} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} K_b^{(r_1)}\left(\hat{S}_i - \hat{S}_j\right) K_b^{(r_2)}\left(\hat{T}_i - \hat{T}_j\right),$$

where b is the associated preliminary bandwidth, $K_b(x) = K(x/b)/b$ and $K_b^{(r)} = \mathrm{d}^r K_b(x)/\mathrm{d}x^r$. Following Duong and Hazelton (2003), I use the product kernel for simplicity, moreover, the univariate Gaussian kernel $K(x) = \phi(x)$ is used as usual. Then Γ_3 can be estimated term by term as above, and I have

(3.21)

$$\hat{\Gamma}_3 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Big\{K_b^{(4)}\left(\hat{S}_i - \hat{S}_j\right) K_b\left(\hat{T}_i - \hat{T}_j\right) + 2K_b^{(2)}\left(\hat{S}_i - \hat{S}_j\right) K_b^{(2)}\left(\hat{T}_i - \hat{T}_j\right) + K_b\left(\hat{S}_i - \hat{S}_j\right) K_b^{(4)}\left(\hat{T}_i - \hat{T}_j\right)\Big\}$$

The estimations of $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ are easier. Both can be easily written as

$$\mathbf{\Gamma}_1 = \mathbb{E}\left(\boldsymbol{B}(S_i,T_i)\boldsymbol{B}(S_i,T_i)^\top g(S_i,T_i)\right)$$
$$\mathbf{\Gamma}_2 = \mathbb{E}\left(\boldsymbol{B}(S_i,T_i)(g^{(2,0)}(S_i,T_i) + g^{(0,2)}(S_i,T_i))\right).$$

Thus I have them estimated nonparametrically by

$$\hat{\mathbf{\Gamma}}_1 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{B}\left(\hat{S}_i, \hat{T}_i\right) \boldsymbol{B}^\top\left(\hat{S}_i, \hat{T}_i\right) K_b\left(\hat{S}_i - \hat{S}_j\right) K_b\left(\hat{T}_i - \hat{T}_j\right) \tag{3.22}$$

and

$$\begin{aligned}\hat{\mathbf{\Gamma}}_2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{B}\left(\hat{S}_i, \hat{T}_i\right) \Big\{ & K_b^{(2)}\left(\hat{S}_i - \hat{S}_j\right) K_b\left(\hat{T}_i - \hat{T}_j\right) \\ & + K_b\left(\hat{S}_i - \hat{S}_j\right) K_b^{(2)}\left(\hat{T}_i - \hat{T}_j\right)\Big\},\end{aligned} \tag{3.23}$$

where b, K_b and $K_b^{(2)}$ are similarly defined as above.

Consequently, the desired smoothing parameters choices are given by

$$\hat{h}_{0,m1} = \left[\frac{1}{2\pi(\hat{\Gamma}_3 - \hat{\mathbf{\Gamma}}_2^\top \hat{\mathbf{\Gamma}}_1^{-1}\hat{\mathbf{\Gamma}}_2)}\right]^{1/6} n^{-1/6} \quad \text{and} \quad \hat{\mathbf{\Theta}}_{0,m1} = -\frac{\hat{h}_{0,m1}^2}{2}\hat{\mathbf{\Gamma}}_1^{-1}\hat{\mathbf{\Gamma}}_2. \tag{3.24}$$

Since Γ_3 is the most difficult to estimate, the choice of b is directed to Duong and Hazelton (2003), which is optimal for $\hat{\Gamma}_3$. Meanwhile this choice provides precise enough $\hat{\mathbf{\Gamma}}_1$ and $\hat{\mathbf{\Gamma}}_2$. Note that I use the same preliminary bandwidth b in $\hat{\mathbf{\Gamma}}_1^{-1}$, $\hat{\mathbf{\Gamma}}_2$ and $\hat{\Gamma}_3$. This is to avoid the situations where $\hat{\Gamma}_3 - \hat{\mathbf{\Gamma}}_2^\top \hat{\mathbf{\Gamma}}_1^{-1}\hat{\mathbf{\Gamma}}_2$ is negative, and thus makes $\hat{h}_{0,m1}$ invalid (Duong and Hazelton 2003).

3.4.2 Profile weighted cross validation

I provides an alternative way to select the smoothing parameters for the MTK. Least square cross validation is a commonly used tool for kernel density estimators, whose objective function is given by

$$\mathrm{WCV}\left(h, \lambda, \mathbf{\Theta}\right) = \int_{\mathcal{I}} \left(\hat{c}_m(u,v)\right)^2 w(u,v)\mathrm{d}u\mathrm{d}v - \frac{2}{n}\sum_{i=1}^{n} \hat{c}_m^{-i}\left(\hat{U}_i, \hat{V}_i\right) w\left(\hat{U}_i, \hat{V}_i\right), \tag{3.25}$$

where $w(u,v)$ is a weight function and $\hat{c}_m^{-i}\left(\hat{U}_i, \hat{V}_i\right)$ is the "leave-one-out" version of the $\hat{c}_m$ evaluated at the pseudo-data point $\left(\hat{U}_i, \hat{V}_i\right)$. As discussed earlier, setting $w(u,v) = 1$ leads to the unweighted cross validation in the UV-domain, while set-

ting $w(u,v) = \phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))$ amounts to the unweighted cross validation in the ST domain. My numerical experiments indicate that, for copula density estimation, the weighted cross validation performs considerably better, demonstrating the merit of smoothing parameters selection in the ST-domain. Therefore, I use the weight function $w(u,v) = \phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))$ hereafter. The second term of (3.25) can be evaluated straightforwardly. If taking the Gaussian kernel (3.6), the first term is shown to admit an analytical form, which is presented in Appendix B.3. However, a direct implementation of the weighted cross validation with respect to $(h, \mathbf{\Theta})$ seems impractical because it requires a difficult 3-dimensional minimization procedure. This motivates us to propose the profile weighted cross validation.

The profile weighted cross validation can be considered as a hybrid method that combines the weighted cross validation and the theoretical asymptotic results. Recall that I have derived $\mathbf{\Theta}_{0,m1} = -\frac{h_{0,m1}^2}{2}\mathbf{\Gamma}_1^{-1}\mathbf{\Gamma}_2$. If in the first place, I have $\mathbf{\Gamma}_1$ and $\mathbf{\Gamma}_2$ estimated as in (3.22) and (3.23), then I can treat $\mathbf{\Theta}_{0,m1}$ as a known function of h and conduct the weighted cross validation with respect to h alone. Below I provide a step-by-step description of this method:

- set $\mathbf{\Theta}_{0,m1}(h) = -\frac{h^2}{2}\hat{\mathbf{\Gamma}}_1^{-1}\hat{\mathbf{\Gamma}}_2$;
- plug $\mathbf{\Theta}_{0,m1}(h)$ into (3.25), conduct the minimization with respect to h and obtain $\hat{h}_{0,m1}$.

This procedure lowers the dimension of numerical optimization from three to one, reducing the computational burden considerably.

3.5 Higher order improvement for Gaussian copulas

If the underlying copula C is Gaussian copula, the transformed density g in the ST-domain is the pdf of bivariate Gaussian distribution. Let ρ be the associated

correlation coefficient. If I set $\boldsymbol{\Theta} = -\frac{h^2}{2}\boldsymbol{\Gamma}_1^{-1}\boldsymbol{\Gamma}_2$ according to (3.19), namely

$$\theta_1 = \frac{1+\rho^2}{(1-\rho^2)^2}\frac{h^2}{2} \quad \text{and} \quad \theta_2 = -\frac{4\rho}{(1-\rho^2)^2}\frac{h^2}{2}, \tag{3.26}$$

it is easy to check that the asymptotic bias term $b_{m1}(u,v)$ in (3.13) vanishes for any $(u,v) \in (0,1)^2$ regardless of the value of h. In fact, under Gaussian copulas, the $\hat{c}_{m1}$ reduces the asymptotic bias from the order $O\left(h^2\right)$ to the order $O\left(h^4\right)$. To see this, I first consider the $\hat{c}^*_{m1}$ as usual. Note that the bivariate Gaussian density is sufficiently smooth; thus it guarantees the subsequent higher order approximations are legitimate. Following the arguments in Charpentier, Fermanian, and Scaillet (2006), the asymptotic bias of $\hat{c}^*_{t1}$ can be extended by

(3.27)

$$\begin{aligned}\text{abias}\left\{\hat{c}^*_{t1}(u,v)\right\} = b^{(G)}_{t1}(u,v) \equiv & \frac{h^2\left(g^{(2,0)}+g^{(0,2)}\right)\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)}{2\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \\ & + \frac{h^4\left(g^{(4,0)}+g^{(0,4)}+2g^{(2,2)}\right)\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)}{8\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))},\end{aligned}$$

where the terms up to the fourth order are explicitly stated. Similarly, Taylor expansion for $J^*(u,v;h,\boldsymbol{\Theta})$ up to the second order yields

(3.28)

$$J^*(u,v;h,\boldsymbol{\Theta}) \approx 1+\boldsymbol{\Theta}^\top \boldsymbol{B}\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)+h^2\boldsymbol{\Theta}^\top\boldsymbol{A}_1+\frac{1}{2}\boldsymbol{\Theta}^\top\boldsymbol{A}_2\left(\Phi^{-1}(u),\Phi^{-1}(v)\right)\boldsymbol{\Theta},$$

where

$$\boldsymbol{A}_1 = \begin{pmatrix} 2 \\ \\ 0 \end{pmatrix}$$

and

$$\boldsymbol{A}_2(s,t) = \begin{pmatrix} (s^2+t^2-2)^2 - 4(1+\rho^2) & (st-\rho)(s^2+t^2-2) - 4\rho \\ (st-\rho)(s^2+t^2-2) - 4\rho & (st-\rho)^2 - (\rho^2+1) \end{pmatrix}$$

Combine the above results as I did previously and plug (3.26) in, one may easily find that

$$\text{abias}\left\{\hat{c}^*_{m1}(u,v)\right\} = b^{(G)}_{m1}(u,v) \equiv h^4 R\left(\Phi^{-1}(u), \Phi^{-1}(v); \rho\right),$$

which is of order $O\left(h^4\right)$ since the terms associated with h^2 are all canceled out. The explicit form of $R(\cdot,\cdot;\rho)$ is given by

$$R(s,t;\rho) = c\left(\Phi(s), \Phi(t)\right)\left\{\frac{-(1+3\rho^2)(s^2+t^2) + 2\rho(\rho^2+3)st + 2(1-\rho^4)}{2(1-\rho^2)^3}\right\}.$$

The asymptotic variance, of course, remains the same. To summarize, the following theorem is presented.

Theorem 2. *If the underlying copula C is Gaussian copula and let $\boldsymbol{\Theta}$ be selected according to* (3.26)*, under Assumptions 1-4 and 6, the MTK estimator $\hat{c}_{m1}$ is such that for any $(u,v) \in (0,1)^2$,*

$$\sqrt{nh^2}\left(\hat{c}_{m1}(u,v) - c(u,v) - b^{(G)}_{m1}(u,v)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{m1}(u,v)\right),$$

where $b^{(G)}_{m1}(u,v)$ is defined above and $\sigma^2_{m1}(u,v)$ is described in (3.14)*.*

Proof. See Appendix B.2. □

Likewise, the global properties of the MTK under Gaussian copulas are studied. The weighted MISE (3.15) is hereby approximated by

$$\text{wmise}^{(G)}\left\{\hat{c}_{m1}\right\} \approx h^8\Lambda + \frac{1}{4\pi nh^2},$$

where $\Lambda = \int_{\mathcal{R}^2} R^2(s,t;\rho)\phi^2(s)\phi^2(t)\mathrm{d}s\mathrm{d}t$. From its first order condition, the optimal

bandwidth is given by

$$h_{0,m1}^{(G)} = (16\pi\Lambda)^{-1/10}\, n^{-1/10},$$

which obviously satisfies the assumptions for Theorem 2. Then, together with the $\boldsymbol{\Theta}$ choices in (3.26), the optimal asymptotic weighted MISE has the form

$$\text{wmise}_0^{(G)}\left\{\hat{c}_{m1}\right\} \approx \frac{5}{16\pi}\,(16\pi\Lambda)^{1/5}\, n^{-4/5}.$$

It is seen that, when the underlying copula is Gaussian copula, the convergence rate of $\hat{c}_{m1}$ in terms of weighted MISE is $O\left(n^{-4/5}\right)$, which is faster than the usual rate $O\left(n^{-2/3}\right)$.

Consider the smoothing parameters selection under Gaussian copulas. Obviously, $\hat{c}_{m1}$ requires the optimal bandwidth $h_{0,m1}^{(G)}$ that has different convergence rate from the $h_{0,m1}$ defined in (3.18). The profile weighted cross validation method is adaptive, thus obtains the optimal bandwidth rate automatically. The plug in method derived under the lower order asymptotic analysis ceases to be optimal, but it remains viable in practice. My simulation experiments indicate that it still delivers performance better than or comparable to other competing copula density estimators, though it is dominated by the profile weighted cross validation method.

The merit that the MTK enjoys higher order convergence rate for Gaussian copulas has many practical implications. In finance, the Gaussian copula family has been widely used in managing risks and pricing portfolios, however, this parametric method is criticized for restrictiveness in practice. In fact, the copulas embedded in many financial data (specially the residuals obtained after some pre-processing models), although unlikely exactly Gaussian, are near Gaussian with deviations such as asymmetry, fat tails, etc. Intuitively, the MTK shall provide superior performance for near Gaussian copulas as well. This is confirmed in my simulation experiments.

Therefore, the MTK is an appealing copula density estimator for such situations.

3.6 Extension to non-diagonal bandwidth matrix

It is often desirable to keep the bandwidth matrix $\boldsymbol{H}$ non-diagonal since the off-diagonal element controls the direction towards which the smoothing is placed. This is sensible in copula density estimation (Geenens, Charpentier, and Paindaveine 2014; Duong and Hazelton 2005). Consider the bandwidth matrix

$$\boldsymbol{H} = h^2 \begin{pmatrix} 1 & \lambda \\ \lambda & 1 \end{pmatrix}, \tag{3.29}$$

where $-1 < \lambda < 1$. Note that I use the same bandwidth for the two margins here; this is mainly for simplicity consideration. Then the MTK becomes, if similarly Gaussian kernel (3.6) is used,

$$\hat{c}_{m2}(u,v) = \frac{\exp\left(-\theta_1\left(\{\Phi^{-1}(u)\}^2 + \{\Phi^{-1}(v)\}^2\right) - \theta_2\Phi^{-1}(u)\Phi^{-1}(v)\right)}{n\eta h^2 \phi\left(\Phi^{-1}(u)\right)\phi\left(\Phi^{-1}(v)\right)} \sum_{i=1}^{n} \phi_2\left(\frac{\Phi^{-1}(u) - \hat{S}_i}{h}, \frac{\Phi^{-1}(v) - \hat{T}_i}{h}\right), \tag{3.30}$$

where ϕ_2 is the bivariate Gaussian density

$$\phi_2(x,y) = \frac{1}{2\pi\sqrt{1-\lambda^2}} \exp\left(-\frac{x^2 + y^2 - 2\lambda xy}{2(1-\lambda^2)}\right).$$

In this case, the normalization term η admits the form

$$\eta = \frac{1}{n\delta} \sum_{i=1}^{n} \exp\left\{-\frac{(4h^2\theta_1^2 - h^2\theta_2^2 + 2\theta_1)(\hat{S}_i^2 + \hat{T}_i^2) + (2\lambda h^2\theta_2^2 - 8\lambda h^2\theta_1^2 + 2\theta_2)\hat{S}_i\hat{T}_i}{2\delta^2}\right\},$$

where

$$\delta = \sqrt{h^4(1-\lambda^2)(4\theta_1^2 - \theta_2^2) + 2h^2(2\theta_1 + \lambda\theta_2) + 1}.$$

The properties of the $\hat{c}^*_{m2}$ computed on $\{(S_i, T_i), i = 1, \cdots, n\}$ are examined and then extended to the feasible $\hat{c}_{m2}$ informally. The asymptotic bias and variance of the $\hat{c}^*_{m2}$ for any $(u, v) \in (0, 1)^2$ are given by

$$\text{abias}\left\{\hat{c}^*_{m2}(u,v)\right\} = \frac{h^2\left(g^{(2,0)} + g^{(0,2)} + 2\lambda g^{(1,1)}\right)\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)}{2\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} + \boldsymbol{\Theta}^\top \boldsymbol{B}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) c(u,v) \tag{3.31}$$

and

$$\text{avar}\left\{\hat{c}^*_{m2}(u,v)\right\} = \frac{c(u,v)}{4\pi n h^2\sqrt{1-\lambda^2}\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}. \tag{3.32}$$

Similarly, the first term in (3.31) is the asymptotic bias of $\hat{c}^*_{t2}$, which denotes the TKE under the bandwidth matrix (3.29). The second term is introduced for bias correction. This adjustment term is the same as that for the $\hat{c}^*_{m1}$ case since the Taylor expansion of $J^*(u, v; h, \boldsymbol{\Theta})$ here is identical to (3.12). For the asymptotic variance (3.32), it is again a duplicate of the corresponding counterpart for $\hat{c}^*_{t2}$.

Intuitively, the asymptotic properties of the $\hat{c}_{m2}$ can be inferred from (3.31) and (3.32) by neglecting the effect of pseudo-sample since the $\hat{c}_{m2}$ with non-diagonal bandwidth matrix is qualitatively equivalent to the $\hat{c}_{m1}$. Then for global properties, the weighted MISE as defined in (3.15), is approximated by

$$\text{wmise}\left\{\hat{c}_{m2}\right\} \approx \frac{h^4}{4}\Gamma_3(\lambda) + h^2\boldsymbol{\Theta}^\top\boldsymbol{\Gamma}_2(\lambda) + \boldsymbol{\Theta}^\top\boldsymbol{\Gamma}_1\boldsymbol{\Theta} + \frac{1}{4\pi n h^2\sqrt{1-\lambda^2}}, \tag{3.33}$$

where $\boldsymbol{\Gamma}_1$ is defined in (3.17) and

$$\boldsymbol{\Gamma}_2(\lambda) = \int_{\mathcal{R}^2} \boldsymbol{B}(s,t)g(s,t)\left(g^{(2,0)}(s,t) + g^{(0,2)}(s,t) + 2\lambda g^{(1,1)}(s,t)\right)\mathrm{d}s\mathrm{d}t$$
$$\Gamma_3(\lambda) = \int_{\mathcal{R}^2} \left(g^{(2,0)}(s,t) + g^{(0,2)}(s,t) + 2\lambda g^{(1,1)}(s,t)\right)^2 \mathrm{d}s\mathrm{d}t.$$

It is noted that when $\lambda = 0$, $\hat{c}_{m2}$ reduces to $\hat{c}_{m1}$, moreover $\boldsymbol{\Gamma}_2(\lambda)$ and $\Gamma_3(\lambda)$ become $\boldsymbol{\Gamma}_2$ and Γ_3 defined in (3.17). To minimize the asymptotic weighted MISE (3.33), the

optimal smoothing parameters are given by, in parallel to (3.18) and (3.19),

$$h_{0,m2} = \left[\frac{1}{2\pi\sqrt{1-\lambda^2}\left(\Gamma_3(\lambda) - \mathbf{\Gamma}_2(\lambda)^\top \mathbf{\Gamma}_1^{-1} \mathbf{\Gamma}_2(\lambda)\right)} \right]^{1/6} n^{-1/6}$$

and

$$\mathbf{\Theta}_{0,m2} = -\frac{h_{0,m2}^2}{2} \mathbf{\Gamma}_1^{-1} \mathbf{\Gamma}_2(\lambda).$$

The first order condition with respect to λ is complicated but it is equivalent to the following minimization problem. The optimal λ, denoted by λ_0, is given by

$$\begin{aligned} \underset{\lambda}{\text{minimize}} \quad & \frac{\Gamma_3(\lambda) - \mathbf{\Gamma}_2(\lambda)^\top \mathbf{\Gamma}_1^{-1} \mathbf{\Gamma}_2(\lambda)}{1-\lambda^2} \\ \text{subject to} \quad & -1 < \lambda < 1. \end{aligned}$$

The optimal asymptotic weighted MISE of $\hat{c}_{m2}$ follows as

$$\text{wmise}_0\left\{\hat{c}_{m2}\right\} \approx \min_{-1<\lambda<1} \quad \frac{1}{4}(2\pi)^{-2/3} \left(\frac{\Gamma_3(\lambda) - \mathbf{\Gamma}_2(\lambda)^\top \mathbf{\Gamma}_1^{-1} \mathbf{\Gamma}_2(\lambda)}{1-\lambda^2} \right)^{1/3} n^{-2/3}.$$

Since the $\text{wmise}_0\left\{\hat{c}_{m1}\right\}$ specified in (3.20) is a special case of the $\text{wmise}_0\left\{\hat{c}_{m2}\right\}$, I have $\text{wmise}_0\left\{\hat{c}_{m1}\right\} \geq \text{wmise}_0\left\{\hat{c}_{m2}\right\}$. Therefore, it is expected that the $\hat{c}_{m2}$ dominates $\hat{c}_{m1}$.

