Spatially Smoothed Empirical Likelihood Kernel Density Estimation with Application to Crop Yield Distributions

Kuangyu Wen, Ximing Wu†, David Leatham‡

Abstract

This study concerns the estimation of many densities, each with a small number of observations. These densities tend to resemble one another if they are spatially proximate. To gain flexibility and improve efficiency, we propose kernel-based estimators that are refined by empirical likelihood probability weights derived under spatially smoothed moment conditions. We construct spatially smoothed moments based on spline functions, which are robust to outliers and facilitate flexible customizations. We use these methods to estimate the corn yield distributions of Iowa counties and to predict the premiums of crop insurance programs. Monte Carlo simulations and an empirical application demonstrate the good performance and usefulness of the proposed methods.

JEL Classification: C14, C16, R12, Q10

Keywords: Crop Yield Distributions; Empirical Likelihood; Insurance; Kernel Density Estimation; Spatial Smoothing
1 Introduction

Estimation of probability density functions plays an important role in exploring the stochastic structure of observed data and has been a powerful tool in economic and statistical analyses. For instance, density estimation is customarily used to examine certain features of the global distribution of labor productivity such as bi-modality; the widely-used Value at Risk and Expected Shortfall characterize the left tail of stock return distributions. In this study we are concerned with estimating densities that share certain similarity due to spatial proximity. Prime examples include the crop yield distributions of many geographic locations (Goodwin and Ker, 1998; Racine and Ker, 2006; Harri et al., 2011; Tack and Holt, 2015) and the housing price indices across cites (Iversen Jr, 2001; Banerjee et al., 2004; Majumdar et al., 2006; Brady, 2011; Baltagi and Li, 2014; Baltagi et al., 2015). The data used in these studies often consist of a large number of geographic units with a relatively small number of observations for each unit.

We aim to design suitable density estimators that meet the two following objectives: (i) flexibility in functional forms to avoid specification errors and (ii) incorporation of spatial smoothing to improve estimation efficiency. Flexibility can be achieved by employing nonparametric density estimators, which are known to excel when there are abundant observations. This approach, however, is hampered by the typical small sample size for each individual unit in the type of investigations considered in the present study. This practical difficulty motivates our second objective of exploiting the similarity—due to spatial proximity—among the densities for possible efficiency gains. Spatial methods have been widely used in parametric analyses, but their applications to nonparametric analyses are relatively few. In particular, see Tran (1990), Carbon et al. (1997), Hallin et al. (2001, 2004a), Lee et al. (2004) for density estimation of spatial processes, Lu and Chen (2004), Hallin et al. (2004b), Gao et al. (2006), Lin et al. (2009), and Robinson (2011) in the context of spatial regressions, and Phillips (2001), Conley and Topa (2002) and Crespo Cuaresma and Feldkircher (2013) for interesting applications.

In this study, we present several density estimators that rely on the Kernel Density Estimator (KDE) for flexibility and employ the method of Empirical Likelihood (EL) to incorporate spatial smoothing into the KDE. The empirical likelihood approach is a non-
parametric likelihood based method of estimation and inference (Owen, 1988, 2001). Chen (1997) introduced the Empirical Likelihood Kernel (ELK) density estimator that incorporates out of sample information via EL observation weights. These weights are obtained through maximizing the empirical likelihood subject to moment conditions implied by out of sample information. He showed that the ELK reduces the asymptotic variance of the KDE. Moreover, the optimal KDE bandwidth and the empirical likelihood weights can be derived separately, adding to the practical appeal of the ELK.

Chen’s (1997) empirical likelihood refinement of kernel-based estimators has been explored in subsequent studies, including Hall and Presnell (1999); Müller et al. (2005); Schick and Wefelmeyer (2009); Zhang (1997, 1998) in quantile or density estimations and Cai (2001); Zhang and Liu (2003); Qin and Tsao (2005) in nonparametric regressions. The current study contributes to this literature in the following ways. We generalize the empirical likelihood refinement of the standard KDE to the Adaptive Kernel Density Estimator (AKDE), wherein the degree of smoothness varies according to the (unknown) underlying density. We apply the ELK and ELAK to the estimation of spatially similar densities. We construct, for each density, a set of spatially smoothed moments. We then maximize the empirical likelihood subject to these auxiliary moment conditions and apply the implied EL observation weights to the KDE and AKDE, resulting in the spatially-smoothed ELK and ELAK.

Polynomial moments, such as mean, variance, skewness and kurtosis, are customarily used in the empirical likelihood adjustment to kernel estimations. However, high order moments are known to be susceptible to possible outliers, especially when sample sizes are small as in the current study. We advocate instead spline based moments in this study. This approach offers both robustness and flexibility. Typically spline functions, for instance low order truncated power series, are less sensitive to outliers than high order polynomials. In addition, we can tailor the configuration of the spline knots according to the objective of investigation. For instance, when the focus is on the lower tail of distributions (as in financial risk management and insurance), we can place more knots on the lower part of distributions to arrive at more accurate estimates in that region. Lastly, our numerical experiments suggest that the maximization of empirical likelihood converges readily under spline based moments, but less so under polynomial moments.
To explore the finite sample performance of the proposed estimators for spatially similar densities, we conduct several Monte Carlo simulations. The experiments are designed to resemble the actual corn yield distributions of all counties in Iowa, the largest corn producing state in the United States. We compare the spatially-smoothed ELK and ELAK with the conventional KDE and AKDE according to the estimation of overall density and crop insurance premium rates, which play a crucial role in the design of actuarially fair insurance programs. In all cases, the ELK and ELAK outperform their conventional counterparts, often by considerable margins. Lastly we apply these methods to the prediction of Iowa county corn yield distributions and their corresponding crop insurance premiums.

The remaining text proceeds as follows. Section 2 introduces the ELK and ELAK estimators. Section 3 describes the construction of spatially smoothed moment conditions. Section 4 presents the Monte Carlo simulations, followed by an empirical application in Section 5. The last section concludes.

2 Empirical likelihood kernel density estimation

Let $X_1, X_2, \ldots, X_n$ denote an I.I.D. sample from a distribution with density $f$. The standard Kernel Density Estimator (KDE), for a given point $x$, is defined as

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

(2.1)

where $K_h(\cdot) = K(\cdot/h)/h$ and the kernel function $K$ usually takes the form of some symmetric and uni-modal density function defined on $\mathbb{R}$ or some finite interval. The bandwidth $h$ controls the degree of smoothness of the KDE. For general treatments of the KDE, see, e.g., Wand and Jones (1995) and Li and Racine (2007).

Oftentimes out-of-sample information about $f$ is available in economic and statistical applications. The KDE, known for its flexibility, can conceivably benefit from incorporating auxiliary information. There exists a body of literature on kernel density estimation under moment or quantile constraints as well as shape restrictions such as monotonicity, unimodality, concavity, etc; see, e.g., Bickel and Fan (1996), Cheng et al. (1999), Eloyan and Ghosh
In this paper, we consider a general framework of incorporating out-of-sample information in the form of moment constraints. Let \( g = (g_1, ..., g_q)^\top \) be a \( q \)-dimensional vector of real valued, bounded and linearly independent functions with expectation

\[
\mathbb{E}\{g(X_i)\} = b. \tag{2.2}
\]

Chen (1997) introduced the Empirical Likelihood Kernel (ELK) density estimator to incorporate moment constraints in the KDE. Note that the KDE (2.1) places a uniform weight \( 1/n \) on each observation. The ELK, instead, applies a general weight \( p = (p_1, ..., p_n)^\top \) to the KDE, yielding

\[
\hat{f}_{el}(x) = \sum_{i=1}^{n} p_i K_h(x - X_i), \tag{2.3}
\]

where \( p_i \geq 0 \) and \( \sum_{i=1}^{n} p_i = 1 \). The observation weights \( p \) are obtained by maximizing the empirical likelihood subject to (2.2) and thus reflect auxiliary distributional knowledge.

The empirical likelihood approach, originally proposed by Owen (1988, 1990), is a non-parametric likelihood based method of estimation and inference. It chooses the sample probability weight \( p_i \) for each observation \( X_i \) according to the following maximization problem:

\[
\max_{(p_1, ..., p_n)} \prod_{i=1}^{n} np_i \\
\text{subject to} \sum_{i=1}^{n} p_i g(X_i) = b, \tag{2.4}
\]

\[0 \leq p_i \leq 1, \quad i = 1, ..., n,
\]

\[\sum_{i=1}^{n} p_i = 1.
\]

The objective function \( \prod_{i=1}^{n} np_i \), the so called empirical likelihood, is a nonparametric version of likelihood ratio. For notational conciseness, we define \( J(X_i) = g(X_i) - b \) and let the Lagrangian multipliers associated with the constraints \( \sum_{i=1}^{n} p_i J(X_i) = 0 \) be \( \lambda = (\lambda_1, ..., \lambda_q)^\top \).
Then the solution to (2.4) takes the following form, for \( i = 1, \ldots, n \),

\[
p_i = n^{-1} \{ 1 + \lambda^\top J(X_i) \}^{-1},
\]

where \( \lambda \) is given by

\[
\sum_{i=1}^{n} \frac{J(X_i)}{1 + \lambda^\top J(X_i)} = 0.
\]