When the underlying copula is Gaussian copula and let $\mathbf{\Theta} = -\frac{h^2}{2}\mathbf{\Gamma}_1^{-1}\mathbf{\Gamma}_2(\lambda)$ regardless of λ value, one can easily check that the $\hat{c}_{m2}$ also reduces the asymptotic bias from the order $O\left(h^2\right)$ to the order $O\left(h^4\right)$. Thus, similar to $\hat{c}_{m1}$, the optimal bandwidth $h_{0,m2}^{(G)} \sim n^{-1/10}$ and the optimal asymptotic weighted MISE $\text{wmise}_0^{(G)}\left\{\hat{c}_{m2}\right\}$ is $O\left(n^{-4/5}\right)$.

The smoothing parameters choices of $\hat{c}_{m2}$ require more work as an additional λ is introduced. To begin with, I need to estimate $\mathbf{\Gamma}_2(\lambda)$ and $\Gamma_3(\lambda)$ in the first place, denoted by $\hat{\mathbf{\Gamma}}_2(\lambda)$ and $\hat{\Gamma}_3(\lambda)$. The estimate of $\mathbf{\Gamma}_1$ has already been given in (3.22). I

obtain $\hat{\mathbf{\Gamma}}_2(\lambda)$ by a simple generalization of (3.23), namely

$$\begin{aligned}\hat{\mathbf{\Gamma}}_2(\lambda) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{B}\left(\hat{S}_i,\hat{T}_i\right)\Big\{K_b^{(2)}\left(\hat{S}_i-\hat{S}_j\right)K_b\left(\hat{T}_i-\hat{T}_j\right) \\ +K_b\left(\hat{S}_i-\hat{S}_j\right)K_b^{(2)}\left(\hat{T}_i-\hat{T}_j\right)+2\lambda K_b^{(1)}\left(\hat{S}_i-\hat{S}_j\right)K_b^{(1)}\left(\hat{T}_i-\hat{T}_j\right)\Big\}.\end{aligned} \tag{3.34}$$

Note that $\Gamma_3(\lambda)$ can be decomposed to

$$\Gamma_3(\lambda) = \psi_{4,0}+\psi_{0,4}+(4\lambda^2+2)\psi_{2,2}+4\lambda\psi_{3,1}+4\lambda\psi_{1,3}.$$

Then each term is estimated separately and together they are combined to form $\hat{\Gamma}_3(\lambda)$ as

$$\begin{aligned}\hat{\Gamma}_3(\lambda) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\Big\{K_b^{(4)}\left(\hat{S}_i-\hat{S}_j\right)K_b\left(\hat{T}_i-\hat{T}_j\right) \\ +(4\lambda^2+2)K_b^{(2)}\left(\hat{S}_i-\hat{S}_j\right)K_b^{(2)}\left(\hat{T}_i-\hat{T}_j\right) \\ +K_b\left(\hat{S}_i-\hat{S}_j\right)K_b^{(4)}\left(\hat{T}_i-\hat{T}_j\right)+4\lambda K_b^{(3)}\left(\hat{S}_i-\hat{S}_j\right)K_b^{(1)}\left(\hat{T}_i-\hat{T}_j\right) \\ +4\lambda K_b^{(1)}\left(\hat{S}_i-\hat{S}_j\right)K_b^{(3)}\left(\hat{T}_i-\hat{T}_j\right)\Big\}.\end{aligned} \tag{3.35}$$

Again, I use the univariate Gaussian kernel $K(x) = \phi(x)$ and applies the same preliminary bandwidth b, as provided in Duong and Hazelton (2003). Then the plug in method for smoothing parameters selection is described as follows.

- obtain $\hat{\lambda}_0$ by numerically minimizing $\left\{\hat{\Gamma}_3(\lambda)-\hat{\mathbf{\Gamma}}_2(\lambda)^\top\hat{\mathbf{\Gamma}}_1^{-1}\hat{\mathbf{\Gamma}}_2(\lambda)\right\}/\left\{1-\lambda^2\right\}$ with the constraint that $-1<\lambda<1$;
- let $\hat{h}_{0,m2} = \left\{2\pi\sqrt{1-\hat{\lambda}_0^2}\left(\hat{\Gamma}_3(\hat{\lambda}_0)-\hat{\mathbf{\Gamma}}_2(\hat{\lambda}_0)^\top\hat{\mathbf{\Gamma}}_1^{-1}\hat{\mathbf{\Gamma}}_2(\hat{\lambda}_0)\right)\right\}^{-1/6}n^{-1/6}$;
- finally have $\hat{\mathbf{\Theta}}_{0,m2} = -\hat{h}_{0,m2}^2\hat{\mathbf{\Gamma}}_1^{-1}\hat{\mathbf{\Gamma}}_2(\hat{\lambda}_0)/2$.

The profile weighted cross validation is easily adapted to the $\hat{c}_{m2}$ case. A step-by-step description is presented as

- similarly obtain $\hat{\lambda}_0$ by following the first step in the above plug in method;
- set $\boldsymbol{\Theta}_{0,m2}(h) = -h^2\hat{\boldsymbol{\Gamma}}_1^{-1}\hat{\boldsymbol{\Gamma}}_2(\hat{\lambda}_0)/2$;
- plug $\hat{\lambda}_0$ and $\boldsymbol{\Theta}_{0,m2}(h)$ into (3.25), conduct the minimization with respect to h and obtain $\hat{h}_{0,m2}$.

3.7 Monte Carlo simulation

I conduct a simulation to compare the finite sample performance of several competing copula density estimators. In this section, some notations may be abused in order for consistency with the original literature; nevertheless, this should not cause any ambiguity. The following estimators are considered.

- The proposed MTK $\hat{c}_{m1}$ and $\hat{c}_{m2}$ based on the plug in method and the profile weighted cross validation.
- Gijbels and Mielniczuk (1990)'s mirror reflection estimator $\hat{c}_r$ with the bandwidth matrix $\boldsymbol{H} = h^2\boldsymbol{I}$, where h is selected by least square cross validation for flexibility.
- The TKE $\hat{c}_{t1}$ and $\hat{c}_{t2}$. The required parameters are chosen analogously to my proposed plug in method and profile weighted cross validation method by setting $\boldsymbol{\Theta} = 0$.
- The Beta kernel estimator considered in Charpentier, Fermanian, and Scaillet (2006) with Chen (1999)'s further bias correction. The selection of h remains vague in the literature. Following Geenens, Charpentier, and Paindaveine (2014), I consider two arbitrary values: $h = 0.02$ denoted by $\hat{c}_{b1}$ and $h = 0.05$ by $\hat{c}_{b2}$.

- The penalized hierarchical B-splines estimator. The parameters d and D are set to 4 and 8 according to Kauermann, Schellhase, and Ruppert (2013). The vector of penalty coefficients are selected as $\lambda = (10, 10)$ in $\hat{c}_{p1}$, $\lambda = (100, 100)$ in $\hat{c}_{p2}$ and $\lambda = (1000, 1000)$ in $\hat{c}_{p3}$.

I simulate data from some parametric copula families with sample sizes $n = 100$ and subsequently $n = 500$. For each family, with appropriate parameters, I look at two copula densities of which the Kendall τ's are 0.3 and 0.6, respectively. Specifically, I consider two groups of parametric copula families. In the first group, I highlight the Gaussian copulas and some near Gaussian copulas.

(A) The Gaussian copula, with parameters $\rho = 0.454$ and $\rho = 0.809$.

(B) The mixture of $\omega_1 = 85\%$ Gaussian copula and $\omega_2 = 15\%$ Clayton copula with two pairs of parameters $\big(\rho = 0.454, \theta = 6/7\big)$ and $(\rho = 0.809, \theta = 3)$. These mixture copulas are asymmetric and they place more dependence in the lower tail than in the upper tail.

(C) The Student t-copula with 15 degrees of freedom, with parameters $\rho = 0.454$ and $\rho = 0.809$. These copulas are close to Gaussian copula but possesses fat tails.

(D) The mixture of $\omega_1 = 85\%$ Student t-copula with 15 degrees of freedom and $\omega_2 = 15\%$ Clayton copula. Similarly, I set the two pairs of parameters to be $\big(\rho = 0.454, \theta = 6/7\big)$ and $(\rho = 0.809, \theta = 3)$. Thus, these mixture copulas are featured by both asymmetry and fat tails.

In the second group, I consider some other commonly used copula families.

(E) The Student t-copula with 5 degrees of freedom, with parameters $\rho = 0.454$ and $\rho = 0.809$.

(F) The Frank copula, with parameters $\theta = 2.92$ and $\theta = 7.93$.

(G) The Gumbel copula, with parameters $\theta = 10/7$ and $\theta = 2.5$.

(H) The Clayton copula, with parameters $\theta = 6/7$ and $\theta = 3$.

For each sample, the corresponding pseudo-sample is obtained for estimation. I evaluate the performance of each competing estimator $\hat{c}$ by the mean integrated squared error (MISE). Given each combination of the estimator $\hat{c}$ and the true copula density c, I compute

$$\mathrm{ISE}(\hat{c}, c) \approx \frac{1}{99^2} \sum_{k=1}^{99} \sum_{l=1}^{99} \left(\hat{c}(k/100, l/100) - c(k/100, l/100)\right)^2$$

and further estimate the corresponding MISE by averaging the obtained ISE's over 1000 simulated samples. These approximated MISE's are reported in Table 3.1 and 3.2. Bold values and underlined values show the minimum and the second minimum MISE's, respectively.

Doubtlessly, the MTK provides the overall best performance. It has improved upon the TKE substantially, demonstrating the success of using tapering method for further boundary biases correction. It is known that the mirror reflection estimator is particularly appropriate when the copula density has zero partial derivatives on the boundaries; the penalized B-splines estimator with large penalty coefficients is a strong competitor for flat copulas; the Beta kernel estimator performs quite well when the dependence is low but behaves badly when the copula is unbounded. However, in general, these estimators fail to compete with the MTK. For Gaussian copulas and near Gaussian copulas in the first group, the MTK beats all other estimators. This is understood because it possesses higher order convergence rate for Gaussian copulas, indicated by my theoretical results. For other common copulas in the second group, the MTK still dominates in most cases at the usual convergence rate. Not

Table 3.1: Simulation results $n = 100$

Copula	$\hat{c}_r$	$\hat{c}_{t1}$		$\hat{c}_{t2}$		$\hat{c}_{b1}$	$\hat{c}_{b2}$	$\hat{c}_{p1}$	$\hat{c}_{p2}$	$\hat{c}_{p3}$	$\hat{c}_{m1}$		$\hat{c}_{m2}$	
		Plug in	CV	Plug in	CV						Plug in	CV	Plug in	CV
A3	0.0711	0.0929	0.0888	0.0920	0.0907	0.2351	0.0781	0.0377	0.0299	0.0387	0.0501	0.0380	0.0459	**0.0212**
A6	0.2925	0.2016	0.1646	0.1586	0.1363	0.2081	0.1751	0.2625	0.4492	0.6666	0.1407	0.1119	0.1041	**0.0751**
B3	0.0768	0.0924	0.0904	0.0932	0.0917	0.2385	0.0793	0.0431	0.0366	0.0460	0.0475	**0.0242**	0.0470	0.0248
B6	0.3332	0.1913	0.1859	0.1715	0.1576	0.2361	0.2202	0.3086	0.4992	0.7152	0.1237	**0.0933**	0.1202	0.1034
C3	0.0829	0.0957	0.0933	0.0953	0.0932	0.2409	0.0815	0.0476	0.0450	0.0554	0.0498	0.0272	0.0478	**0.0265**
C6	0.3408	0.1830	0.1808	0.1636	0.1483	0.2222	0.2186	0.3246	0.5266	0.7438	0.1175	**0.0851**	0.1088	0.0933
D3	0.0896	0.0971	0.0916	0.0960	0.0939	0.2468	0.0852	0.0505	0.0497	0.0616	0.0500	**0.0286**	0.0490	0.0297
D6	0.3878	0.2003	0.2086	0.1799	0.1738	0.2555	0.2616	0.3730	0.5713	0.7882	0.1391	**0.1113**	0.1299	0.1206
E3	0.1221	0.1114	0.1065	0.1075	0.1029	0.2551	0.1012	0.0776	0.0912	0.1098	0.0615	0.0608	0.0600	**0.0477**
E3	0.4770	0.2364	0.2333	0.1850	0.1875	0.2830	0.3388	0.4957	0.7192	0.9391	0.1693	0.2147	**0.1401**	0.1626
F3	0.0509	0.0927	0.0961	0.0954	0.1019	0.2340	0.0718	0.0273	**0.0159**	0.0251	0.0462	0.0224	0.0475	0.0242
F6	0.1459	0.1897	0.1825	0.1738	0.1655	0.2007	0.0917	0.1007	0.2726	0.4823	0.1163	**0.0769**	0.1100	0.0791
G3	0.1426	0.1045	0.1082	0.1042	0.1042	0.2511	0.1130	0.0987	0.1104	0.1273	0.0693	0.0832	0.0659	**0.0651**
G6	0.6469	0.3098	0.3463	0.2469	0.2820	0.4034	0.5032	0.6762	0.8920	1.1118	0.2531	0.3972	**0.2320**	0.3028
H3	0.1793	0.1064	0.1151	0.1001	0.1103	0.2474	0.1314	0.1297	0.1564	0.1840	0.0799	0.1211	**0.0757**	0.0951
H6	1.2853	0.6482	0.7560	**0.5488**	0.6086	0.8790	1.1358	1.4016	1.6351	1.8615	0.6178	1.0193	0.5989	0.7520

surprisingly, I observe the $\hat{c}_{p2}$ provides the best performance for the Frank copula with Kendall $\tau = 0.3$ since it is flat enough. For the Clayton copula with Kendall $\tau = 0.6$, $\hat{c}_{t2}$ seems slightly outperform the MTK. This particular copula density is featured by extremely high lower tail dependence, thus tends to infinity sharply in the $(0, 0)$-corner. The MTK that tries to prevent its estimates from exploding then becomes suboptimal for this case, as explained in Geenens, Charpentier, and Paindaveine (2014). It is clear that all simulation results show the superior performance of my proposed estimator.

Next I take a closer look at the MTK. In general, the $\hat{c}_{m2}$ produces more accurate estimates than the $\hat{c}_{m1}$, confirming that non-diagonal bandwidth matrix which allows directional smoothing is sensible for copula density estimation. Still, some exceptions are present: when the sample size $n = 100$, $\hat{c}_{m1}$ hits the best performance for several

Table 3.2: Simulation results $n = 500$

Copula	$\hat{c}_r$	$\hat{c}_{t1}$		$\hat{c}_{t2}$		$\hat{c}_{b1}$	$\hat{c}_{b2}$	$\hat{c}_{p1}$	$\hat{c}_{p2}$	$\hat{c}_{p3}$	$\hat{c}_{m1}$		$\hat{c}_{m2}$	
		Plug in	CV	Plug in	CV						Plug in	CV	Plug in	CV
A3	0.0299	0.0365	0.0354	0.0361	0.0358	0.0494	0.0205	0.0207	0.0203	0.0310	0.0185	0.0093	0.0180	**0.0083**
A6	0.1223	0.0803	0.0689	0.0701	0.0576	0.0553	0.1321	0.1711	0.3049	0.5232	0.0528	0.0443	0.0483	**0.0249**
B3	0.0339	0.0373	0.0363	0.0367	0.0361	0.0503	0.0229	0.0240	0.0255	0.0376	0.0191	**0.0095**	0.0187	0.0097
B6	0.1570	0.0882	0.0849	0.0775	0.0693	0.0812	0.1730	0.2171	0.3539	0.5723	0.0604	0.0437	0.0562	**0.0418**
C3	0.0366	0.0386	0.0367	0.0381	0.0376	0.0508	0.0247	0.0262	0.0303	0.0463	0.0196	0.0105	0.0195	**0.0101**
C6	0.1490	0.0833	0.0791	0.0742	0.0623	0.0685	0.1729	0.2286	0.3746	0.6007	0.0552	0.0338	0.0513	**0.0321**
D3	0.0403	0.0390	0.0381	0.0385	0.0377	0.0509	0.0271	0.0302	0.0351	0.0517	0.0202	**0.0116**	0.0199	**0.0116**
D6	0.1881	0.0952	0.0935	0.0799	0.0726	0.0965	0.2140	0.2763	0.4216	0.6453	0.0671	0.0534	0.0624	**0.0500**
E3	0.0582	0.0442	0.0426	0.0418	0.0402	0.0551	0.0400	0.0477	0.0642	0.0961	0.0258	0.0235	0.0241	**0.0211**
E3	0.2270	0.1041	0.1010	0.0851	0.0768	0.1169	0.2891	0.3831	0.5529	0.7978	0.0719	0.0976	0.0637	**0.0602**
F3	0.0175	0.0376	0.0379	0.0379	0.0402	0.0476	0.0153	0.0135	**0.0069**	0.0162	0.0191	0.0115	0.0194	0.0119
F6	0.0514	0.0824	0.0803	0.0719	0.0684	0.0468	0.0536	0.0464	0.1397	0.3419	0.0544	0.0423	0.0489	**0.0384**
G3	0.0758	0.0440	0.0442	0.0412	0.0415	0.0582	0.0547	0.0673	0.0859	0.1146	0.0300	0.0401	**0.0297**	0.0318
G6	0.3653	0.1456	0.1594	0.1125	0.1197	0.2356	0.4524	0.5554	0.7302	0.9680	0.1221	0.2462	**0.1073**	0.1418
H3	0.0958	0.0450	0.0470	0.0423	0.0438	0.0599	0.0716	0.0925	0.1213	0.1651	0.0367	0.0616	**0.0340**	0.0458
H6	0.8372	0.3778	0.3863	**0.2631**	0.2837	0.6956	1.0772	1.2435	1.4625	1.7118	0.3719	0.4547	0.3009	0.3501

occasions. This may be because a small sample has difficulty in detecting the optimal λ. For Gaussian copulas and near Gaussian copulas, the profile weighted cross validation method is clearly doing better than the plug in method, which again has verified my belief that the former is adaptive but the latter is not optimal any more since it is derived based on smaller order asymptotic approximation. Nevertheless, the plug in method still provides good enough estimates, in most cases even better than other estimators. For the copulas in the second group, the plug in method seems to excel for most times. This is also understandable because it converges to the true smoothing parameters faster than the profile weighted cross validation method, which is well known in the literature. Since there is not a method that can totally dominates in all cases, I suggest first trying both the plug in method and the profile weighted cross validation method in practice and then making a choice based on

one's specific purposes or prior beliefs. From my experience, more often, the profile weighted cross validation method produces more visually pleasant estimates.