Remark 1. In the absence of auxiliary information, \( p_i \) reduces to \( 1/n \) for all observations. Similarly, if only sample information is used as moment constraints, i.e. if we set \( b = \frac{1}{n} \sum_{i=1}^{n} g(X_i) \) in (2.4), the same solution of uniform weights results. Intuitively, since the sample moment conditions are satisfied by the sample automatically, the Lagrangian multipliers are zero, yielding \( p_i = 1/n \) for each \( i = 1, \ldots, n \) according to (2.5). In either case, the ELK degenerates to the KDE.

The standard KDE is characterized by its uniform weight \( 1/n \) and a global bandwidth \( h \) for all observations. The ELK generalizes the KDE by allowing non-uniform weights. In this study, we consider a further generalization to the ELK by introducing non-constant bandwidths via the Adaptive Kernel Density Estimator (AKDE). For a density with varying degrees of smoothness over its support, a KDE with a global bandwidth is sometimes inadequate and might produce spurious bumps or other aberrants. The AKDE uses varying bandwidths across observations. Letting \( h = h(X_i) \) be a positive function of \( X_i \), one can construct a flexible KDE as follows:

\[
\hat{f}(x) = \frac{1}{n} \sum_{i=1}^{n} K_{h(X_i)}(x - X_i).
\]

Abramson (1982) proposed to set \( h(X_i) = h \nu(X_i) \), where \( \nu(\cdot) = f^{-1/2}(\cdot) \). It can be shown that under some regularity conditions, the AKDE reduces the order of bias to \( O(h^4) \) from the usual \( O(h^2) \). For detailed treatments of the AKDE and its extensions, see Hall and Marron (1988), Hall (1990), Jones (1990), Terrell and Scott (1992), and Hall et al. (1995). In practice the AKDE requires a pilot estimate of \( f(X_i) \) to calculate its observation specific bandwidth \( h(X_i) \). Silverman (1986) suggested using the KDE for this purpose as well and prescribed a detailed step-by-step implementation procedure. Abramson (1982), Silverman
(1986), and Jones and Signorini (1997) demonstrated good finite sample performance of the AKDE.

In order to exploit out of sample information, we replace the uniform weight $1/n$ of the conventional AKDE with observation specific weights implied by the empirical likelihood. The resultant Empirical Likelihood Adaptive Kernel (ELAK) density estimator takes the form

$$
\hat{f}_{el}(x) = \sum_{i=1}^{n} p_i K_{h(X_i)}(x - X_i),
$$

where the empirical likelihood observation weights $p_i$'s are obtained according to (2.4).

The asymptotic properties of the ELK have been established by Chen (1997). In particular, we have

$$
bias\left\{ \hat{f}_{el}(x) \right\} = bias\left\{ \hat{f}(x) \right\} + o(n^{-1}), \quad (2.9)
$$

and

$$
var\left\{ \hat{f}_{el}(x) \right\} = var\left\{ \hat{f}(x) \right\} - \mathbf{J}(x)^\top \Sigma^{-1} \mathbf{J}(x)f^2(x)n^{-1} + o(n^{-1}), \quad (2.10)
$$

where $\Sigma$ is the covariance matrix of $\mathbf{J}(X_i)$ such that $\Sigma_{l,m} = \text{cov}(J_l(X_i), J_m(X_i)), 1 \leq l, m \leq q$. Analogously, these results can be shown to hold for the ELAK as well.

**Remark 2.** The results above suggest that the difference in bias is of order $o(n^{-1})$ between the KDE and the ELK, which is asymptotically negligible. On the other hand, there is an order $O(n^{-1})$ reduction in the variance of the ELK relative to that of the KDE as suggested by (2.10), as the second term on the right hand side is clearly negative. These results are consistent with the general belief that the empirical likelihood reduces estimation variation, although in the current case this reduction only occurs in the small order term due to nonparametric smoothing. Nonetheless as pointed out by Chen (1997), the extent of this reduction can be substantial when sample size is relatively small.

**Remark 3.** An important practical implication of the above results is that one can use the optimal data-driven bandwidths for the KDE and AKDE for their empirical likelihood adjusted counterparts, for which those bandwidths remain asymptotically optimal up to order $O(n^{-1})$. Therefore, one can select the optimal bandwidth and calculate the empirical likelihood weights separately in the construction of empirical likelihood adjusted estimates as given by
(2.3) or (2.8).

3 Spatially smoothed moment constraints

Denote by $\mathcal{L}$ a collection of geographic locations. Suppose that for each location $l \in \mathcal{L}$, we observe an I.I.D. random sample $X_i(l), i = 1, \ldots, n$, generated by an unknown distribution with density $f_l$. Our goal is to seek flexible estimators of $f_l$'s using a kernel type estimator such as the KDE and AKDE discussed in the previous section. In the presence of small sample sizes, however, separate estimation of individual densities via nonparametric methods may not be satisfactory. Nonetheless, if spatially proximate densities in the collection tend to resemble one another, we might be able to improve the estimates via spatial smoothing of the kernel estimates. In this section we present an approach of incorporating spatial information in kernel density estimations through the empirical likelihood. In particular, our strategy is to incorporate spatial information as moment constraints to kernel density estimates via the ELK and ELAK described above. This approach entails two modeling components: the specification of the moment functions $g$ and their estimation based on spatial smoothing, which are discussed below.

**Specification of moment functions**

The ELK and ELAK can be viewed as information-theoretic refinement of their corresponding counterparts with uniform observation weights. The refinement via Empirical Likelihood (EL) observation weights are remarkably flexible since no restrictions are imposed on the specification of the moment functions $g$ other than finite expectation and linear independence. Nonetheless, we advocate the following general guidelines in the configurations of $g$.

First, it is well known that a smooth density can be approximated arbitrarily well by a series type estimator. Natural candidates for $g$ are basis functions of series estimators, such as the power series, trigonometric series and splines. Despite the popularity of the power series in the EL adjustment of kernel estimations, their high order moments may be sensitive to possible outliers, especially when sample sizes are small. Instead we focus on the spline
basis functions for their noted flexibility and robustness. For instance, the commonly used s-degree truncated power series take the form

$$g(x) = (1, x, x^2, \cdots, x^s, (x - \tau_1)_+^s, \cdots, (x - \tau_m)_+^s),$$

where $$(x)_+ = \max(x, 0)$$, and $$\tau_1 < \cdots < \tau_m$$ are spline knots. Due to its piecewise nature, splines are more flexible than power series. At the same time, since typical spline estimations employ low order piecewise power series, they are less sensitive to possible outliers and do not require the existence of high order moments as the power series do. Lastly we note that the maximization of empirical likelihood is a difficult nonlinear optimization problem. Our numerical experiments indicate that imposing power series moment conditions, even with moderate orders no greater than $$\mathbb{E}[X^4]$$, sometimes impedes the convergence of the empirical likelihood maximization. In contrast, moment conditions given by the commonly used linear, quadratic or cubic spline basis functions appear to be immune from this numerical difficulty.

Second, we recognize that the need of refinement of the kernel density estimates, in the current case, arises from the small sample sizes of individual densities. Although the finite sample properties of the KDE and AKDE might not be satisfactory, they nonetheless provide reasonable initial estimates. Consequently a moderate number of moments usually suffice to provide the desired improvement.

Third, we stress that the moment functions $$g$$ can be tailored to suit the need of investigation. For instance, the crop insurance programs are primarily concerned with the lower part of crop yield distributions. Correspondingly in the configuration of spline basis functions, we can place more knots in the lower part of the distribution support. Our simulations below illustrate the merit of this customized knot placement in the estimation of crop yield distributions and calculation of crop insurance premium rates.

**Moment estimation via spatial smoothing**

Equipped with moment functions $$g$$, we proceed to calculate the sample analog of $$\mathbb{E}[g]$$ for individual densities based on their own observations and spatial neighbors. Denote by $$\mathcal{F}$$ the collection of densities $$\{f_l : l \in \mathcal{L}\}$$. Let $$s_l$$ be the coordinate vector of location $$l$$ in $$\mathcal{L}$$. For instance, it is common that $$s_l$$ consists of longitude and latitude. Denote by
\(d(l, l')\) the distance between two locations \(l\) and \(l'\) in \(\mathcal{L}\); for example, the Euclidean distance \(d(l, l') = \|s_l - s_{l'}\|_2\) or Manhattan distance \(d(l, l') = \|s_l - s_{l'}\|_1\).