3.8 Empirical applications

3.8.1 Loss and ALAE data

I first consider the classic loss and ALAE data in an insurance context, which reports the logarithms of the indemnity payment and allocated loss adjustment expense from 1500 insurance claims. Copulas are employed to model the dependence between these two variables. This data has been widely used in the literature to illustrate copula fitting and goodness-of-fit testing, see Frees and Valdez (1998); Klugman and Parsa (1999); Chen et al. (2010), among others. It is generally agreed that the Gumbel copula with parameter 1.453 provides the best fit out of the usual parametric copula families. I use my proposed $\hat{c}_{m2}$ to estimate the copula and select the smoothing parameters according to the profile weighted cross validation. Then, I compare $\hat{c}_{m2}$ with the above parametric fit as well as $\hat{c}_{t2}$. Before estimation, I have removed 34 censored observations from the dataset, thus the sample size here is $n = 1466$. The estimation results are reported in Figure 3.1. From the 3-d plots, $\hat{c}_{t2}$ is obviously quite wiggly in the corners. On the contrary, $\hat{c}_{m2}$ has produced very smooth estimates that are similar to the parametric fit; visually, it is difficult to distinguish between the two. This can be also seen from the MTK contour plot, where lines almost match each other. However, several notable differences are still present. For example, near the $(1, 1)$-corner, $\hat{c}_{m2}$ grows more slowly than the parametric fit does, moreover $\hat{c}_{m2}$ estimates are not symmetric about the 45 degrees diagonal line of $\mathcal{I}$. Of course, I have no idea what the true underlying copula is. Since $\hat{c}_{m2}$ is fully nonparametric and does not require any prior assumption, it should be closer to the truth with confidence.

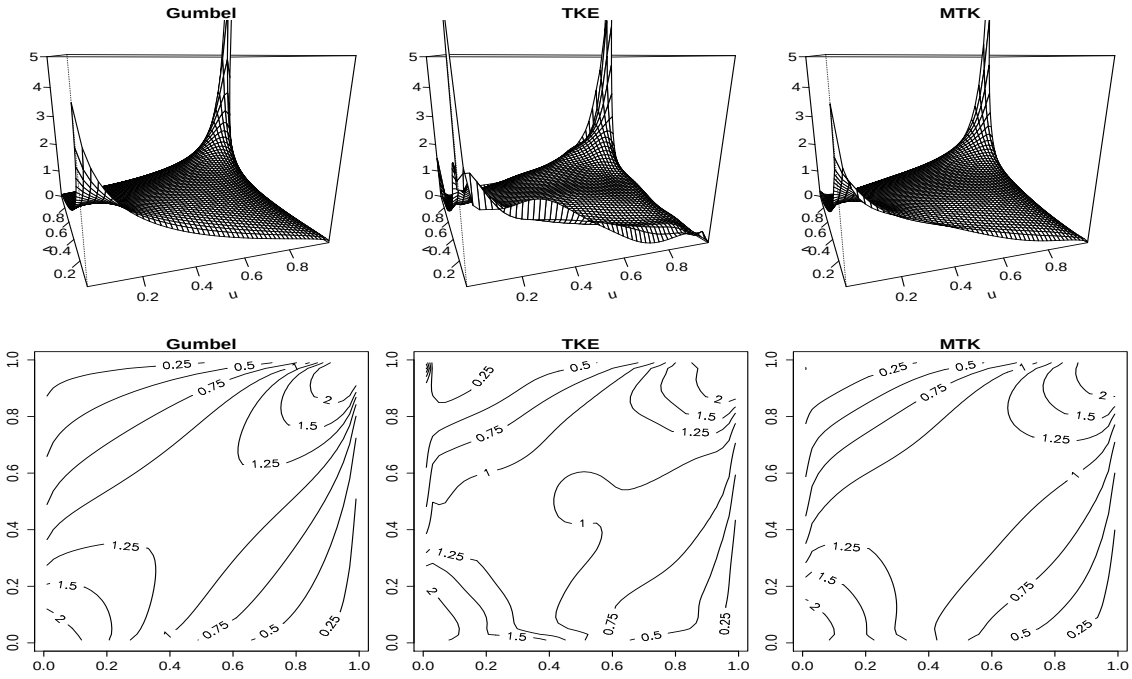


Figure 3.1: Loss and ALAE data: in the two contour plots, black line denotes parametric estimates and blue line denotes TKE or MTK estimates.

3.8.2 Uranium exploration data

As a second illustration, I consider the uranium exploration data which was originally studied in Cook and Johnson (1981, 1986) and later in the copula literature, for example Genest, Quessy, and Rmillard (2006); Chen and Huang (2007). This data was collected from water samples in the Montrose quadrangle of west Colorado and consists of 655 concentrations measured for seven chemical elements including uranium (U) and lithium (Li). My interest is to model the dependence between U and Li by estimating their associated copula. Based on a Cramér-Von Mises type test statistic, Chen and Huang (2007) conclude the Student t-copula with 59 degrees of freedom and correlation parameter 0.17 seems to provide the best dependence description for U and Li. Based on this parametric fit, it is reasonable to expect that the copula of U and Li is near Gaussian and possesses fat tails. Similarly, I use $\hat{c}_{m2}$

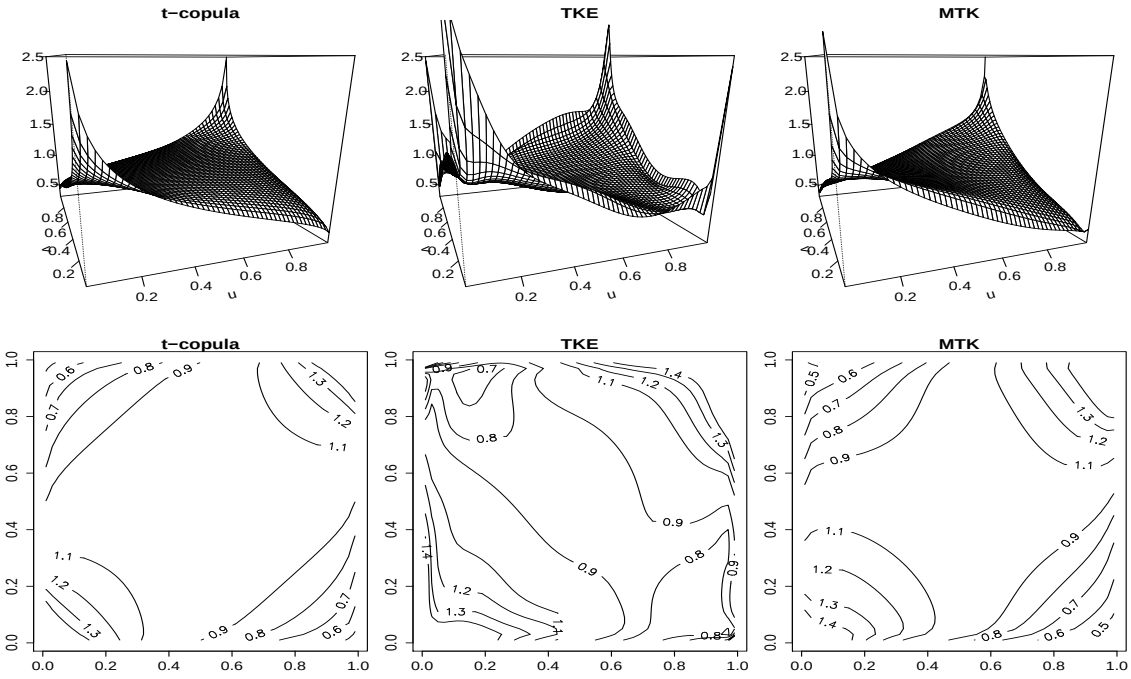


Figure 3.2: Uranium exploration data: in the two contour plots, black line denotes parametric estimates and blue line denotes TKE or MTK estimates.

to estimate the copula and compare it with this t-copula fit as well as $\hat{c}_{t2}$. Figure 3.2 displays my estimation results. From the 3-d plots, $\hat{c}_{m2}$ again yields very pleasant appearance that closely resembles the t-copula fit. In $(0,1)$- and $(1,0)$-corners, both the $\hat{c}_{m2}$ and the t-copula fit tend towards 0.5 approximately, suggesting the feature of fat tails for the underlying copula. In fact, from the MTK contour plot, it is obvious that the $\hat{c}_{m2}$ produces even 'fatter' tails around the four corners than the t-copula fit does. Not surprisingly, $\hat{c}_{t2}$ still performs badly for this data, with very irregular behaviors. For example, it spuriously explodes upwards in the $(0,1)$- and $(1,0)$-corners. As I have stressed earlier, this is caused by the unbounded multiplier associated with back transformation. Thus, it seems that using tapering method to mitigate the influence of this multiplier is quite successful from this application.

3.8.3 FTSE 100 and Hang Seng indexes

Lastly I consider estimating copula to detect the dynamic relationship between two financial time series. Specifically, I look at the weekly log returns of FTSE 100 (London Stock Exchange) and Hang Seng (Hong Kong Stock Market) indexes covering the period from January 2010 till December 2013. In total, I have $n = 729$ observations. For each series $r_t : t = 1, \cdots, n$, I assume a GARCH(1,1) model, i.e. $r_t = \mu + h_t$, $h_t = \sigma_t \epsilon_t$ and $\sigma_t^2 = \kappa + \alpha h_{t-1}^2 + \beta \sigma_{t-1}^2$ where the standardized residuals $\epsilon_t : t = 1, \cdots, n$ are assumed i.i.d. from Student t-distribution with zero mean and unity variance. The primary task here is to model the copula embedded in the standardized residuals obtained from the two return series. Some related graphs are presented in Figure 3.3. It is clear from the pseudo-sample scatterplot that the bottom left quarter of $\mathcal{I}$ contains considerably more observations compared to the top right quarter, indicating the London and the Hong Kong stock markets exhibit stronger lower tail dependence but relatively weaker upper tail dependence. This is consistent with our prior belief in practice: bear markets move together more likely than bull markets do. Therefore, it is reasonable to expect the true underlying copula should be asymmetric and its tail in the $(0, 0)$-corner should be higher than the tail in the $(1, 1)$-corner. From the 3-d plots, both $\hat{c}_{m2}$ and $\hat{c}_{t2}$ have successfully revealed this feature. Again, $\hat{c}_{m2}$ produces very smooth appearance and its contour lines looks regular. In contrast, $\hat{c}_{t2}$ is less smooth with obvious bumps, which are particularly clear from the contour plot. This financial data another time confirms the good performance of my proposed estimator.

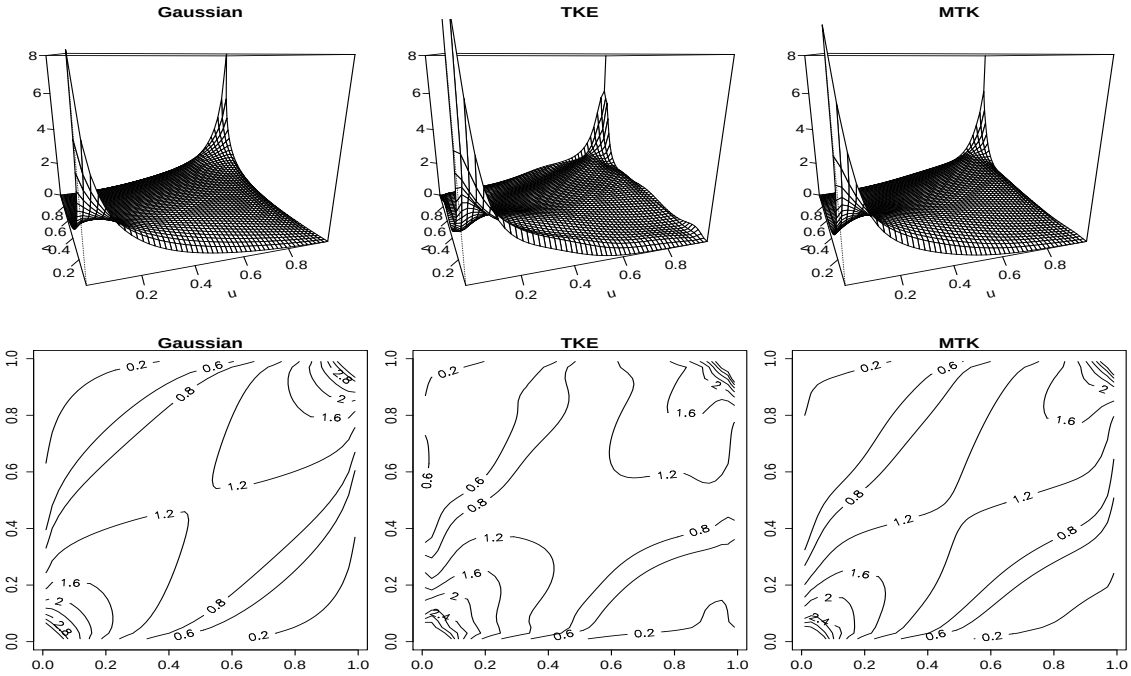


Figure 3.3: FTSE 100 and Hang Seng indexes

4. ESTIMATION OF SPATIALLY DEPENDENT CROP YIELD DISTRIBUTIONS AND CROP INSURANCE PREMIUM RATES: AN EMPIRICAL LIKELIHOOD KERNEL APPROACH

4.1 Introduction

Reliable calculation of crop insurance rates is essential for the US federal crop insurance program, which has been a vital part of agricultural industry. Its utmost importance is partly due to the policy consideration. For instance, the 2014 Farm Act allocates $89.8 billion federal budget for crop insurance programs over the next ten years. This massive resource is directed to help farmers pay their crop insurance premiums. Moreover, accurate crop insurance rates are necessary to avoid moral hazard or adverse selection, thus are fundamental for the health of crop insurance markets. Among a variety of crop insurance plans, area-yield insurance such as Group Risk Plan (GRP) has attracted significant attention. GRP is based on county-yield data provided by the National Agricultural Statistics Service (NASS) of the US Department of Agriculture. After participating in the plan, crop producers select the coverage levels based on a county's average yield and will collect indemnities if the realized average yield is lower than a pre-specified trigger level. GRP is designed to reduce the potential adverse selection behaviors because a single producer is considered incapable to manipulate a county's average yield. In this article, I restrict ourselves to GRP insurance rates estimation.

The past two decades have seen persistently increasing interest in modeling crop yield distribution and calculating crop insurance rates. Reliable estimation of yield distribution, or more relevantly its lower tail, is crucial in deriving accurate insurance rates. However, as highlighted by many studies, these objectives are challenging. The

main obstacle is essentially a data sparseness issue. In practice, crop yield data is featured by very short panel structure: the number of counties is numerous, but given a county of interest, the historical yield data is typically limited, only covering around fifty years or less. Most past studies evaluated insurance rates county by county separately, therefore the panel structure was seldom taken into consideration. In reality, although the yield data among nearby counties are governed by different distributions, they exhibit strong correlation and share certain similarities. These facts may provide additional information that is beneficial to estimating the yield distribution for a particular county of interest. Some attempts to exploit the panel structure of yield data have been made in the literature, for example, Ker (1996); Ker and Goodwin (2000); Ozaki et al. (2008).

This study aims to propose a new method that pools information effectively under mild assumptions and estimates yield distributions as well as insurance rates across counties in a given area, e.g., a state. It is recognized that the yield distributions of nearby counties share some stylized features. If characterized properly, borrowing information from neighbors improves statistical efficiency. To proceed, some reasonable presuppositions are necessary for a sound pooling scheme. It is often restrictive to make assumptions directly on the yield distribution functions of nearby counties; nevertheless, it may be reasonable to assume their certain moments, which summarize the feature of the yield distributions concisely, resemble each other. It's well known that given a set of similar moment conditions, the underlying distributions still allow great flexibility in their exact functions. Under this premise, I uncover the required moments of a county's yield distribution by smoothing the data from all its neighbor counties with carefully selected weight functions. Moreover, I take spatial information into account when constructing these weight functions in order to achieve further efficiency gains. Then the uncovered moment conditions, with extra infor-

mation from the neighbor counties incorporated, are imposed as constraints when estimating the yield distribution and calculating the insurance rates for a particular county of interest.

Specifically, the empirical likelihood kernel density estimator (ELK) is employed to estimate yield distributions for my purpose. ELK is specially designed for the situation where extra distributional information is available. The standard kernel density estimator (KDE), which has already been introduced in the crop insurance literature, fails to do so because it attaches the equal probability weight to each data point. ELK simply replaces those equal probability weights by some unequal counterparts obtained by maximizing empirical likelihood subject to a set of moment conditions. I present more details regarding this in the following sections. After estimating the yield distributions based on ELK, the insurance rates are evaluated in the usual way.

The proposed method that combines spatially smoothed moment conditions and ELK has several desired advantages. First, like other nonparametric methods, ELK does not require any parametric assumption for the underlying yield distribution; thus it is flexible and free of misspecification error. Second, the performance of KDE is hampered by small sample size, resulting in limited applications in the literature. In contrast, ELK is able to decrease the variance significantly in small samples. Third, with possibly erratic bumps, KDE is known to behave poorly at tails. This difficulty aggravates the estimation of insurance rates because it only addresses the left tail of the yield distribution. In contrast, the imposed moment conditions in the ELK are found to have the tails regularized, resulting in more reliable tail estimates. Fourth, the proposed method effectively pools information from neighbor counties under less restrictive assumptions, thus the estimated insurance rates shall be more accurate. Fifth, the proposed method is intuitively simple and easy to implement.

Then, I apply the proposed method to Iowa corn. Iowa was chosen because it is the largest corn planting state in the US. I estimate the yield distributions and insurance rates for the ninety-nine Iowa counties. My results have demonstrated that large differences can arise solely from the proposed information pooling method. Furthermore, I conduct an empirical simulation study and assess the performance of each competing estimator in estimating yield distributions and insurance rates. It suggests that, in general, the proposed method is substantially better than its counterpart without information pooling. Hereby, this has demonstrated the soundness of the proposed method for applications in practice.

The rest of text is organized as follows. In section 4.2, I outline the procedures for estimating yield distributions and insurance rates; some existing approaches are briefly introduced. Section 4.3 describes the empirical likelihood kernel density estimator. In Section 4.4, the construction of the four spatially smoothed moment conditions and a step-by-step description of the proposed method are presented. I report the empirical simulation in Section 4.5, followed by an application to Iowa corn in Section 4.6.

4.2 Literature

Typically, researchers consider a two-step procedure to model yield distribution and calculate insurance rates. In the first step, it is necessary to remove the trend effect as crop yield is inclined to trend upwards due to technological advancement and yield distribution should solely represent random factors. Various detrending models have been used in the literature and they generally fall into two categories: stochastic and deterministic approaches. Denote $Y_t : t = 1, \cdots, T$ to the yield series of a county of interest. The stochastic approach considers fitting an autoregressive integrated moving average model to the yield series Y_t, for example, see Goodwin

and Ker (1998); Ker and Goodwin (2000). The main argument for the stochastic approach may be best explained in (Goodwin and Ker 1998, p.143): "because drought or excessive moisture effects may persist from year to year, it is important that any autoregressive or moving average effects also be recognized". Nevertheless, Harri et al. (2009) found limited empirical evidence to support the stochastic approach. Therefore, the recent literature opted to the deterministic approach instead. Just and Weninger (1999) regressed Y_t against a polynomial of time t up to the fifth order. Sequential t-tests were used to select the polynomial order. Harri et al. (2011) used a two-knot linear spline regression. The coefficients and the two knots are treated as unknown parameters and estimated by outliers robust M-estimation with Huber function and bisquare function. The nonparametric kernel regression was also employed in the literature, where the trend is treated as an unknown smooth function of the time t, i.e.

$$(4.1)\quad Y_t = m(t) + e_t.$$

Local linear estimator is a common choice for kernel regression, and it is used to estimate $m(t)$ in Claassen and Just (2011). For a general introduction of the local linear estimator, see Fan and Gijbels (1996). This model is unrestrictive as it does not specify any functional form for detrending. In fact, erroneously imposing linear or nonlinear trend function may lead to false conclusions in modeling yield distribution (Just and Weninger 1999; Atwood, Shaik, and Watts 2003). Therefore, I use (4.1) to remove the trend effect in this article.