For each location \(l \in \mathcal{L}\), denote by \(\bar{b}_l = \frac{1}{n} \sum_{i=1}^{n} g(X_i(l))\) the sample analog of \(g\) based on observations from the \(l\)-th location. Suppose that the elements in \(\mathcal{F}\) resemble their neighbors and the degree of resemblance gradually declines with spatial distance. We consider the following estimator of spatially smoothed moments,

\[
\hat{b}_l = \sum_{l' \in \mathcal{L}} \omega(l, l') \bar{b}_{l'}. \tag{3.1}
\]

The spatial weight \(\omega(l, l')\), as a function of spatial distance \(d(l, l')\), is constructed as

\[
\omega(l, l') = \frac{K \left( \frac{d(l, l')}{\theta} \right)}{\sum_{l'' \in \mathcal{L}} K \left( \frac{d(l, l'')}{\theta} \right)}, \tag{3.2}
\]

where \(K\) is a univariate kernel function, for instance the Gaussian kernel, and \(\theta > 0\) is a spatial smoothing parameter. Since the weights \(\omega(l, l')\)'s are normalized such that they add up to unity, \(\hat{b}_l\) is a weighted average of own-sample averages \(\bar{b}_{l'}\)'s of all units in \(\mathcal{L}\). The farther away is location \(l'\) from \(l\), the smaller is its weighted in the spatially smoothed moment of location \(l\); the smoothing parameter \(\theta\) governs the decaying rate of spatial discounting.

We note that (3.1) together with (3.2) take the form of Nadaraya-Watson kernel estimator of the regression equation \(g(X_i(l)) = b(s_l) + e_{i,l}\), where \(b(s) = \mathbb{E}[g(X_i)|s]\) and \(e_{i,l}\) is an additive error term with mean zero and finite variance. Thus the spatially smoothed moments are essentially nonparametric kernel estimates of the location-specific conditional mean \(\mathbb{E}[g(X)|s = s_l]\) given the full sample \(\{X_i(l)\}_{i=1}^{n}, l \in \mathcal{L}\).

It is well-known that the smoothing parameter plays a crucial role in kernel estimations. Here we present a data-driven method motivated by cross validation in kernel regressions. We consider the following least-square cross validation criterion

\[
CV(\theta) = \sum_{l \in \mathcal{L}} \|\hat{b}_l - \hat{b}_{-l}\|^2, \tag{3.3}
\]

where \(\| \cdot \|\) denotes the Euclidean norm, \(\hat{b}_l\) is the own-sample average at location \(l\), and \(\hat{b}_{-l}\)
is a leave-one-location-out version of the spatially smoothed \( \hat{b}_l \), which is given by

\[
\hat{b}_{-l} = \sum_{l' \in \mathcal{L}, l' \neq l} \omega^*(l, l') \hat{b}_{l'}
\]

with

\[
\omega^*(l, l') = \frac{K \left( \frac{d(l, l')}{\theta} \right)}{\sum_{l'' \in \mathcal{L}, l'' \neq l} K \left( \frac{d(l, l'')}{\theta} \right)}, \ l' \neq l.
\]

It is seen that \( \hat{b}_{-l} \) can be viewed as a prediction of \( \mathbb{E}[g(X)] \) at location \( l \) based on observations from all other locations in \( \mathcal{L} \). The optimal \( \theta \) is selected as \( \theta_{opt} = \arg \min_{\theta > 0} CV(\theta) \).

4 Simulations

To explore the performance of the proposed methods and possible benefits of spatial smoothing in kernel density estimations, we conduct a number of Monte Carlo simulations. Rather than generating an ad hoc set of densities, we base our design on a real world situation. In particular, we estimate parametrically the historical corn yield distributions of 99 Iowa counties, and generate random samples from the estimated densities for our simulations. This design allows us to mimic not only the real world distribution of crop yields, but more importantly the spatial similarity in yield distributions shared by counties that are spatially proximate.

Simulation design

We focus on corn yield distribution of Iowa because corn is a major commodity of U.S. agriculture and Iowa is the largest corn producing state. In our experiment, \( \mathcal{L} \) consists of the 99 counties in Iowa. Corn yield data from 1957 through 2010 for each county in Iowa, obtained from the National Agricultural Statistics Service, are used in the estimation. Denote by \( Y_t(l) \) the average corn yield of county \( l \) at time \( t = 1, 2, ..., T \), where \( T = 54 \) in this example.

Apparently crop yield data are not I.I.D. over time. To remove the influence of technology advance, the crop yield literature has customarily employed the following model
\[ Y_t(l) = m_t(l) + e_t(l), \]  

(4.1)

where \( m_t(l) \) is a trend function of county \( l \) and \( e_t(l), t = 1, 2, ..., T \) are residuals with mean zero and finite variance. To avoid rigid functional form assumptions, we follow a common practice and use the LOESS method to estimate the trend. Details of the LOESS method are provided in the next section. Some existing studies suggested that the errors may violate the Homoskedasticity (HOM) assumption and considered instead the Constant Coefficient of Variation (CCV) specification, wherein the standard deviation of \( e_t(l) \) varies proportionally with \( m_t(l) \) such that \( \mathbb{E}[e_t^2(l)] = \sigma^2 m_t^2(l) \), where \( \sigma < \infty \). For detailed discussions of crop yield distributions, see, e.g., Claassen and Just (2011), Harri et al. (2011) and references therein.