Many previous studies reported the obtained residuals $e_t : t = 1, \cdots, T$ have violated the homoscedasticity assumption required in the detrending model. The existing literature commonly relates the source of heteroscedasticity to the fitted yield trend $\hat{Y}_t$. Two primary heteroscedasticity assumptions have been extensively used in

yield distribution modeling: one is the simple homoscedasticity assumption (Mahul 1999; Coble, Heifner, and Zuniga 2000), the other is the constant coefficient of variation assumption (Miranda and Glauber 1997; Skees, Black, and Barnett 1997; Ker and Coble 2003), which asserts that the standard deviation of e_t varies proportionally to the changes of fitted yield trend, namely

$$\mathbb{E}\left(e_t^2\right) = \sigma^2 \left\{\mathbb{E}\left(Y_t\right)\right\}^2 = \sigma^2 \hat{Y}_t^2.$$

Despite both the assumptions get support in the literature, neither of them holds universally across location or crop type. Harri et al. (2011) studied the effect of heteroscedasticity assumptions on the estimated area-yield insurance rates and their results have revealed significant differences under different assumptions. Therefore, the literature still remains ambiguity on this heteroscedasticity issue. In the empirical simulation of this study, I consider both the assumptions and demonstrate that the superior performance of the proposed method is robust to different heteroscedasticity assumptions.

For a county of interest, let $\mathcal{F}_T$ be the σ-algebra generated by the observed yield series $Y_1, \cdots, Y_T$, then $\mathcal{F}_T$ represents all the past information up to time T. In the second step, the yield distribution, i.e. the density of Y_{T+2} conditional on $\mathcal{F}_T$, needs to be estimated. This two-step ahead variable Y_{T+2} is typical in the literature because insurance rates are evaluated about six months before farmers purchase the insurance contracts and plant their crops, and moreover, yield data from the NASS has some time lag. In order for estimation, a sample is recovered by

$$\hat{Y}_{T+2} + \epsilon_t,\ t = 1, \cdots, T$$

where $\hat{Y}_{T+2}$ is a two-period ahead forecast from the detrending model and $\epsilon_t : t =$

$1, \cdots, T$ are adjusted residuals defined as

$$\epsilon_t = \begin{cases} e_t & \text{if homoscedasticity assumption;} \\ e_t \frac{\hat{Y}_{T+2}}{\hat{Y}_t} & \text{if constant of variation assumption.} \end{cases} \tag{4.2}$$

This sample is viewed i.i.d. from the yield distribution, which is estimated by some appropriate method. Then the insurance rates, at the percentage level, are calculated by

$$R_\theta = \mathbb{P}\left(Y_{T+2} < \theta \hat{Y}_{T+2}\right) \left\{ 1 - \frac{\mathbb{E}\left(Y_{T+2} | Y_{T+2} < \theta \hat{Y}_{T+2}\right)}{\theta \hat{Y}_{T+2}} \right\} \times 100\%, \tag{4.3}$$

where θ is the coverage level and $\theta \hat{Y}_{T+2}$ is the trigger yield. Similarly, this formula is evaluated conditionally on the past information $\mathcal{F}_T$.

Doubtlessly, the quality of yield distribution estimation will heavily affect the accuracy of insurance rates. However, this task is challenged by sparse yield data. To ensure statistical soundness, the majority of past studies used parametric approach. The literature generally reported the yield distribution is non-normal: skewness and excess kurtosis are observed. Therefore, many flexible distribution families have been proposed, including gamma distribution (Gallagher 1987), beta distribution (Nelson 1990), inverse hyperbolic sine transformation (Moss and Shonkwiler 1993), log-normal distribution (Stokes 2000), weibull distribution (Sherrick et al. 2004), and etc. In contrast, Just and Weninger (1999) pointed out that some methodology and data limitations may have inappropriately rejected the normality hypothesis, and they suggested calling the normal distribution back to attention. One limitation of the parametric approach is that the distribution specification may be incorrect, resulting in misleading insurance rates. Sherrick et al. (2004) studied various distribution specifications and concluded that unexamined specification may lead to

significant errors in crop insurance policy. Some efforts to empirically rank the different distribution specifications have been made, see Norwood, Roberts, and Lusk (2004); Ramírez and McDonald (2006).

In contrast, the nonparametric approach does not require any distribution specification. Goodwin and Ker (1998) introduced the standard kernel density estimator to model yield distribution and estimate insurance rates. Ker and Goodwin (2000) emphasized that lower tail probabilities are crucial in deriving insurance rates and used empirical Bayes kernel density estimator with adaptive bandwidth. Ker and Coble (2003) pursued a semiparametric way: they took normal and beta distributions as a prior guess for the yield distribution, and then used the nonparametric kernel method to correct them. In general, the nonparametric approach has a slower convergence rate than a parametric approach. Therefore, its performance may be more sensitive to small samples. To overcome this potential disadvantage, exploring the panel nature of yield data in the nonparametric context is promising. In Goodwin and Ker (1998), residuals from neighbor counties were weighted and then pooled together to estimate the yield distribution directly. Though simple, this strategy makes rather restrictive assumptions on the yield distributions from neighbor counties. Ker and Goodwin (2000) assumed a baseline yield distribution across counties by a hierarchical model. Given a fixed support point, function values from the different yield distributions deviate from the baseline according to a normal distribution. This method requires the yield data from different counties to be independent. Thus, in their framework, pooling is done among remote counties.

4.3 Empirical likelihood kernel density estimation

I make a detour to introducing empirical likelihood kernel density estimator (ELK) in this section. Assume, in a general sense, I have an i.i.d. sample $X_1, \cdots, X_n$

from a univariate distribution with density function f. The standard kernel density estimator (KDE) gives, at a support point x,

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K_h\left(x - X_i\right),$$

where $K_h(\cdot) = K(\cdot/h)/h$. The kernel function $K(\cdot)$ is taken to be some symmetric and uni-modal density function defined on $\mathbb{R}$ or some finite interval. Though Epanechnikov kernel is optimal, the choice of kernel function is of little importance regarding the general performance of KDE. In practice, the density function of the standard normal distribution is a common choice. The bandwidth h controls the amount of smoothing, and is the most crucial decision one needs make. In KDE, h is global and fixed, indicating equal smoothing over all the data points. See Wand and Jones (1995) for a general treatment of kernel type density estimation.

KDE is essentially a weighted average of individual kernels at a given support point; see Ker and Goodwin (2000) for a graphical illustration of this. Specifically, for each data point X_i, two things are attached. One is the kernel $K_h\left(x - X_i\right)$ that centers at X_i and the other is the associated weight $1/n$. KDE applies the equal weight to all kernels, thus can not represent any additional information about the underlying distribution f. Motivated by this, Chen (1997) introduced ELK in a general setup, i.e.

$$\hat{f}_{el}(x) = \sum_{i=1}^{n} w_i K_h\left(x - X_i\right), \tag{4.4}$$

where w_i is the weight attached to X_i. The unequal weights w_i's are determined by maximizing empirical likelihood subject to a set of constraints that reveals the extra distributional knowledge.

Empirical likelihood (EL), originally proposed by Owen (1988, 1990), has been a popular and powerful tool in statistics and econometrics literature. For a complete

introduction, see Owen (2001). EL nonparametrically chooses the probability weight w_i for the ith data point X_i according to some constraints. Assume I have the following q known moment conditions about f:

$$\text{(4.5)}\quad \mathbb{E}\left\{g_l(X_i)\right\} = b_l, \quad l = 1, ...q,$$

where $g_l(X_i)$ is some well-defined function of the random variable X_i, and b_l is the corresponding moment value. Then the EL probability weights can be obtained by the following maximization problem

$$\text{(4.6)}\quad \begin{aligned} \underset{(w_1,...,w_n)}{\text{maximize}} \quad & \prod_{i=1}^{n} nw_i \\ \text{subject to} \quad & \sum_{i=1}^{n} w_i g_l(X_i) = b_l, \quad l = 1, ..., q, \\ & 0 \le w_i \le 1, \quad i = 1, ..., n, \\ & \sum_{i=1}^{n} w_i = 1. \end{aligned}$$

Note that the moment conditions (4.5) are replaced by their sample analogs in the above maximization problem. The objective function $\prod_{i=1}^{n} nw_i$ in (4.6) is called empirical likelihood. EL can be seen as a nonparametric version of likelihood ratio, which we may have been familiar with in a parametric situation. Since w_i is the probability associated with the observation X_i under those constraints, the likelihood of observing $X_1, X_2, ..., X_n$ is $\prod_{i=1}^{n} w_i$. Owen (2001) has pointed out that, if there is no constraint, the maximum likelihood is $(1/n)^n$, i.e. equal probability mass $1/n$ is placed at each data point. Then the likelihood ratio is $\prod_{i=1}^{n} w_i/(1/n)^n$, which is exactly the formula of EL.

In order for the notational conciseness, let $J(X_i) = \left[g_1(X_i) - b_1, ..., g_q(X_i) - b_q\right]^\top$ be a q-dimensional vector of functions constructed from the moment conditions. Then $\sum_{i=1}^{n} J(X_i) = 0$ is equivalent to the first set of constraints in (4.6). Also let

their corresponding Lagrangian multipliers be $\lambda = [\lambda_1, ..., \lambda_q]^\top$. Then the solution of (4.6) is characterized by, for $i = 1, ..., n$,

$$(4.7)\quad w_i = n^{-1}\left\{1 + \lambda^\top J(X_i)\right\}^{-1},$$

and λ is the solution of

$$(4.8)\quad \sum_{i=1}^{n} \frac{J_l(X_i)}{1 + \lambda^\top J(X_i)} = 0.$$

These optimal w_i's are then plugged back into (4.4) for ELK.

We may consider KDE as a special case of ELK. If no additional prior knowledge is available or only the in-sample information is used, ELK reduces to KDE. In the case where we do not have those moment conditions in (4.5), the solution of (4.6) will always be $w_i = 1/n, i = 1, ..., n$, as said above. Also if we only use the in-sample information such as the corresponding sample counterparts, i.e., let $b_l = \frac{1}{n}\sum_{i=1}^{n} g_l(X_i)$ in (4.6), one may easily check that (4.7) together with (4.8) give $w_i = 1/n$ too. Therefore, prior or out-of-sample information distinguishes ELK from KDE.

The asymptotic properties of ELK are introduced in the next, compared to those of KDE. Chen (1997) has established its bias and variance as

$$(4.9)\quad \text{bias}\left\{\hat{f}_{el}(x)\right\} = \text{bias}\left\{\hat{f}(x)\right\} + o(n^{-1}),$$

and

$$(4.10)\quad \text{var}\left\{\hat{f}_{el}(x)\right\} = \text{var}\left\{\hat{f}(x)\right\} - J(x)^\top \Sigma^{-1} J(x) f^2(x) n^{-1} + o(n^{-1}),$$

where $\Sigma = \Big(\text{cov}\left(J_l(X_i), J_m(X_i)\right)\Big)$ is a covariance matrix of $J(X_i)$. It is obvious that there is only $o(n^{-1})$ difference between KDE and ELK in terms of bias. If considering mean integrated squared error (MISE), this difference is negligible. Although the

dominant terms in the variance of KDE and ELK are the same, we do see an $O(n^{-1})$ reduction in ELK case since the coefficient of n^{-1} in (4.10) is negative. This confirms the general belief that empirical likelihood decreases an estimator's variance. As a result of smoothing, this reduction only occurs in the small order term. Howerver, as Chen (1997) pointed out, the extent of the reduction can be subtantial when sample size is small or medium. Combining (4.9) and (4.10), it yields

$$\text{MISE}\left\{\hat{f}_{el}(x)\right\} = \text{MISE}\left\{\hat{f}(x)\right\} - J(x)^{\top}\Sigma^{-1}J(x)f^2(x)n^{-1} + o(n^{-1}).$$

This suggests that a reduction in MISE is present by using extra information as well. Here, I do not write down the explicit expressions for the asymptotic bias, variance and MISE of KDE. For more details, one may refer to Wand and Jones (1995).

The bandwidth selection is known to be crucial in kernel type density estimation problems. Since EL reduces MISE at the smaller order $O(n^{-1})$, asymptotically speaking, ELK does not require a distinct rule for the optimal h. For simplicity, we may replicate any bandwidth selection method for KDE directly to the ELK situation. There are a variety methods available in the literature, for example, rule of thumb, cross validation, Sheather and Jones' plug in method, and etc. In this study, the sample size is usually small. Therefore, following Goodwin and Ker (1998), I feel Silverman's rule of thumb method seems to be a reasonable choice towards robust results. Thus, I set the optimal bandwidth of ELK to be

$$h = 0.9 \cdot \min\left\{\hat{\sigma}, \text{IQR}/1.34\right\} \cdot n^{-1/5},$$

where $\hat{\sigma}$ is the sample standard deviation and IQR stands for interquantile range.

Ker and Goodwin (2000) argued that the fixed bandwidth in KDE may be sometimes problematic in rating crop insurance because it often yields too much spurious bumps in the tails. For a type of long-tailed densities, this kind of undersmoothing

in the tails becomes particularly serious. Therefore, they recommended the adaptive kernel density estimation (AKDE) featured by local bandwidth, namely

$$\check{f}(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h\nu_i} K\left(\frac{x - X_i}{h\nu_i}\right).$$

From the Abramson rule, the optimal ν_i is equal to $f(X_i)^{-1/2}$. Thus, AKDE requires a pilot estimate of $f(X_i)$, and KDE may serve this purpose well. See Silverman (1986) or Ker and Goodwin (2000) for a detailed step-by-step procedure. Compared to KDE, AKDE typically has faster convergence rate and has more complex global properties. To my best knowledge, there is not an empirical likelihood version of AKDE in the literature. Still, I can construct one analogously. I call it empirical likelihood adaptive kernel density estimation (ELAK) in the following, namely,

$$\check{f}_{el}(x) = \sum_{i=1}^{n} \frac{w_i}{h\nu_i} K\left(\frac{x - X_i}{h\nu_i}\right).$$

Similarly, the w_i's therein are obtained from (4.6). The properties of this ELAK remain unclear. It is generally believed that empirical likelihood helps reduce an estimator's variance. Thus, I may reasonably expect that an $O(n^{-1})$ variance reduction occurs in ELAK as well, compared to AKDE. Because this reduction is still of smaller order, I generally follow the procedures used in AKDE. In this study, I also take AKDE and ELAK into consideration for completeness.

4.4 Spatially smoothed moment conditions

A set of moment conditions is required when implementing empirical likelihood, for this purpose, I describe the selected four moments and propose a spatial smoothing procedure to uncover their values. To proceed, the panel structure of yield data is incorporated and some notations are introduced as follows,

(1) $\mathcal{N}$ consists of all counties of interest and let i and j denote two distinct counties

therein; in total, $\mathcal{N}$ includes N elements;

(2) $Y_{i,t} : t = 1, \cdots, T$ are a sequence of random variables denoting the observed yield series from county i;

(3) $\epsilon_{i,t} : t = 1, \cdots, T$ are the corresponding adjusted residuals for county i; they are extracted from the detrending model (4.1) and then adjusted according to (4.2) under an appropriate heteroscedasticity assumption;

(4) $\hat{Y}_{i,T+2}$ is a two-period ahead forecast based on the detrending model for county i; it is an estimate of the expected yield $\mathbb{E}\left(Y_{i,T+2}\right)$;

(5) f_i is the yield distribution of county i; the sample $\hat{Y}_{i,T+2} + \epsilon_{i,t} : t = 1, \cdots, T$ is constructed for its estimation;

(6) f_i^{ϵ} is the distribution of $\epsilon_{i,t}$; it can be estimated from the sample $\epsilon_{i,t} : t = 1, \cdots, T$;

(7) d_{ij} is the Euclidean distance between counties i and j calculated from the longitude and latitude data;

(8) $\mathcal{N}_i^k$ is a set including county i and its k nearest neighbors based on the above distance metric;

(9) $\mathcal{X}_i^k$ is a set of adjusted residuals for all counties in $\mathcal{N}_i^k$,
i.e. $\left\{\epsilon_{j,t} : j \in \mathcal{N}_i^k, t = 1, \cdots, T\right\}$.

It is argued that only the adjusted residuals $\epsilon_{i,t}$ that purely capture the random effects share stylized similarities among nearby counties, while the fixed trend effects $\hat{Y}_{i,T+2}$ may be quite different. Therefore, instead of targeting at f_i directly, the ELK or ELAK is used to estimate f_i^{ϵ} as the first step in the proposed method. Then in

the second step, the desired yield distribution is recovered by the following change of variable formula, i.e.

$$f_i(y) = f_i^{\epsilon}\left(y - \hat{Y}_{i,T+2}\right). \tag{4.11}$$

If $\mathbb{E}(\epsilon_{i,t}^2) < \infty$, the four moment conditions, parallel to the form in (4.5), are specified for each county i as

$$\mathbb{E}\boldsymbol{g}\left(\epsilon_{i,t}\right) = \mathbb{E}\begin{bmatrix} g_1\left(\epsilon_{i,t}\right) \\ g_2\left(\epsilon_{i,t}\right) \\ g_3\left(\epsilon_{i,t}\right) \\ g_4\left(\epsilon_{i,t}\right) \end{bmatrix} = \mathbb{E}\begin{bmatrix} \epsilon_{i,t} \\ \epsilon_{i,t}^2 \\ \left(\epsilon_{i,t} - \kappa_{1,i}\right)_+^2 \\ \left(\epsilon_{i,t} - \kappa_{2,i}\right)_+^2 \end{bmatrix} = \begin{bmatrix} b_{1,i} \\ b_{2,i} \\ b_{3,i} \\ b_{4,i} \end{bmatrix} = \boldsymbol{b}_i, \tag{4.12}$$

where $(x)_+ = x$ if $x > 0$, otherwise $(x)_+ = 0$. The two knots κ_{1i} and κ_{2i} are defined as the sample quantiles of $\mathcal{X}_i^k$ with probabilities 0.33 and 0.67, respectively. Following the idea of spline, these moment conditions are piecewise defined. This construction embraces more information than just the first two moments do, at the same time, it avoids the difficulties of modeling higher moments, e.g., more restrictive assumptions and numerical unstableness.

The four moment values $\boldsymbol{b}_i$ need to be uncovered, based on the assumption that, if two counties i and j are close, the specified moments $\boldsymbol{b}_i$ and $\boldsymbol{b}_j$ shall resemble each other. Under this premise, the moment values for a particular county i can be estimated by spatially smoothing over all its neighbors. Specifically, $\boldsymbol{b}_i$ is estimated by

$$\boldsymbol{b}_i = \frac{1}{T}\sum_{t=1}^{T}\sum_{j\in\mathcal{N}}\boldsymbol{g}(\epsilon_{jt})\omega_{ij}. \tag{4.13}$$

This is essentially a weighted average procedure across all counties and all time periods. The weight ω_{ij} is invariant across time and is solely determined by the

distance from the targeted county i to other counties, and it admits the form

$$\omega_{ij} = \frac{\exp\left(-d_{ij}^2/\tau_i^2\right) \mathbb{1}\{j \in \mathcal{N}_i^k\}}{\sum_{j=1}^N \exp\left(-d_{ij}^2/\tau_i^2\right) \mathbb{1}\{j \in \mathcal{N}_i^k\}}, \tag{4.14}$$

where $\mathbb{1}\{a\} = 1$ if a is true, otherwise $\mathbb{1}\{a\} = 0$. It is easy to check that $\sum_{j=1}^N \omega_{ij} = 1$ for any county i. Particularly, as the distance d_{ij} increases, the weight ω_{ij} becomes smaller; moreover, if d_{ij} exceeds some threshold, the corresponding weight is set to zero. This construction is based on the perception that, the closer a county is to the targeted county i, the greater extent they resemble each other; if these two counties are far away enough, we may view them mutually independent. Note that I use the Gaussian type function $\exp\left(-d_{ij}^2/\tau_i^2\right)$ to conduct the smoothing; one can use the exponential type function $\exp\left(-d_{ij}/\tau_i\right)$ as well. The Gaussian type function allocates more weights to the several closest neighbors than the exponential type function does. Nevertheless, in my simulation experiments, I observe no significant difference between these two options. The parameter τ_i serves like the bandwidth and it controls the amount of smoothing. I employ the k-nearest neighbor idea and relate τ_i to the parameter k in the following way

$$\tau_i = \frac{1}{k} \sum_{j \in \mathcal{N}_i^k} d_{ij}, \tag{4.15}$$

which is exactly the average distance between county i and its k nearest neighbors. It is noticed that the parameter k plays a crucial role in my proposed method and I set k to the integer that is the closest to $\sqrt{N}$. Though arbitrary, this selection rule generally yields good performance in practice.

The procedure of my proposed method is briefly summarized here. For each county i, the following steps are replicated, *a*) I remove the trend effect and obtain the corresponding two-period forecast individually; *b*) I adjust the extracted residuals

based on some reasonable heteroscedasticity assumption; *c*) I follow the four moment conditions in (4.12) and uncover their values by the above spatial smoothing procedure; by construction, these moment values contain information from neighbor counties; *d*) I use ELK or ELAK to estimate f_i^ϵ based on the four moment conditions; *e*) the desired yield distribution f_i is recovered and the insurance rates are calculated according to (4.3).

4.5 Empirical simulation

In this section, I investigate the performance of the proposed method in practice. Two pairs of competing estimators, namely KDE and ELK, as well as AKDE and ELAK, are used to estimate the yield distributions, respectively, and I examine if the two empirical likelihood based estimators that pool information among neighbor counties improve the accuracy of the estimated insurance rates.

Following the idea in Ker and Ergun (2005); Ker and Goodwin (2000), I conduct an empirical simulation in the context of Iowa corn. The yield data from 1957 to 2010, i.e. $T = 54$, for all $N = 99$ counties in Iowa is available from the NASS. I remove the trend effects and get the two-period ahead forecast $\hat{Y}_{i,T+2}$ for each county $i \in \{1, \cdots, N\}$. The extracted residuals are adjusted according to both the homoscedasticity assumption and subsequently the constant coefficient of variation assumption. Then, for each county i, I obtain a pilot estimate of f_i^ϵ by assuming that $\epsilon_{i,t} : t = 1, \cdots, T$ follow the skew normal distribution. Denote each pilot estimate by $f_i^{\epsilon,p}$ and it admits the form

$$f_i^{\epsilon,p}(y) = \frac{2}{\sigma_i}\phi\left(\frac{y-\mu_i}{\sigma_i}\right)\Phi\left(\alpha_i\left(\frac{y-\mu_i}{\sigma_i}\right)\right),$$

where ϕ and Φ are the pdf and cdf of standard normal distribution, moreover μ_i, σ_i and α_i are location, scale and skewness parameters, respectively. The skew normal

distribution is used for two reasons: first, it has a regular and smooth parametric form; second, it allows skewness which is often highlighted in the literature. Then, each pilot estimate $f_i^{\epsilon,p}$ is treated as the true f_i^{ϵ} and used for the empirical simulation.

The proposed method has assumed that the distributions of the adjusted residuals among neighbor counties resemble each other, moreover the extent of this resemblance is negatively related to their distances. In reality, I attempt to justify these assumptions empirically by examining whether the pilot estimates $f_i^{\epsilon,p} : i \in \{1, \cdots, N\}$ exhibit such spatial structures. Specifically, I proceed as follows.

(A) Denote h_{ij} to the Hellinger distance between $f_i^{\epsilon,p}$ and $f_j^{\epsilon,p}$, where i and j are two distinct counties. Consider a simple regression

$$h_{ij} = \beta_0 + \beta_1 d_{ij} + \eta_{ij}. \tag{4.16}$$

Then, I should expect β_1 positive and statistically significant.

(B) Let $\boldsymbol{\mu} = [\mu_1, \cdots, \mu_N]^\top$, $\boldsymbol{\sigma} = [\sigma_1, \cdots, \sigma_N]^\top$ and $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_N]^\top$ be three vectors of the estimated parameters. Consider fitting a spatial error model to $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\alpha}$, respectively. Take $\boldsymbol{\mu}$ as an illustrating example, the spatial error model is represented as

$$\begin{aligned} \boldsymbol{\mu} &= m + \boldsymbol{\zeta} \\ \boldsymbol{\zeta} &= \lambda \mathbf{W} \boldsymbol{\zeta} + \boldsymbol{\xi}, \end{aligned} \tag{4.17}$$

where *a)* m is the intercept and can be interpreted as the baseline level; *b)* λ is the spatial error coefficient; if $\lambda \neq 0$, then spatial effect is present; *c)* $\mathbf{W}$ is a $N \times N$ spatial weight matrix; *d)* $\boldsymbol{\xi}$ is a N-dimensional vector of random errors and $\boldsymbol{\xi} \sim \mathcal{N}\left(\mathbf{0}, \sigma_\xi^2 \mathbf{I}_N\right)$. The spatial weight matrix $\mathbf{W}$ is constructed based on the distance of each pair of counties, i.e. 1) let $\mathbf{W}^* = \left[w_{ij}^*\right]$, where $w_{ij} = 1/d_{ij}$ if $i \neq j$, otherwise $w_{ij} = 0$; 2) row normalization, i.e. $\mathbf{W} = \left[w_{ij}\right]$, where

$w_{ij} = w^*_{ij} / \sum_{j=1}^{99} w^*_{ij}$. The spatial error model has been commonly used in the literature. I should expect all $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\alpha}$ possess spatial structures.

Table 4.1 briefly summarizes the estimated coefficients in the above models. The results are as one would expect. The estimated $\hat{\beta}_1$'s are all positive and statistically significant, indicating the distributions of nearby counties are more resembled than those of distant counties. The estimated $\hat{\lambda}$'s are generally around 0.9, thus strong spatial structures are embedded in $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\alpha}$. Therefore, the assumptions used in the proposed method should be valid.

Table 4.1: Estimation results of the model (4.16) and model (4.17)

	$\hat{\beta}_0$	$\hat{\beta}_1$	$\boldsymbol{\mu}$		$\boldsymbol{\sigma}$		$\boldsymbol{\alpha}$	
			$\hat{m}$	$\hat{\lambda}$	$\hat{m}$	$\hat{\lambda}$	$\hat{m}$	$\hat{\lambda}$
homoscedasticity	0.0857*	0.0187*	-3.6363	0.9536*	17.4935*	0.9543*	-0.7585*	0.8719*
	(0.0021)	(0.0008)	(3.1397)	(0.0461)	(4.5741)	(0.0457)	(0.1320)	(0.1246)
const. coef. of variation	0.0892*	0.0144*	-5.2234	0.9299*	26.3672*	0.9427*	-0.7491*	0.8772*
	(0.0019)	(0.0007)	(3.0826)	(0.0692)	(5.0547)	(0.0572)	(0.1236)	(0.1188)

The estimated standard errors are marked in parentheses; * denotes the significance level < 0.001.

The empirical simulation is managed as follows.

(1) I generate $\{\epsilon_{i,t} : t = 1, \cdots, T_0\}$ for each county $i \in \{1, \cdots, N\}$ independently, then combine them into the panel form $\{\epsilon_{i,t} : i = 1, \cdots, N; t = 1, \cdots, T_0\}$ for information pooling. Consider sample size $T_0 = 30$ and subsequently $T_0 = 50$ in order to highlight the small sample case.

(2) Based on the above simulated residuals and the two-period ahead forecast $\hat{Y}_{i,T+2}$, the four competing estimators, namely KDE, ELK, AKDE and ELAK,

are used to obtain an estimated yield distribution $\hat{f}_i^p$ and three estimated insurance rates $\hat{R}_{\theta,i}^P$ with coverage levels $\theta = 70\%$, 80% and 90%, respectively.

(3) I compare the performances of the four estimators in estimating yield distributions, based on the average MISE across N counties, i.e. $\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\int\left(\hat{f}_i^p - f_i^p\right)^2$, where f_i^p is the pilot yield distribution recovered from $f_i^{\epsilon,p}$ by the formula (4.11). The integral is evaluated numerically and the expectation is approximated by averaging 500 simulated samples.

(4) Finally, I consider the performances in estimating insurance rates, based on the average MSE, i.e. $\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(\hat{R}_{\theta,i}^p - R_{\theta,i}^p\right)^2$, where $R_{\theta,i}^p$ are pilot insurance rates derived based on f_i^p. The expectation is approximated analogously.

The empirical simulation results are described in Table 4.2. For ease of interpretation, I take KDE as the benchmark and all the reported numbers are in relative scale. Also for reference, the exact values for the KDE are shown in *italic* and parenthesis. Some conclusions are summarized here, which are robust to different sample sizes as well as different heteroscedasticity assumptions. First, the proposed information pooling method is confirmed to be successful. ELK and ELAK have substantially improved upon their counterparts, namely KDE and AKDE. Second, the performance of ELAK in estimating yield distributions is the best, while the performance of ELK is comparable. This is understandable because adaptive kernel method has faster convergence rate in theory. Third, in estimating insurance rates, ELK dramatically outperforms when the coverage levels are 70% and 80%, while ELAK dominates when the coverage level is 90%. Interestingly, it seems that ELK has better performance in the left tail of yield distributions; in contrast, ELAK excels in the global performance. To summarize, these results have demonstrated the potential large gains in estimating yield distributions and insurance rates based on

Table 4.2: Empirical simulation results

		KDE	ELK	AKDE	ELAK
Panel 1: $T_0 = 30$					
homoscedasticity	yield distribution	*(0.0010)*	0.7911	1.1904	**0.7603**
	$\theta = 70\%$	*(0.0875)*	**0.2240**	1.3672	0.5069
	$\theta = 80\%$	*(0.3652)*	**0.3443**	1.0305	0.4788
	$\theta = 90\%$	*(1.1245)*	0.4671	0.9406	**0.4371**
const. coef. of variation	yield distribution	*(0.0007)*	0.7926	1.1862	**0.7609**
	$\theta = 70\%$	*(0.8132)*	**0.3432**	1.0042	0.5027
	$\theta = 80\%$	*(1.8191)*	**0.4612**	0.9250	0.4831
	$\theta = 90\%$	*(3.1537)*	0.5000	0.9577	**0.4543**
Panel 2: $T_0 = 50$					
homoscedasticity	yield distribution	*(0.0006)*	0.8752	1.1680	**0.8136**
	$\theta = 70\%$	*(0.0544)*	**0.2148**	1.4537	0.6247
	$\theta = 80\%$	*(0.2277)*	**0.3804**	1.0370	0.5797
	$\theta = 90\%$	*(0.7024)*	0.5705	0.9149	**0.5394**
const. coef. of variation	yield distribution	*(0.0004)*	0.8914	1.1664	**0.8220**
	$\theta = 70\%$	*(0.5089)*	**0.3779**	0.9855	0.6115
	$\theta = 80\%$	*(1.1400)*	**0.5516**	0.8877	0.5932
	$\theta = 90\%$	*(1.9637)*	0.6400	0.9225	**0.5853**

the proposed method.

4.6 Example: Iowa corn

I estimate the yield distributions and derive the 2012 corn insurance rates for the ninety-nine counties in Iowa. To be consistent with the empirical simulation, the same dataset is used here. Similarly, I remove the trend effect for each county by model (4.1). In practice, local linear regression and the corresponding two-period ahead forecast can be easily implemented, for example, by the locfit package in R. I collect the residuals after detrending and perform a simple regression based heteroscedasticity test that is proposed and used in Harri et al. (2011). My results show that in 42 counties the homoscedasticity assumption has been rejected at the significant level 5%; for the rest 57 counties, I fail to do so. In contrast, the constant coefficient of variation assumption has been rejected in only 5 counties at the 5% significant level. Based on this observation, I adopt the constant coefficient of variation assumption in my application, though it is difficult to be validated. Then, the residuals are adjusted accordingly for subsequent estimation of yield distributions and insurance rates.

For illustration, in Figure 4.1, I plot the estimated yield distributions from the four estimators for Adair County. First consider the difference between adaptive kernel estimates and standard kernel estimates. Clearly, both KDE and ELK suffer a bump in their left tails. This feature seems undesirable and does not reveal the truth. Instead, AKDE and ELAK provide more smooth tail estimates, which is consistent with my prior beliefs. This again confirmed the observation made in Ker and Goodwin (2000). Second, consider the difference induced by the proposed information pooling method. Obviously, compared to their counterparts, ELK and ELAK have reduced the likelihood around the mode but allocated more probability

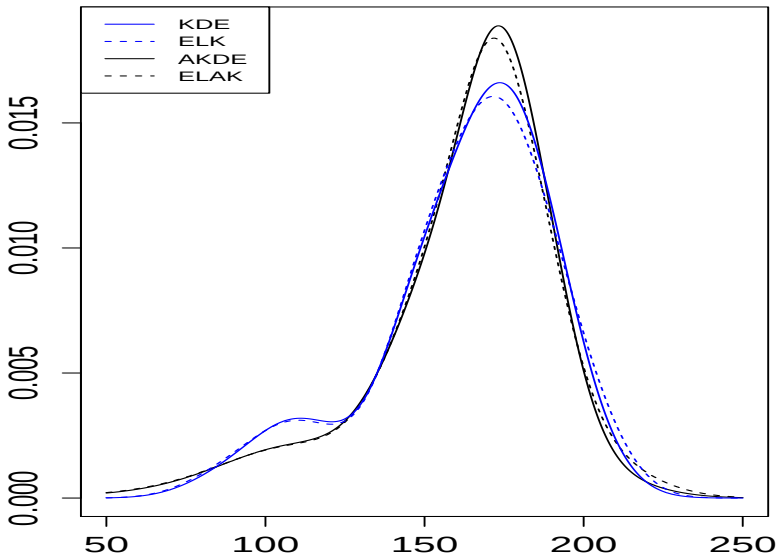


Figure 4.1: Yield distributions estimated by the four estimators for Adair County

mass along the tails, especially the right part. Admittedly, there is no way of knowing the true yield distribution here; nevertheless, I expect ELAK produce the most accurate estimates, according to the above empirical simulation.

To avoid lengthy report, only the coverage level 80% is considered for illustration. I provide some histograms in Figure 4.2 and the summary statistics in Table 4.3 to show how the estimated insurance rates are distributed across counties. The exact numbers for the estimated insurance rates are gathered in Appendix C. The following three major points are observed. First, on average, insurance rates by AKDE or ELAK are 6% higher than those by KDE or ELK; specifically, this is the case in 86 counties. This is understood because adaptive kernel estimators typically place larger amount of smoothing in the tails. Second, the proposed information pooling method may result in significantly different insurance rates from its counterpart

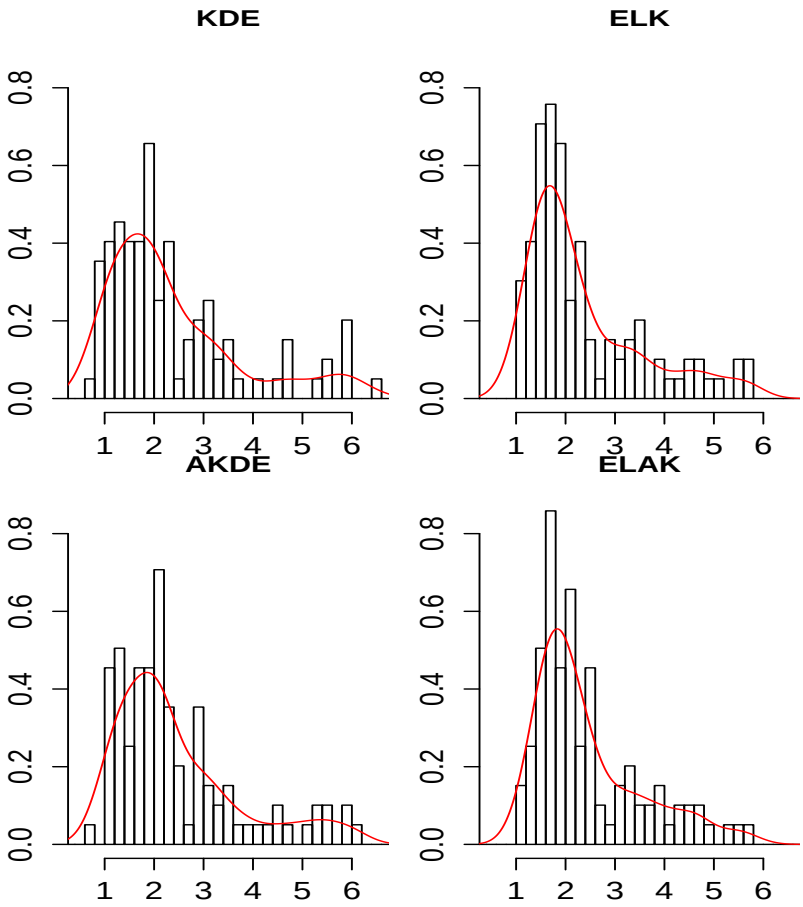


Figure 4.2: Histograms of the Iowa corn insurance rates at the coverage level 80%

Table 4.3: Summary statistics of the Iowa corn insurance rates at the coverage level 80% (all numbers are at percentage level)

	Min	1st Quantile	Median	Mean	3rd Quantile	Max
KDE	0.671	1.407	1.945	2.365	2.899	6.416
ELK	1.020	1.576	1.910	2.353	2.904	5.741
AKDE	0.799	1.599	2.082	2.449	2.957	6.104
ELAK	1.127	1.718	2.051	2.439	2.998	5.726

from a dot plot in Figure 4.3 which visually reveals the relative differences among the estimated insurance rates. For example, compared to KDE, ELK has increased the insurance rates in 50 counties with percentage increase ranging from 0.2% to 109.39%; in contrast, for the rest 49 counties, ELK has decreased the insurance rates instead with percentage decrease varying from -0.2% to -35.02%. Third, the insurance rates estimated by ELK or ELAK are more concentrated in a narrow interval than those estimated by KDE or AKDE, which is quite obvious from Figure 4.2 and Table 4.3. For example, the KDE insurance rates range from 0.671 to 6.416 and the corresponding histogram is relatively flat; in contrast, the ELK insurance rates merely range from 1.020 to 5.741 and the histogram exhibits a obviously sharp peak around 1.6. This reduction in variability is expected and again is the consequence of information pooling across neighbor counties. Since ELK is demonstrated to have the best performance in the above empirical simulation, I feel its estimated insurance rates shall be reasonably close to the true ones with a higher degree of confidence.