In our simulations, we consider both the HOM and CCV specifications. Our estimation is based on the de-trended and studentized (as is needed under CCV) residuals given by

\[ X_t(l) = \begin{cases} 
\hat{e}_t(l), & \text{if HOM;} \\
\frac{\hat{e}_t(l)}{\hat{m}_t(l)}, & \text{if CCV.} 
\end{cases} \]  

(4.2)

Note that for \( s \)-step-ahead predictive density estimation under CCV, we shall also multiply \( X_t(l) \) by \( \hat{m}_{T+s}(l) \).

For each residual series \( X_t(l), t = 1, \ldots, T \), we estimate a skewed normal distribution, whose density is given by

\[ f_l(x) = \frac{2}{\sigma_l} \phi \left( \frac{x - \mu_l}{\sigma_l} \right) \Phi \left( \alpha_l \left( \frac{x - \mu_l}{\sigma_l} \right) \right), \]  

(4.3)

where \( \phi \) and \( \Phi \) are the PDF and CDF of the standard normal distribution, and \( \mu_l, \sigma_l \) and \( \alpha_l \) are the location, scale and skewness parameters respectively. We stress that the skewed normal distribution is chosen mainly to accommodate the stylized fact that residuals from crop yield regressions are usually skewed. This choice does not imply that the true underlying crop yield distributions are skewed normal and more importantly it does not compromise the validity of the subsequent simulations.
Verification of spatial similarity among yield distributions

Equipped with the estimated skew-normal densities $\hat{f}_l, l \in \mathcal{L}$, we can then generate random samples from these densities to conduct simulations. To justify the spatial smoothing advocated in this study, we start our analysis with verifying the spatial similarity among $\hat{f}_l$’s. Let $d(l, l')$ be the Euclidean distance between two counties $l$ and $l'$ based on their longitudes and latitudes, and $\hat{h}(l, l') \equiv \left[ \int \{ \hat{f}^{\frac{1}{2}}_l(x) - \hat{f}^{\frac{1}{2}}_{l'}(x) \}^2 dx \right]^{\frac{1}{2}}$ the Hellinger distance between $\hat{f}_l$ and $\hat{f}_{l'}$. We consider the following simple linear regression

$$\hat{h}(l, l') = \beta_0 + \beta_1 d(l, l') + \epsilon(l, l'), \quad (l, l') \in \mathcal{L}, \quad (4.4)$$

where $\epsilon(l, l')$ is a random error term. A significantly positive $\beta_1$ should indicate spatial similarity among the densities. Table 1 reports the estimation results of model (4.4) under both the HOM and CCV specifications. The estimated coefficient $\hat{\beta}_1$ in all models are positive and statistically highly significant, suggesting a high degree of similarity among neighboring distributions and thus lending support to the proposed spatial smoothing of densities. We have considered alternative discrepancy measures other than the Hellinger distance; in additional to model (4.4), we also experimented with the log-linear, log-log and more general Box-Cox models. All estimations strongly support the existence of spatial similarity among yield distributions.

Spatially smoothed moments

Having verified the spatial similarity among the densities, we proceed to conduct numerical simulations based on the estimated densities. For each estimated density $\hat{f}_l, l \in \mathcal{L}$, we generate random samples $X^*_i(l) : i = 1, ..., T_0$ and calculate the proposed empirical likelihood based kernel density estimators, the ELK and ELAK. For comparison, we also estimate the densities using the conventional KDE and AKDE. We consider two sample sizes $T_0 = 30$ and 50, which are typical in crop yield studies. We repeat each experiment 500 times.

One advantage of the empirical likelihood based kernel estimators is that the selection of kernel bandwidth and the calculation of empirical likelihood weights can be conducted separately. For each estimator, we select the bandwidth $h$ according to Silverman’s rule of
thumb. We experimented with alternative bandwidth selection rules, the overall results are virtually identical to those from Silverman’s rule and hence not reported.