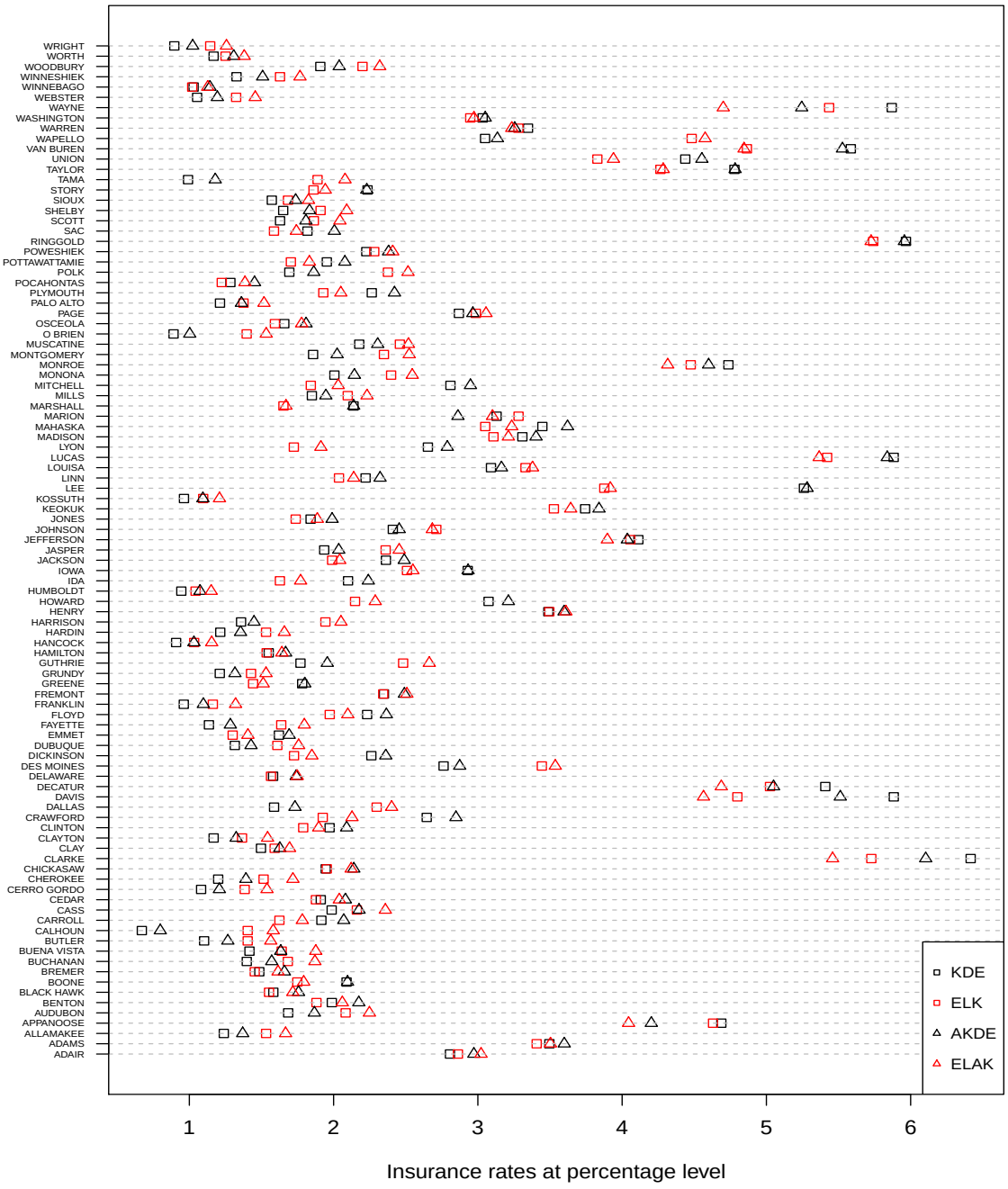


Figure 4.3: Dot plot of the Iowa corn insurance rates at the coverage level 80%

5. CONCLUSION

In the first essay, I have proposed a modified transformation-based kernel estimator for densities with bounded supports. The new estimator introduces a multiplicative factor while maintains the simplicity of transformation based kernel estimator with a single global bandwidth. I have established the theoretical properties of the proposed estimator and shown that it dominates the conventional transformation based estimator. I proposed three methods of bandwidth selection. My simulations demonstrate its good finite sample performance, especially for densities with poles on the boundaries. Extension of the proposed estimator to multivariate density estimations and regressions will be pursued in my future work. Another possibility, as suggested by a referee, is to develop an analog of SiZer for the modified transformation-based estimator, using the scale space ideas proposed by Chaudhuri and Marron (1999) and Hannig and Marron (2006).

Kernel type copula density estimation seems less developed in the literature because the standard kernel estimator suffers severe boundary biases. The transformation-based kernel estimator is a natural solution for boundary correction but may result in erratic estimates because of the unbounded multiplier associated with the back transformation. In the second essay, I propose the modified transformation-based kernel estimator that employs the tapering method to mitigate the consequences of the multiplier while maintains the simplicity of the fixed transformation and a single global bandwidth. I establish the theoretical properties of the proposed estimator and show it dominates the transformation-based kernel estimator. I further show that the proposed estimator enjoys higher order convergence rate under Gaussian copulas. Therefore, my estimator should provide outstanding performance for Gaus-

sian copulas and near Gaussian copulas which are practically relevant in financial data analyses. Extensions to non-diagonal bandwidth matrix are sensible and have been briefly discussed. I propose two methods to select the optimal smoothing parameters. My simulation results demonstrate its superior finite sample performance. I apply the proposed estimator to three real world datasets and it produces very smooth and well-behaved estimates. Consequently, the proposed estimator should be an appealing choice in practice.

Short yield series has limited the use of nonparametric approach to estimate yield distributions and insurance rates. An appropriate exploitation of the panel structure possessed by yield data is promising but not well developed in the literature. In the third essay, I therefore have proposed a new method in response to this demand. The proposed method begins with effectively incorporating information among neighbor counties to construct the spatially smoothed moment conditions. Then, these conditions are imposed as constraints when estimating yield distributions by well-established empirical likelihood kernel density estimator. The insurance rates are finally calculated in the usual way. The extension to empirical likelihood adaptive kernel estimator is also provided for completeness. Based on an empirical simulation, I have demonstrated the superior performance of the proposed method for small samples; moreover, the improvement in estimating insurance rates is substantial. If a nonparametric approach is necessary for flexibility reasons in practice, I feel the proposed method enables one to proceed with reliable results.

REFERENCES

Atwood, J., S. Shaik, and M. Watts. 2003. "Are Crop Yields Normally Distributed? A Reexamination." *American Journal of Agricultural Economics* 85:888–901.

Autin, F., E.L. Pennec, and K. Tribouley. 2010. "Thresholding Methods to Estimate Copula Density." *Journal of Multivariate Analysis* 101:200 – 222.

Bouezmarni, T., A. El Ghouch, and A. Taamouti. 2013. "Bernstein Estimator for Unbounded Copula Densities." *Statistics & Risk Modeling* 30:343–360.

Bouezmarni, T., J.V. Rombouts, and A. Taamouti. 2010. "Asymptotic Properties of the Bernstein Density Copula Estimator for α-Mixing Data." *Journal of Multivariate Analysis* 101:1–10.

Boyd, S., and L. Vandenberghe. 2004. *Convex Optimization*. Cambridge: Cambridge University Press.

Charpentier, A., J.D. Fermanian, and O. Scaillet. 2006. "The Estimation of Copulas: Theory and Practice." In J. Rank, ed. *Copulas, from Theory to Application in Finance*. London: Risk Books, pp. 35–64.

Chaudhuri, P., and J.S. Marron. 2000. "Scale Space View of Curve Estimation." *Annals of Statistics* 28:408–428.

—. 1999. "SiZer for Exploration of Structure in Curves." *Journal of the American Statistical Association* 94:807–823.

Chen, S.X. 1999. "Beta Kernel Estimators for Density Functions." *Computational Statistics and Data Analysis* 31:131–145.

—. 1997. "Empirical Likelihood-Based Kernel Density Estimation." *Australian Journal of Statistics* 39:47–56.

Chen, S.X., and T.M. Huang. 2007. "Nonparametric Estimation of Copula Functions

for Dependence Modelling." *The Canadian Journal of Statistics* 35:265–282.

Chen, X., Y. Fan, D. Pouzo, and Z. Ying. 2010. "Estimation and Model Selection of Semiparametric Multivariate Survival Functions under General Censorship." *Journal of Econometrics* 157:129 – 142.

Chen, X., Y. Fan, and V. Tsyrennikov. 2006. "Efficient Estimation of Semiparametric Multivariate Copula Models." *Journal of the American Statistical Association* 101:1228–1240.

Cheng, M.Y. 1997. "Boundary Aware Estimators of Integrated Density Derivative Products." *Journal of the Royal Statistical Society. Series B (Methodological)* 59:191–203.

Claassen, R., and R.E. Just. 2011. "Heterogeneity and Distributional Form of Farm-Level Yields." *American Journal of Agricultural Economics* 93:144–160.

Cline, D.B.H., and J.D. Hart. 1991. "Kernel Estimation of Densities with Discontinuities or Discontinuous Derivatives." *Statistics* 22:69–84.

Coble, K.H., R.G. Heifner, and M. Zuniga. 2000. "Implications of Crop Yield and Revenue Insurance for Producer Hedging." *Journal of Agricultural and Resource Economics* 25:539–551.

Cook, R.D., and M.E. Johnson. 1981. "A Family of Distributions for Modelling Non-Elliptically Symmetric Multivariate Data." *Journal of the Royal Statistical Society. Series B (Methodological)* 43:210–218.

—. 1986. "Generalized Burr-Pareto-Logistic Distributions with Applications to a Uranium Exploration Data Set." *Technometrics* 28:123–131.

Cowling, A., and P. Hall. 1996. "On Pseudodata Methods for Removing Boundary Effects in Kernel Density Estimation." *Journal of the Royal Statistical Society. Series B (Methodological)* 58:551–563.

Duong, T., and M. Hazelton. 2003. "Plug-in Bandwidth Matrices for Bivariate Kernel

Density Estimation." *Journal of Nonparametric Statistics* 15:17–30.

Duong, T., and M.L. Hazelton. 2005. "Convergence Rates for Unconstrained Bandwidth Matrix Selectors in Multivariate Kernel Density Estimation." *Journal of Multivariate Analysis* 93:417 – 433.

Fan, J., and I. Gijbels. 1996. *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall.

Fermanian, J.D. 2005. "Goodness-of-Fit Tests for Copulas." *Journal of Multivariate Analysis* 95:119–152.

—. 2012. "An Overview of the Goodness-of-Fit Test Problem for Copulas." In P. Jaworski, F. Durant, and W. Härdle, eds. *Copulae in Mathematical and Quantitative Finance*. Berlin: Springer, pp. 61–89.

Frees, E.W., and E.A. Valdez. 1998. "Understanding Relationships Using Copulas." *North American Actuarial Journal* 2:1–25.

Gallagher, P. 1987. "U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Disturbances." *American Journal of Agricultural Economics* 69:796–803.

Geenens, G. 2014. "Probit Transformation for Kernel Density Estimation on the Unit Interval." *Journal of the American Statistical Association* 109:346–358.

Geenens, G., A. Charpentier, and D. Paindaveine. 2014. "Probit Transformation for Nonparametric Kernel Estimation of the Copula Density." Unpublished, ECARES working paper, Université Libre de Bruxelles, Brussels, Belgium.

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

Genest, C., E. Masiello, and K. Tribouley. 2009. "Estimating Copula Densities through Wavelets." *Insurance: Mathematics and Economics* 44:170 – 181.

Genest, C., J.F. Quessy, and B. Rmillard. 2006. "Goodness-of-Fit Procedures for

Copula Models Based on the Probability Integral Transformation." *Scandinavian Journal of Statistics* 33:337–366.

Genest, C., and J. Segers. 2010. "On the Covariance of the Asymptotic Empirical Copula Process." *Journal of Multivariate Analysis* 101:1837–1845.

Gijbels, I., and J. Mielniczuk. 1990. "Estimating the Density of a Copula Function." *Communications in Statistics. Theory and Methods* 19:445–464.

Goodwin, B.K., and A.P. Ker. 1998. "Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk Crop Insurance Contracts." *American Journal of Agricultural Economics* 80:139–153.

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

Hall, P., and B.U. Park. 2002. "New Methods for Bias Correction at Endpoints and Boundaries." *The Annals of Statistics* 30:1460–1479.

Hannig, J., and J.S. Marron. 2006. "Advanced Distribution Theory for SiZer." *Journal of the American Statistical Association* 101:484–499.

Harri, A., K.H. Coble, A.P. Ker, and B.J. Goodwin. 2011. "Relaxing Heteroscedasticity Assumptions in Area-Yield Crop Insurance Rating." *American Journal of Agricultural Economics* 93:707–717.

Harri, A., C. Erdem, K.H. Coble, and T.O. Knight. 2009. "Crop Yield Distributions: A Reconciliation of Previous Research and Statistical Tests for Normality." *Applied Economic Perspectives and Policy* 31:163–182.

Janssen, P., J. Swanepoel, and N. Veraverbeke. 2014. "A Note on the Asymptotic Behavior of the Bernstein Estimator of the Copula Density." *Journal of Multivariate Analysis* 124:480 – 487.

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

Jones, M. 1993. "Simple Boundary Correction for Kernel Density Estimation." *Statistics and Computing* 3:135–146.

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

Jones, M., and D. Henderson. 2007. "Kernel-Type Density Estimation on the Unit Interval." *Biometrika* 94:977–984.

Jones, M., J. Marron, and S. Sheather. 1996a. "Progress in Data-based Bandwidth Selection for Kernel Density Estimation." *Computational Statistics* 11:337–381.

Jones, M.C., J.S. Marron, and S.J. Sheather. 1996b. "A Brief Survey of Bandwidth Selection for Density Estimation." *Journal of the American Statistical Association* 91:401–407.

Just, R.E., and Q. Weninger. 1999. "Are Crop Yields Normally Distributed?" *American Journal of Agricultural Economics* 81:287–304.

Karunamuni, R., and T. Alberts. 2006. "A Locally Adaptive Transformation Method of Boundary Correction in Kernel Density Estimation." *Journal of Statistical Planning and Inference* 136:2936–2960.

Karunamuni, R.J., and T. Alberts. 2005. "A Generalized Reflection Method of Boundary Correction in Kernel Density Estimation." *Canadian Journal of Statistics* 33:497–509.

Kauermann, G., C. Schellhase, and D. Ruppert. 2013. "Flexible Copula Density Estimation with Penalized Hierarchical B-splines." *Scandinavian Journal of Statistics* 40.

Ker, A.P. 1996. "Using SNP Maximum Likelihood Techniques to Recover Conditional Densities: A New Approach to Recovering Premium Rates." Unpublished, Department of Agricultural and Resource Economics,University of Arizona.

Ker, A.P., and K. Coble. 2003. "Modeling Conditional Yield Densities." *American*

Journal of Agricultural Economics 85:291–304.

Ker, A.P., and A. Ergun. 2005. "Empirical Bayes Nonparametric Kernel Density Estimation." *Statistics & Probability Letters* 75:315 – 324.

Ker, A.P., and B.K. Goodwin. 2000. "Nonparametric Estimation of Crop Insurance Rates Revisited." *American Journal of Agricultural Economics* 82:463–478.

Klugman, S.A., and R. Parsa. 1999. "Fitting Bivariate Loss Distributions with Copulas." *Insurance: Mathematics and Economics* 24:139 – 148.

Koekemoer, G., and J.W. Swanepoel. 2008. "Transformation Kernel Density Estimation With Applications." *Journal of Computational and Graphical Statistics* 17:750–769.

Mahul, O. 1999. "Optimum Area Yield Crop Insurance." *American Journal of Agricultural Economics* 81:75–82.

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

Miranda, M.J., and J.W. Glauber. 1997. "Systemic Risk, Reinsurance, and the Failure of Crop Insurance Markets." *American Journal of Agricultural Economics* 79:206–215.

Moss, C.B., and J.S. Shonkwiler. 1993. "Estimating Yield Distributions with a Stochastic Trend and Nonnormal Errors." *American Journal of Agricultural Economics* 75:1056–1062.

Müller, H.G. 1991. "Smooth Optimum Kernel Estimators near Endpoints." *Biometrika* 78:521–530.

Nelsen, R.B. 2006. *An Introduction to Copulas*. New York: Springer.

Nelson, C.H. 1990. "The Influence of Distributional Assumptions on the Calculation of Crop Insurance Premia." *North Central Journal of Agricultural Economics*

12:71–78.

Norwood, B., M.C. Roberts, and J.L. Lusk. 2004. "Ranking Crop Yield Models Using Out-of-Sample Likelihood Functions." *American Journal of Agricultural Economics* 86:1032–1043.

Owen, A.B. 2001. *Empirical Likelihood*. London: Chapman and Hall.

—. 1988. "Empirical Likelihood Ratio Confidence Intervals for a Single Functional." *Biometrika* 75:237–249.

—. 1990. "Empirical Likelihood Ratio Confidence Regions." *The Annals of Statistics* 18:90–120.

Ozaki, V.A., S.K. Ghosh, B.K. Goodwin, and R. Shirota. 2008. "Spatio-Temporal Modeling of Agricultural Yield Data with an Application to Pricing Crop Insurance Contracts." *American Journal of Agricultural Economics* 90:951–961.

Qu, L., and W. Yin. 2012. "Copula Density Estimation by Total Variation Penalized Likelihood with Linear Equality Constraints." *Computational Statistics & Data Analysis* 56:384 – 398.

Ramírez, O.A., and T. McDonald. 2006. "Ranking Crop Yield Models: A Comment." *American Journal of Agricultural Economics* 88:1105–1110.

Ruppert, D., and D.B.H. Cline. 1994. "Bias Reduction in Kernel Density Estimation by Smoothed Empirical Transformations." *The Annals of Statistics* 22:185–210.

Schuster, E.F. 1999. "Incorporating Support Constraints into Nonparametric Estimators of Densities." *Communications in Statistics - Theory and Methods* 14:1123–1136.

Shen, X., Y. Zhu, and L. Song. 2008. "Linear B-spline Copulas with Applications to Nonparametric Estimation of Copulas." *Computational Statistics & Data Analysis* 52:3806 – 3819.

Sherrick, B.J., F.C. Zanini, G.D. Schnitkey, and S.H. Irwin. 2004. "Crop Insurance

Valuation under Alternative Yield Distributions." *American Journal of Agricultural Economics* 86:406–419.

Silverman, B.W. 1986. *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

Simonoff, J. 1996. *Smoothing Methods in Statistics*. New York: Springer.

Skees, J.R., J.R. Black, and B.J. Barnett. 1997. "Designing and Rating an Area Yield Crop Insurance Contract." *American Journal of Agricultural Economics* 79:430–438.

Sklar, A. 1959. "Fonctions de répartition à n dimensions et leurs marges." *Publ. Inst. Statist. Univ. Paris* 8:229–231.

Stokes, J.R. 2000. "A Derivative Security Approach to Setting Crop Revenue Coverage Insurance Premiums." *Journal of Agricultural and Resource Economics* 25:159–176.

Wand, M., and M. Jones. 1995. *Kernel Smoothing*. London: Chapman and Hall.

Wand, M.P., J.S. Marron, and D. Ruppert. 1991. "Transformations in Density Estimation." *Journal of the American Statistical Association* 86:343–353.

Yang, L., and J.S. Marron. 1999. "Iterated Transformation-Kernel Density Estimation." *Journal of the American Statistical Association* 94:580–589.

Zhang, S., and R.J. Karunamuni. 1998. "On Kernel Density Estimation near Endpoints." *Journal of Statistical Planning and Inference* 70:301–316.

—. 2000. "On Nonparametric Density Estimation at the Boundary." *Journal of Nonparametric Statistics* 12:197–221.

Zhang, S., R.J. Karunamuni, and M.C. Jones. 1999. "An Improved Estimator of the Density Function at the Boundary." *Journal of the American Statistical Association* 94:1231–1241.

APPENDIX A

APPENDIX MATERIAL FOR SECTION 2

A.1 Positive semi-definiteness of $A_3 - \mathbf{A}_2^T \mathbf{A}_1^{-1} \mathbf{A}_2$

Define a 3×3 matrix $\mathbf{B}$ by

$$\int_{-\infty}^{\infty} \left[(EY_1 - y)f_Y(y), (EY_1^2 - y^2)f_Y(y), f_Y''(y)\right]^T \left[(EY_1 - y)f_Y(y), (EY_1^2 - y^2)f_Y(y), f_Y''(y)\right] dy$$

It is obvious that $\mathbf{B}$ is symmetric and positive semi-definite and can be rewritten as a block matrix:

$$\mathbf{B} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_2^T & A_3, \end{bmatrix},$$

where $A_1, \mathbf{A}_2$ and $\mathbf{A}_3$ are given in (2.12).

Note that $A_3 - \mathbf{A}_2^T \mathbf{A}_1^{-1} \mathbf{A}_2$ is the Schur complement of A_3. According to the Schur complement lemma (see e.g, Boyd and Vandenberghe (2004)), the Schur complement of A_3 in $\mathbf{B}$ is positive semi-definite if and only if $\mathbf{B}$ is positive semi-definite. It can be verified readily that if f_X is uniform, f_Y is standard normal and $A_3 - \mathbf{A}_2^T \mathbf{A}_1^{-1} \mathbf{A}_2 = 0$.