Implementations of the empirical likelihood based estimators require the specification of moment functions and calculation of spatially smoothed moments. We consider two spline-based specifications of the moment functions. For each density, denote by $\hat{Q}_p$, its $p$-th sample quantile, $0 < p < 100$. The first specification is a linear spline basis given by

$$g(x) = (x, (x - \hat{Q}_{20})_+, (x - \hat{Q}_{40})_+, (x - \hat{Q}_{60})_+),$$

and the second is a quadratic one

$$g(x) = (x, x^2, (x - \hat{Q}_{33})_+^2, (x - \hat{Q}_{67})_+^2).$$

Some remarks are in order. We use these two simple specifications to illustrate the flexibility of the proposed method. In particular, the linear spline basis features an uneven placement of knots, emphasizing the lower part of the distribution. This is motivated by that farm risk management and crop insurance are primarily concerned with the lower part of crop yield distributions. The second specification is a quadratic spline with evenly spaced knots. We restrict the number of basis functions to be the same in these two specifications to facilitate their comparison. Generalizations in terms of alternative polynomial degrees, knot placements and types of spline basis are readily available.

The accuracy of estimated moments is crucial for subsequent empirical likelihood based estimation. We therefore first inspect the quality of the spatially smoothed moments calculated according to (3.1) and (3.2). Given a set of moment functions $g$, let $b_l = \int g(x) \hat{f}_l(x) dx$ for location $l$. We consider estimating $b_l$ by (i) the proposed spatially smoothed moments which are based on the entire simulated sample $X_i^* (l) : i = 1, \ldots, T_0, l \in L$ and (ii) the simple own-sample average $\frac{1}{T_0} \sum_{i=1}^{T_0} X_i^*(l)$. We compare the performances of these two moment estimates according to the total weighted mean squared error $\sum_{l \in L} \mathbb{E}(\hat{b}_l - b_l)\Omega^{-1}(\hat{b}_l - b_l)$, where $\hat{b}_l$ is a generic estimate of $b_l$ and $\Omega$ is the covariance matrix $\int (g(x) - b_l)(g(x) - b_l)^\top \hat{f}_l(x) dx$. We report in Table 2 the results for two specifications of $g$, i.e. the linear and quadratic spline moment functions, as proposed in previous paragraphs. The spatially
smoothed moments clearly dominate those based on own sample averages. Note that the KDE and AKDE can be seen as special cases of ELK and ELAK by setting moment estimates to their corresponding own-sample averages. We therefore expect that the proposed EL-weighted estimators should provide better performance thanks to the incorporation of spatially smoothing moment conditions, which are shown to be more accurate than simple own sample averages.

**Estimation performance**

We next evaluate the performance of various density estimators according to two criteria. Let \( \hat{f}_l \) be a generic estimate of \( f_l, l \in \mathcal{L} \). We examine their global performance according to the total mean squared errors \( \sum_{l \in \mathcal{L}} \mathbb{E} \int \{ \hat{f}_l(x) - \hat{f}_l(x) \}^2 dx \). We calculate the total mean squared error numerically for each experiment and report their averages across all experiments. The results are reported in the third column of Table 3, where the ELK and ELAK estimators with the linear spline basis and quadratic spline basis are denoted by (A) and (B) respectively. We signify the estimator with the best performance within each experiment design with the bold font. The spatially adjusted estimators clearly outperform their non-adjusted counterparts, oftentimes by a considerable margin. In particular, the ELAK with linear spline moment functions provides the best overall performance in all experiments. The results underscore the merit of spatial smoothing of spatially similar densities, and demonstrate that the benefits of incorporating out-of-sample information in the KDE via the empirical likelihood extend to the adaptive kernel density estimations.

We next evaluate the usefulness of the proposed methods by examining how they fare in the estimation of crop insurance premium rates. Let \( m_t(l) \) be the predicted mean output of crop yield \( Y_t(l) \) for location \( l \) at time \( t \). A crop insurance policy allows a farmer to choose a coverage level, which typically ranges from 50 – 85% of the predicted output \( m_t(l) \). An actuarially fair premium rate (at the percentage level) for a policy with a coverage level \( c \) is given by

\[
R_l(c) = \Pr(Y_t(l) \leq cm_t(l)) \left\{ 1 - \frac{\mathbb{E}[Y_t(l) | Y_t(l) \leq cm_t(l)]}{cm_t(l)} \right\} \times 100, \ c \in (0, 1).
\]

Accurate prediction of the crop insurance premiums plays a crucial role in the federal crop
insurance program. Therefore we also evaluate the various estimators in terms of their accuracy in premium estimations. In particular, we consider four commonly chosen coverage levels \(c = 70\%, 75\%, 80\%\) and \(85\%\) and evaluate the prediction of the premium rates for the last year of the sample, based on observations prior to that year. We gauge the performance using the total mean squared error \(\sum_{t \in \mathcal{L}} \mathbb{E}\{\tilde{R}_t(c) - R_t(c)\}^2\) across all regions, where \(\tilde{R}_t(c)\) is the estimated premium with coverage \(c\) based on a generic estimate \(\tilde{f}_t\).