A.2 Estimation of $\mathbf{A_1}, \mathbf{A_2}, A_3$

The following notations are needed for the derivations in this appendix.

$$\mu_{s,t} = \int_{-\infty}^{\infty} G_{s,t}(Y) f_Y(y) dy,$$
$$V_{s,t} = \int_{-\infty}^{\infty} G_{s,t}^2(Y) f_Y(y) dy,$$
$$C_{s,t}^{(r)} = \int_{-\infty}^{\infty} G_{s,t}^2(y) f_Y(y) f_Y^{(r)}(y) dy,$$

$$D_{s,t}^{(r)} = \int_{-\infty}^{\infty} G_{s,t}^2(y) f_Y(y) \{f_Y^{(r)}(y)\}^2 dy,$$

$$\theta_r = K^{(r)}(0)$$

$$\kappa_r = \int_{-\infty}^{\infty} K^{(r)}(y)^2 dy.$$

Recall that the quantities in (2.18) are defined by $A_{s,t}^{(r)} = \int G_{s,t}(y) f_Y(y) f_Y^{(r)}(y) dy$. I first consider a "non-leave-one-one" estimator

$$\hat{A}_{s,t}^{(r)} = \frac{1}{n^2 b^{r+1}} \sum_{i=1}^{n} \sum_{j=1}^{n} G_{s,t}(Y_i) K^{(r)} \left(\frac{Y_i - Y_j}{b} \right),$$

where $K^{(r)}(\cdot)$ is the rth derivative of the kernel function $K(\cdot)$ that is taken to be the standard normal density function. Note that $\hat{A}_{s,t}^{(r)}$ can be decomposed into two parts

$$\hat{A}_{s,t}^{(r)} = \frac{1}{n^2 b^{r+1}} \sum_{i=1}^{n} G_{s,t}(Y_i) K^{(r)}(0) + \frac{1}{n^2 b^{r+1}} \sum_{i=1}^{n} \sum_{j \neq i} G_{s,t}(Y_i) K^{(r)} \left(\frac{Y_i - Y_j}{b} \right).$$

To approximate its bias, I have

$$\begin{aligned}
E[\hat{A}_{s,t}^{(r)}] &= \frac{\theta_r \mu_{s,t}}{n b^{r+1}} + \frac{n-1}{n b^{r+1}} E \left\{ G_{s,t}(Y_i) K^{(r)} \left(\frac{Y_i - Y_j}{b} \right) \right\} \\
&= \frac{\theta_r \mu_{s,t}}{n b^{r+1}} + \frac{n-1}{n b^{r+1}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_{s,t}(x) K^{(r)} \left(\frac{x-y}{b} \right) f_Y(x) f_Y(y) dx dy \\
&= \frac{\theta_r \mu_{s,t}}{n b^{r+1}} + \frac{n-1}{n b} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_{s,t}(x) K \left(\frac{x-y}{b} \right) f_Y(x) f_Y^{(r)}(y) dx dy \\
&= \frac{\theta_r \mu_{s,t}}{n b^{r+1}} + \frac{n-1}{n} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_{s,t}(x) K(u) f_Y(x) f_Y^{(r)}(x+bu) dx du \\
&\approx \frac{\theta_r \mu_{s,t}}{n b^{r+1}} + \frac{n-1}{n} \left\{ A_{s,t}^{(r)} + \frac{b^2}{2} \int_{-\infty}^{\infty} G_{s,t}(x) f_Y(x) f_Y^{(r+2)}(x) dx \right\}
\end{aligned}$$

Thus the asymptotic bias of $\hat{G}_{s,t}$ is

$$\text{Abias}[\hat{B}_r] = \frac{\theta_r \mu_{s,t}}{n b^{r+1}} + \frac{b^2}{2} A_{s,t}^{(r+2)}.$$

To derive the asymptotic variance of $\hat{A}_{s,t}^{(r)}$, I have

$$\begin{aligned}
\mathrm{Var}[\hat{A}_{s,t}^{(r)}] =& \frac{1}{n^4 b^{2r+2}} \mathrm{Var}\left\{ \sum_i \sum_j G_{s,t}(Y_i) K^{(r)}\left(\frac{Y_i - Y_j}{b}\right) \right\} \\
=& \frac{1}{n^4 b^{2r+2}} \mathrm{Cov}\left\{ \sum_i \sum_j G_{s,t}(Y_i) K^{(r)}\left(\frac{Y_i - Y_j}{b}\right), \right. \\
& \left. \sum_i \sum_j G_{s,t}(Y_i) K^{(r)}\left(\frac{Y_i - Y_j}{b}\right) \right\} \\
=& \frac{1}{n^4 b^{2r+2}} \left\{ n\mathrm{Var}\left(G_{s,t}(Y_1) K^{(r)}(0)\right) \right. \\
& + 4n(n-1)\mathrm{Cov}\left(G_{s,t}(Y_1) K^{(r)}(0), G_{s,t}(Y_1) K^{(r)}\left(\frac{Y_1 - Y_2}{b}\right)\right) \\
& + n(n-1)\mathrm{Var}\left(G_{s,t}(Y_1) K^{(r)}\left(\frac{Y_1 - Y_2}{b}\right)\right) \\
& + n(n-1)\mathrm{Cov}\left(G_{s,t}(Y_1) K^{(r)}\left(\frac{Y_1 - Y_2}{b}\right), G_{s,t}(Y_2) K^{(r)}\left(\frac{Y_2 - Y_1}{b}\right)\right) \\
& + 4n(n-1)(n-2)\mathrm{Cov}\left(G_{s,t}(Y_1) K^{(r)}\left(\frac{Y_1 - Y_2}{b}\right), \right. \\
& \left. \left. G_{s,t}(Y_1) K^{(r)}\left(\frac{Y_1 - Y_3}{b}\right)\right)\right\}
\end{aligned}$$

I can approximate each term above using the following

$$\begin{aligned}
& \mathrm{Var}\left(G_{s,t}(Y_1) K^{(r)}(0)\right) \\
& \approx \theta_r^2 (V_{s,t} - \mu_{s,t}^2) \\
& \mathrm{Cov}\left(G_{s,t}(Y_1) K^{(r)}(0), G_{s,t}(Y_1) K^{(r)}\left(\frac{Y_1 - Y_2}{b}\right)\right) \\
& \approx b^{r+1} \theta_r (C_{s,t}^{(r)} - \mu_{s,t} A_{s,t}^{(r)}) \\
& \mathrm{Var}\left(G_{s,t}(Y_1) K^{(r)}\left(\frac{Y_1 - Y_2}{b}\right)\right)
\end{aligned}$$

$$\approx b\kappa_r C_{s,t}^{(0)} - b^{2r+2}\{A_{s,t}^{(r)}\}^2$$

$$\mathrm{Cov}\left(G_{s,t}(Y_1)K^{(r)}\left(\frac{Y_1-Y_2}{b}\right), G_{s,t}(Y_2)K^{(r)}\left(\frac{Y_2-Y_1}{b}\right)\right)$$

$$\approx b\kappa_r C_{s,t}^{(0)} - b^{2r+2}\{A_{s,t}^{(r)}\}^2$$

$$\mathrm{Cov}\left(G_{s,t}(Y_1)K^{(r)}\left(\frac{Y_1-Y_2}{b}\right), G_{s,t}(Y_1)K^{(r)}\left(\frac{Y_1-Y_3}{b}\right)\right)$$

$$\approx b^{2r+2}(D_{s,t}^{(r)} - \{A_{s,t}^{(r)}\}^2).$$

Plugging the approximations into the asymptotic variance of $\hat{A}_{s,t}^{(r)}$ yields

$$\mathrm{Avar}[\hat{A}_{s,t}^{(r)}] = \frac{\theta_r^2(V_{s,t}-\mu_{s,t}^2)}{n^3 b^{2r+2}} + \frac{4\theta_r(C_{s,t}^{(0)} - \mu_{s,t}A_{s,t}^{(r)})}{n^2 b^{b+1}} + \frac{2\kappa_r C_{s,t}^{(0)}}{n^2 b^{2r+1}} - \frac{2\{A_{s,t}^{(r)}\}^2}{n^2} + \frac{4(D_{s,t}^{(r)} - \{A_{s,t}^{(r)}\}^2)}{n}.$$

This "non-leave-one-out" estimator is mainly designed for $A_{s,t}^{(r)}$'s that satisfy a certain condition such that the leading bias can be removed. In particular, if

$$\text{(A.1)}\quad \frac{\theta_r \mu_{s,t}}{A_{s,t}^{(r+2)}} < 0,$$

I can set

$$\frac{\theta_r \mu_{s,t}}{n b^{r+1}} + \frac{b^2}{2} A_{s,t}^{(r+2)} = 0.$$

It follows that an optimal bandwidth is then given by

$$\text{(A.2)}\quad b_{s,t}^* = \left[-\frac{2\theta_r \mu_{s,t}}{A_{s,t}^{(r+2)}}\right]^{1/(r+3)} n^{-1/(r+3)}.$$

With this optimal bandwidth, the third term of the $\mathrm{Avar}[\hat{A}_{s,t}^{(r)}]$ becomes the leading term which is of order $O\left(n^{-5/(r+3)}\right)$.

For $A_{s,t}^{(r)}$'s that do not satisfy condition (A.1), I use the "leave-one-out" estimator

to estimate them consistently, which is given by

$$\hat{A}_{s,t}^{(r)} = \frac{1}{n(n-1)b^{r+1}} \sum_{i=1}^{n} \sum_{j \neq i} G_{s,t}(Y_i) K^{(r)} \left(\frac{Y_i - Y_j}{b} \right).$$

The asymptotic bias and variance can be obtained in a similar manner as those of the "non-leave-one-out" estimators. I have

$$\begin{aligned} \text{Abias}[\hat{A}_{s,t}^{r}] &= \frac{b^2}{2} A_{s,t}^{(r+2)}, \\ \text{Avar}[\hat{A}_{s,t}^{(r)}] &= \frac{2\kappa_r C_{s,t}^{(0)}}{n^2 b^{2r+1}} - \frac{2\{A_{s,t}^{(r)}\}^2}{n^2} + \frac{4(D_{s,t}^{(r)} - A_{s,t}^{(r)}\}^2)}{n}. \end{aligned}$$

An optimal bandwidth is then given by

$$\text{(A.3)} \quad b_{s,t}^{*} = \left[\frac{(4r+2)\kappa_r C_{s,t}^{(0)}}{\{A_{s,t}^{(r+2)}\}^2} \right]^{1/(2r+5)} n^{-2/(2r+5)}.$$

Under this bandwidth, the MSE of the leave-one-out estimator is of order $O\left(n^{-8/(2r+5)}\right)$, which is larger than that of the "non-leave-one-out" estimator for large r. Thus, the "non-leave-one-out" estimator shall be preferred if condition (A.1) is satisfied.

Next I investigate which estimators are suitable to estimate the various $A_{s,t}^{(r)}$'s given in (2.18). Consider first $A_3 = A_{0,0}^{(4)}$, where $G_{0,0} = 1$. I have $\theta_4 = 3/\sqrt{2\pi} > 0$ and $A_{0,0}^{(6)} = -\int \{f_Y^{(r)}(y)\}^2 dy < 0$. Thus condition (A.1) is satisfied and the "non-leave-one-out" estimator shall be used, as recommended by Wand and Jones (1995). For other quantities given by (2.18), it is straightforward to verify that condition (A.1) is not satisfied and thus the "leave-one-out" estimator shall be used.

Lastly regarding the bandwidth used in these estimations, our experiments indicate that A_3 is the most difficult to estimate while the estimations of others are not sensitive to the bandwidth. I therefore use the optimal bandwidth for A_3 in the estimation of all quantities in (2.18). The optimal bandwidth given by (A.2)

requires the estimation of $A^4_{0,0}$, which for simplicity is calculated using the rule of thumb principle under the assumption of normality. Thus the optimal bandwidth is given by

$$b^* = s \left[\frac{16\sqrt{2}}{5} \right]^{1/7} n^{-1/7},$$

where s is the standard deviation of the transformed data $Y_1, \ldots, Y_n$.

A.3 Simulation details

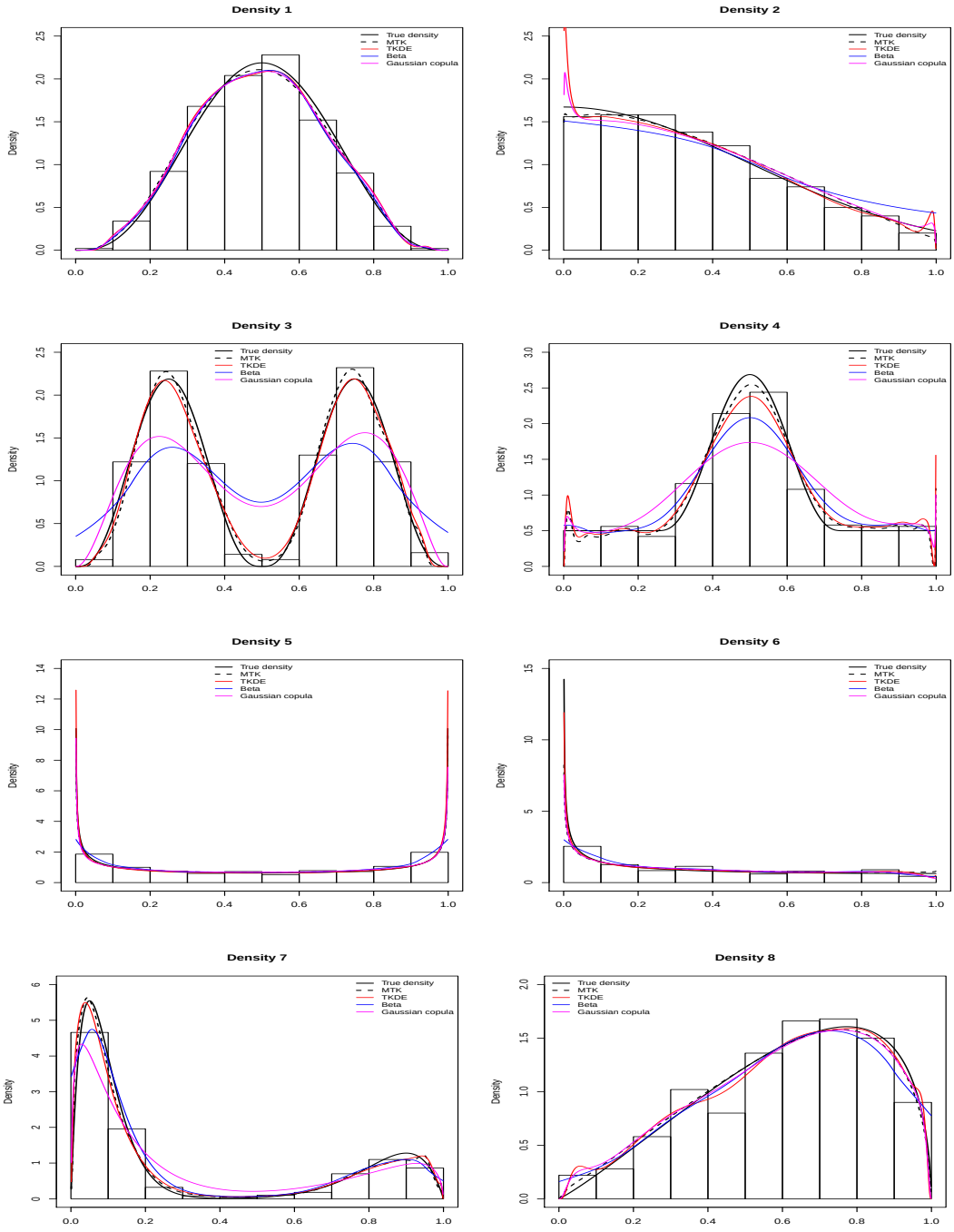
I present the densities of the distributions used in our simulations in the table below. In this table, $\mathbf{1}_{\{.\}}$ is the indicator function. $\text{Beta}(\cdot, \cdot)$ denotes the beta density and $\text{Beta}_{[a,b]}(\cdot, \cdot)$ denotes the corresponding beta density rescaled to the interval $[a, b]$. "Truncated" means the original density is truncated to the interval $[0, 1]$.

Table A.1: Densities used in the simulation

#	$f_x(x), x \in [0,1]$	Description
1	$140x^3(1-x)^3$	Beta(4,4)
2	$2\exp(-2x^2)[\sqrt{2\pi}\{\Phi(2) - 1/2\}]^{-1}$	Truncated $2\phi(2x)$
3	$1120\left[x^3(1-2x)^3\mathbf{1}_{\{x\leq 1/2\}} + 8(x-1/2)^3(1-x)^3\mathbf{1}_{\{x\geq 1/2\}}\right]$	$\frac{1}{2}\text{Beta}_{[0,1/2]}(4,4) + \frac{1}{2}\text{Beta}_{[1/2,1]}(4,4)$
4	$1/2 + 140(2x-1/2)^3(3/2-2x)^3\mathbf{1}_{\{1/4\leq x\leq 3/4\}}$	$\frac{1}{2}\text{Beta}(1,1) + \frac{1}{2}\text{Beta}_{[1/4,3/4]}(4,4)$
5	$\left[\pi\sqrt{x(1-x)}\right]^{-1}$	$\text{Beta}(1/2, 1/2)$
6	$2\left[\pi\sqrt{x(2-x)}\right]^{-1}$	Truncated $\text{Beta}_{[0,2]}(1/2, 1/2)$
7	$294x(1-x)^{19} + 33x^9(1-x)$	$\frac{7}{10}\text{Beta}(2,20) + \frac{3}{10}\text{Beta}(10,2)$
8	$c(x, 0.7; 0.7)$	Conditional Gaussian copula with $\rho = 0.7$

A.4 Examples of estimated densities

Below I report a random example of estimated densities for $n = 500$. Also reported are the true densities and histograms of data.


Density 1
Density 2
Density 3
Density 4
Density 5
Density 6
Density 7
Density 8
True density
MTK
TKDE
Beta
Gaussian copula
Density

APPENDIX B

APPENDIX MATERIAL FOR SECTION 3

B.1 Assumptions

Assumption 1. *The sample* $\{(X_i, Y_i)^\top, i = 1, \cdots, n\}$ *is an i.i.d. sample from the joint distribution* F *that is absolutely continuous. The associated marginal distributions* F_X *and* F_Y *are strictly increasing on their support.*

Assumption 2. *The copula* C *of* F *is such that* $(\partial C/\partial u)(u, v)$ *and* $(\partial^2 C/\partial u^2)(u, v)$ *exist and are continuous on* $\{(u, v) : u \in (0, 1), v \in [0, 1]\}$*, and* $(\partial C/\partial v)(u, v)$ *and* $(\partial^2 C/\partial v^2)(u, v)$ *exist and are continuous on* $\{(u, v) : u \in [0, 1], v \in (0, 1)\}$*. In addition, there are constants* K_1 *and* K_2 *such that*

$$\left|\frac{\partial^2 C}{\partial u^2}(u, v)\right| \leq \frac{K_1}{u(1-u)} \quad \forall (u, v) \in (0, 1) \times [0, 1]$$

and

$$\left|\frac{\partial^2 C}{\partial v^2}(u, v)\right| \leq \frac{K_2}{v(1-v)} \quad \forall (u, v) \in [0, 1] \times (0, 1).$$

Assumption 3. *The copula density* c *exists, is positive and admits continuous second order partial derivatives on the interior of* $\mathcal{I}$*. In addition, there is a constant* K_{00} *such that*

$$c(u, v) \leq K_{00} \min\left(\frac{1}{u(1-u)}, \frac{1}{v(1-v)}\right) \quad \forall (u, v) \in (0, 1)^2.$$

Assumption 4. *As* $n \to \infty$*,* $\boldsymbol{\Theta} \to 0$*. In particular, assume* $\boldsymbol{\Theta} \sim h^2$*, which is optimal.*

Assumption 5. *Under the diagonal bandwidth matrix* $\boldsymbol{H} = h^2 \boldsymbol{I}$*,* $h \sim n^{-a}$ *where* $a \in \left[\frac{1}{6}, \frac{1}{4}\right)$*.*

Assumption 6. *Under the diagonal bandwidth matrix* $\boldsymbol{H} = h^2\boldsymbol{I}$, $h \sim n^{-a}$ *where* $a \in \left[\frac{1}{10}, \frac{1}{4}\right)$.

B.2 Proofs

B.2.1 Proof of Theorem 1

Proof. First note that

$$\begin{aligned}\hat{c}_{m1}(u,v) &= J(u,v;h,\boldsymbol{\Theta})\hat{c}_{t1}(u,v)\\ &= \left(J(u,v;h,\boldsymbol{\Theta}) - J^*(u,v;h,\boldsymbol{\Theta})\right)\hat{c}_{t1}(u,v) + J^*(u,v;h,\boldsymbol{\Theta})\hat{c}_{t1}(u,v),\end{aligned}$$

and

$$b_{m1}(u,v) = b_{t1}(u,v) + \boldsymbol{\Theta}^\top \boldsymbol{B}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) c(u,v),$$

it follows

$$\begin{aligned}&\sqrt{nh^2}\left(\hat{c}_{m1}(u,v) - c(u,v) - b_{m1}(u,v)\right)\\ =&\left(J(u,v;h,\boldsymbol{\Theta}) - J^*(u,v;h,\boldsymbol{\Theta})\right)\sqrt{nh^2}\hat{c}_{t1}(u,v)\\ &+\sqrt{nh^2}\left(J^*(u,v;h,\boldsymbol{\Theta})\hat{c}_{t1}(u,v) - c(u,v) - b_{t1}(u,v)\right.\\ &\left.-\boldsymbol{\Theta}^\top \boldsymbol{B}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) c(u,v)\right)\\ \equiv& I_1 + I_2.\end{aligned}$$

For I_1, I have

$$\begin{aligned}I_1 = &\left(J(u,v;h,\boldsymbol{\Theta}) - J^*(u,v;h,\boldsymbol{\Theta})\right)\sqrt{nh^2}\left(\hat{c}_{t1}(u,v) - c(u,v) - b_{t1}(u,v)\right)\\ &+\left(J(u,v;h,\boldsymbol{\Theta}) - J^*(u,v;h,\boldsymbol{\Theta})\right)\sqrt{nh^2}\left(c(u,v) + b_{t1}(u,v)\right).\end{aligned}$$

Both $J(u,v;h,\boldsymbol{\Theta})$ and $J^*(u,v;h,\boldsymbol{\Theta})$ are in the form of sample average. It is easy to show that, by a simple Taylor expansion, $J(u,v;h,\boldsymbol{\Theta}) - J^*(u,v;h,\boldsymbol{\Theta}) = o_p\left(n^{-1/2}\right) = o_p\left((nh^2)^{-1/2}\right)$. From (3.9), I have $\sqrt{nh^2}\left(\hat{c}_{t1}(u,v) - c(u,v) - b_{t1}(u,v)\right) = O_p(1)$; together with the above result, the first term in I_1 is $o_p(1)$. Since $J(u,v;h,\boldsymbol{\Theta}) -$

$J^*(u,v;h,\boldsymbol{\Theta}) = o_p\left((nh^2)^{-1/2}\right)$, the second term in I_1 is obviously $o_p(1)$ as well. Therefore, I have $I_1 = o_p(1)$.

Consider I_2, after plugging (3.12) in, and it follows

$$\begin{aligned} I_2 &= \sqrt{nh^2}\left(\hat{c}_{t1}(u,v) - c(u,v) - b_{t1}(u,v)\right) \\ &+ \sqrt{nh^2}\boldsymbol{\Theta}^\top \boldsymbol{B}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)\left(\hat{c}_{t1}(u,v) - c(u,v)\right) \\ &+ \sqrt{nh^2}\hat{c}_{t1}(u,v) o\left(\boldsymbol{\Theta}\right) \\ &\equiv I_{21} + I_{22} + I_{23}. \end{aligned}$$

For I_{21}, according to (3.9), I have

$$I_{21} \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{m1}(u,v)\right),$$

since $\sigma^2_{m1}(u,v) = \sigma^2_{t1}(u,v)$. For I_{22}, I have

$$\begin{aligned} I_{22} = \sqrt{nh^2}\boldsymbol{\Theta}^\top \boldsymbol{B}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right)\left(\hat{c}_{t1}(u,v) - c(u,v) - b_{t1}(u,v)\right) \\ + \sqrt{nh^2}\boldsymbol{\Theta}^\top \boldsymbol{B}\left(\Phi^{-1}(u), \Phi^{-1}(v)\right) b_{t1}(u,v). \end{aligned}$$

It is easy to see that the first term in I_{22} is $o(1)\cdot O_p(1) = o_p(1)$. Since $b_{t1}(u,v) = O\left(h^2\right)$ and by assumption $\boldsymbol{\Theta} = O\left(h^2\right)$, I have the second term in I_{22} being $O\left(nh^{10}\right) = o(1)$ by the fact that $h \propto n^{-a}$ where $a \in [1/6, 1/4)$. Therefore, $I_{22} = o_p(1)$. For I_{23}, I have

$$I_{23} = \sqrt{nh^2}\left(\hat{c}_{t1}(u,v) - c(u,v) - b_{t1}(u,v)\right) o\left(\boldsymbol{\Theta}\right) + \sqrt{nh^2}\left(c(u,v) + b_{t1}(u,v)\right) o\left(\boldsymbol{\Theta}\right).$$

The first term in I_{23} is again $o(1)\cdot O_p(1) = o_p(1)$. For the second term in I_{23}, I have

$$\sqrt{nh^2}c(u,v)o\left(\boldsymbol{\Theta}\right) = \sqrt{nh^6}c(u,v)o(1) = o(1)$$

since $\sqrt{nh^6} = O(1)$; similarly I also have $\sqrt{nh^2}b_{t1}(u,v)o\left(\boldsymbol{\Theta}\right) = o(1)$. Thus, I have $I_{23} = o_p(1)$.

The proof concludes by combining the above results. □

B.2.2 Proof of Theorem 2

Proof. I first slightly extend the results stated in (3.9). Under the assumptions in Theorem 2, I have, for any $(u, v) \in (0, 1)^2$,

$$\sqrt{nh^2}\left(\hat{c}_{t1}(u,v) - c(u,v) - b_{t1}^{(G)}(u,v)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_{m1}^2(u,v)\right),$$

where $b_{t1}^{(G)}(u, v)$ is defined in (3.27). Since $b_{t1}^{(G)}(u, v)$ includes the higher order bias term that is associated with h^4, I can relax the condition to $h \sim n^{-a}$ where $a \in [1/10, 1/4)$.

Note that $b_{m1}^{(G)}(u, v) = b_{m1}(u, v) + h^4 R\left(\Phi^{-1}(u), \Phi^{-1}(v); \rho\right)$ since $b_{m1}(u, v) = 0$ in this case and use the higher order Taylor expansion of $J^*(u, v; h, \boldsymbol{\Theta})$, see (3.28). Then the proof is analogous to that in the Theorem 1 above, thus is omitted here. □

B.2.3 Proof of $\Gamma_3 - \boldsymbol{\Gamma}_2^\top \boldsymbol{\Gamma}_1^{-1} \boldsymbol{\Gamma}_2 \geq 0$

Proof. Define a 3-dimensional vector $q(s, t) = \left(\boldsymbol{B}(s, t)^\top g(s, t), g^{(2,0)}(s, t) + g^{(0,2)}(s, t)\right)^\top$ and a matrix

$$Q = \int_{\mathcal{R}^2} q(s,t) q(s,t)^\top \mathrm{d}s\mathrm{d}t.$$

By construction, I have $q(s, t)q(s, t)^\top$ is positive-semidefinite and so is Q. Note that Q can also be written in the block matrix form, namely

$$Q = \begin{pmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2 \\ \boldsymbol{\Gamma}_2^\top & \Gamma_3 \end{pmatrix}.$$

Note that $\Gamma_3 - \boldsymbol{\Gamma}_2^\top \boldsymbol{\Gamma}_1^{-1} \boldsymbol{\Gamma}_2$ is the Schur complement of Γ_3. According to Schur complement lemma, see Boyd and Vandenberghe (2004), the Schur complement of Γ_3 in Q is positive-semidefinite if and only if Q is positive-semidefinite. Therefore, I

have $\Gamma_3 - \mathbf{\Gamma}_2^\top \mathbf{\Gamma}_1^{-1} \mathbf{\Gamma}_2 \geq 0$. It is easy to check that when the underlying copula is the Gaussian copula, i.e. g is the pdf of bivariate Gaussian distribution, I have $\Gamma_3 - \mathbf{\Gamma}_2^\top \mathbf{\Gamma}_1^{-1} \mathbf{\Gamma}_2 = 0$. □

B.3 Exact formula of $\int_{\mathcal{I}} \left(\hat{c}_m(u,v)\right)^2 \phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))\mathrm{d}u\mathrm{d}v$

For the $\hat{c}_{m2}$ with the non-diagonal bandwidth matrix (3.29), I have

$$\int_{\mathcal{I}} \left(\hat{c}_{m2}(u,v)\right)^2 \phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))\mathrm{d}u\mathrm{d}v = \frac{1}{4\pi n^2 \eta^2 h^2 \delta \sqrt{1-\lambda^2}} \sum_{i=1}^{n}\sum_{j=1}^{n}$$
$$\exp\left\{ \frac{\alpha_1 \left(\hat{S}_i^2 + \hat{S}_j^2 + \hat{T}_i^2 + \hat{T}_j^2\right) + \alpha_2 \left(\hat{S}_i\hat{T}_i + \hat{S}_j\hat{T}_j\right)}{4h^2(1-\lambda^2)\delta^2} \right.$$
$$\left. + \frac{\alpha_3 \left(\hat{S}_i\hat{T}_j + \hat{S}_j\hat{T}_i\right) + \alpha_4 \left(\hat{S}_i\hat{S}_j + \hat{T}_i\hat{T}_j\right)}{4h^2(1-\lambda^2)\delta^2} \right\},$$

where

$$\alpha_1 = -2h^4(1-\lambda^2)(4\theta_1^2 - \theta_2^2) + 2h^2((\lambda^2-3)\theta_1 - \lambda\theta_2) - 1$$
$$\alpha_2 = 4h^4(1-\lambda^2)(4\theta_1^2 - \theta_2^2) + 2h^2(4\lambda\theta_1 + (3\lambda^2-1)\theta_2) + 2\lambda$$
$$\alpha_3 = -2h^2(4\lambda\theta_1 + (1+\lambda^2)\theta_2) - 2\lambda$$
$$\alpha_4 = 4h^2((1+\lambda^2)\theta_1 + \lambda\theta_2) + 2.$$

Then for the $\hat{c}_{m1}$, the result can be immediately obtained by setting $\lambda = 0$.

APPENDIX C

APPENDIX MATERIAL FOR SECTION 4

Below, I present the detailed 2012 insurance rates at the coverage level 80% estimated by the four competing estimators.

Table C.1: Iowa corn insurance rates at the coverage level 80% (all numbers are at percentage level)

County	KDE	ELK	% change	AKDE	ELAK	% change	County	KDE	ELK	% change	AKDE	ELAK	% change
ADAIR	2.8066	2.8619	1.97%	2.9723	3.0221	1.67%	JEFFERSON	4.1128	4.0569	-1.36%	4.0365	3.8976	-3.44%
ADAMS	3.4958	3.4077	-2.52%	3.5989	3.5032	-2.66%	JOHNSON	2.4105	2.7123	12.52%	2.4546	2.6844	9.36%
ALLAMAKEE	1.2389	1.5321	23.66%	1.3698	1.6672	21.72%	JONES	1.8404	1.7373	-5.60%	1.9896	1.8870	-5.16%
APPANOOSE	4.6877	4.6303	-1.23%	4.2023	4.0448	-3.75%	KEOKUK	3.7436	3.5270	-5.79%	3.8397	3.6434	-5.11%
AUDUBON	1.6851	2.0834	23.64%	1.8669	2.2483	20.43%	KOSSUTH	0.9627	1.0961	13.85%	1.0940	1.2089	10.50%
BENTON	1.9876	1.8811	-5.36%	2.1746	2.0604	-5.25%	LEE	5.2601	3.8764	-26.31%	5.2821	3.9181	-25.82%
BLACK HAWK	1.5823	1.5525	-1.88%	1.7568	1.7175	-2.23%	LINN	2.2217	2.0373	-8.30%	2.3221	2.1393	-7.87%
BOONE	2.0898	1.7483	-16.34%	2.0966	1.7928	-14.49%	LOUISA	3.0909	3.3301	7.74%	3.1637	3.3789	6.80%
BREMER	1.4850	1.4534	-2.13%	1.6589	1.6149	-2.65%	LUCAS	5.8826	5.4212	-7.84%	5.8379	5.3659	-8.09%
BUCHANAN	1.3981	1.6828	20.36%	1.5710	1.8704	19.06%	LYON	2.6537	1.7244	-35.02%	2.7883	1.9118	-31.43%
BUENA VISTA	1.4167	1.6403	15.78%	1.6333	1.8763	14.88%	MADISON	3.3085	3.1083	-6.05%	3.4033	3.2115	-5.64%
BUTLER	1.1024	1.4044	27.39%	1.2667	1.5644	23.51%	MAHASKA	3.4471	3.0510	-11.49%	3.6216	3.2346	-10.69%
CALHOUN	0.6707	1.4044	109.39%	0.7987	1.5814	98.00%	MARION	3.1300	3.2821	4.86%	2.8619	3.1017	8.38%
CARROLL	1.9154	1.6245	-15.19%	2.0704	1.7832	-13.87%	MARSHALL	2.1383	1.6523	-22.72%	2.1390	1.6697	-21.94%
CASS	1.9865	2.1613	8.80%	2.1765	2.3587	8.37%	MILLS	1.8494	2.0980	13.45%	1.9463	2.2316	14.66%
CEDAR	1.9110	1.8773	-1.77%	2.0816	2.0396	-2.02%	MITCHELL	2.8099	1.8420	-34.45%	2.9486	2.0320	-31.09%
CERRO GORDO	1.0815	1.3838	27.95%	1.2082	1.5384	27.33%	MONONA	2.0055	2.3987	19.60%	2.1441	2.5453	18.71%
CHEROKEE	1.2000	1.5147	26.22%	1.3923	1.7180	23.40%	MONROE	4.7369	4.4757	-5.51%	4.5998	4.3160	-6.17%
CHICKASAW	1.9454	1.9526	0.37%	2.1402	2.1218	-0.86%	MONTGOMERY	1.8578	2.3489	26.44%	2.0244	2.5239	24.67%
CLARKE	6.4159	5.7273	-10.73%	6.1043	5.4590	-10.57%	MUSCATINE	2.1776	2.4594	12.94%	2.3056	2.5188	9.25%
CLAY	1.4966	1.5914	6.33%	1.6271	1.6947	4.16%	O BRIEN	0.8906	1.3974	56.91%	1.0026	1.5329	52.89%
CLAYTON	1.1684	1.3652	16.84%	1.3247	1.5424	16.43%	OSCEOLA	1.6583	1.5931	-3.93%	1.8091	1.7784	-1.70%
CLINTON	1.9723	1.7893	-9.28%	2.0905	1.8968	-9.27%	PAGE	2.8685	2.9858	4.09%	2.9654	3.0565	3.07%
CRAWFORD	2.6455	1.9248	-27.24%	2.8482	2.1276	-25.30%	PALO ALTO	1.2124	1.3755	13.45%	1.3613	1.5177	11.49%
DALLAS	1.5873	2.2985	44.81%	1.7322	2.4023	38.68%	PLYMOUTH	2.2647	1.9288	-14.83%	2.4222	2.0500	-15.36%
DAVIS	5.8830	4.7989	-18.43%	5.5122	4.5642	-17.20%	POCAHONTAS	1.2846	1.2248	-4.66%	1.4519	1.3856	-4.57%
DECATUR	5.4082	5.0250	-7.08%	5.0497	4.6875	-7.17%	POLK	1.6930	2.3764	40.37%	1.8625	2.5160	35.09%
DELAWARE	1.5805	1.5660	-0.92%	1.7405	1.7479	0.42%	POTTAWATTAMIE	1.9530	1.7038	-12.76%	2.0781	1.8321	-11.84%
DES MOINES	2.7625	3.4435	24.65%	2.8740	3.5368	23.06%	POWESHIEK	2.2264	2.2818	2.49%	2.3809	2.4099	1.22%
DICKINSON	2.2616	1.7257	-23.70%	2.3617	1.8493	-21.70%	RINGGOLD	5.9673	5.7407	-3.80%	5.9567	5.7260	-3.87%
DUBUQUE	1.3153	1.6100	22.41%	1.4275	1.7577	23.13%	SAC	1.8187	1.5867	-12.75%	2.0054	1.7421	-13.13%
EMMET	1.6208	1.3003	-19.77%	1.6921	1.4061	-16.90%	SCOTT	1.6289	1.8631	14.38%	1.8080	2.0439	13.05%
FAYETTE	1.1358	1.6370	44.13%	1.2842	1.7972	39.95%	SHELBY	1.6507	1.9096	15.69%	1.8329	2.0910	14.08%
FLOYD	2.2324	1.9732	-11.61%	2.3659	2.0985	-11.30%	SIOUX	1.5717	1.6837	7.13%	1.7369	1.8259	5.12%
FRANKLIN	0.9615	1.1629	20.95%	1.0971	1.3205	20.37%	STORY	2.2352	1.8608	-16.75%	2.2300	1.9423	-12.90%
FREMONT	2.3490	2.3443	-0.20%	2.4908	2.5080	0.69%	TAMA	0.9909	1.8888	90.62%	1.1796	2.0794	76.28%
GREENE	1.7821	1.4418	-19.10%	1.8001	1.5119	-16.01%	TAYLOR	4.7786	4.2652	-10.74%	4.7833	4.2849	-10.42%
GRUNDY	1.2109	1.4278	17.92%	1.3161	1.5315	16.37%	UNION	4.4378	3.8284	-13.73%	4.5527	3.9399	-13.46%
GUTHRIE	1.7704	2.4820	40.20%	1.9562	2.6632	36.14%	VAN BUREN	5.5852	4.8650	-12.90%	5.5273	4.8451	-12.34%
HAMILTON	1.5501	1.5359	-0.92%	1.6676	1.6414	-1.57%	WAPELLO	3.0499	4.4822	46.96%	3.1354	4.5754	45.93%
HANCOCK	0.9084	1.0334	13.76%	1.0318	1.1551	11.95%	WARREN	3.3476	3.2812	-1.98%	3.2553	3.2350	-0.62%
HARDIN	1.2143	1.5317	26.14%	1.3550	1.6587	22.41%	WASHINGTON	3.0331	2.9471	-2.83%	3.0512	2.9729	-2.57%
HARRISON	1.3586	1.9432	43.03%	1.4484	2.0511	41.61%	WAYNE	5.8690	5.4350	-7.39%	5.2448	4.7009	-10.37%
HENRY	3.4884	3.4956	0.20%	3.5974	3.6099	0.35%	WEBSTER	1.0539	1.3244	25.67%	1.1943	1.4567	21.97%
HOWARD	3.0742	2.1487	-30.11%	3.2116	2.2888	-28.73%	WINNEBAGO	1.0296	1.0198	-0.96%	1.1430	1.1274	-1.37%
HUMBOLDT	0.9443	1.0432	10.47%	1.0741	1.1513	7.19%	WINNESHIEK	1.3270	1.6285	22.72%	1.5080	1.7671	17.18%
IDA	2.1001	1.6278	-22.49%	2.2406	1.7713	-20.94%	WOODBURY	1.9079	2.2004	15.33%	2.0389	2.3197	13.77%
IOWA	2.9297	2.5086	-14.37%	2.9322	2.5495	-13.05%	WORTH	1.1684	1.2510	7.06%	1.3077	1.3804	5.56%
JACKSON	2.3625	1.9898	-15.77%	2.4897	2.0424	-17.97%	WRIGHT	0.8959	1.1446	27.76%	1.0231	1.2572	22.88%
JASPER	1.9348	2.3608	22.01%	2.0357	2.4546	20.58%							