The last four columns of Table 3 report the results of premium estimation. Similar to the comparison of overall density estimation, the EL-based estimators dominate their non-adjusted counterparts, this time by even larger margins. In most cases, the EL estimator with linear spline moment functions provides the best performance, but the difference between the two moment specifications are generally rather small. Although the ELAK provides the best global performance as discussed above, the ELK tend to outperform the ELAK in the estimation of premiums, especially when the coverage is low.

In sum, our simulations suggest that the proposed EL-weighted kernel density estimators outperform their non-weighted estimators in the estimation of spatially similar densities. The ELAK tends to provide the best overall performance, while the ELK excels in the estimation of crop insurance premiums, which emphasize the lower part of the distributions. These overall patterns appear to be persistent across different sample sizes, specifications of EL moment conditions, and heteroskedasticity specifications.

5 Empirical application

We apply the proposed methods to the prediction of Iowa corn yield distributions and crop insurance premiums. The crop insurance program has been the centerpiece of the U.S. agriculture safety net. According to the Risk Management Agency of the U.S. Department of Agriculture, the estimated government cost for the federal crop insurance programs were about $15.8 billion for crop year 2012. Given its massive size and the substantial moral hazard and adverse selection that were reported to have plagued this program, accurate calculation of crop insurance premiums are of crucial importance to promote the fairness and efficiency of this program.

In this empirical investigation, we estimate corn yield distributions of Iowa counties
using observations 1957–2010, i.e. in total we have $T = 54$ observations from each county. The estimates are then used to predict corn yield distributions of year 2012 and associated crop insurance premiums. Two-year-ahead prediction is the standard practice in the crop insurance industry since modern farming and farm risk management require a considerable amount of planning time. We stress that unlike the simulations in the previous section where the data are generated from synthetic parametric distributions, here we base our estimations on historical crop yields.

We first remove the time trend for each county according to the model (4.1). The trend function $m_t(l)$ is estimated by the LOESS procedure (Cleveland, 1979; Cleveland and Devlin, 1988). Specifically, at each evaluation time point $t_0$ whose time trend value needs to be estimated, we locally fit a linear polynomial to $\alpha$ proportion of the data with the corresponding time variable values being closest to $t_0$. Each local fit is conducted by weighted least squares, placing more weights to data points near $t_0$ but less weights to points further away. The usual tri-cube function is used to construct the weights. The parameter $\alpha$ controls the degree of smoothing and it is selected by the corrected Akaike information criterion (Hurvich et al., 1998) or generalized cross validation. We note that the estimated yield trend curves during the sample periods are nearly linear, with varying slopes across counties.

We then conduct a regression based heteroscedasticity test as in Harri et al. (2011) to the residuals $\hat{e}_t(l): t = 1,\ldots,T$ for each county. The homoscedasticity assumption is rejected, at the 5% significance level, for 42 counties; in contrast, the constant coefficient of variation assumption is rejected for only 5 counties. We therefore adopt the constant coefficient of variation specification and adjust the residuals accordingly for the subsequent estimations. That is, for each $t = 1,\ldots,T$ and county $l$, we set $X_t(l) = \hat{e}_t(l) \frac{\hat{m}_{T+2}(l)}{\hat{m}_t(l)}$ where $\hat{m}_{T+2}(l)$ is a two-step-ahead forecast of yield trend from the model (4.1). Equipped with thus constructed studentized residuals $\{X_t(l): t = 1,\ldots,T, l \in L\}$, we again consider four estimators: the KDE, ELK, AKDE and ELAK, for each density of $X(l)$. We use the linear spline moment conditions described in the previous section for the ELK and the ELAK estimators. Denote by $\hat{f}_l$ one of the four estimators. The resultant predictive yield density of the county $l$, evaluated at a point $x$, is then given by $\hat{f}_l(x - \hat{m}_{T+2}(l))$. 

16
We calculate the premium rates of each county for Year 2012 based on the estimated predictive densities and present the results at four coverage levels, i.e. 70\%, 75\%, 80\% and 85\%. We report some summary statistics, across all counties, of the estimated premiums in Table 4. Comparison between non-weighted estimates and EL-adjusted estimates suggest that there is little difference in the central tendency (in terms of mean and median) between these two types of estimators. On the other hand, the EL-adjusted estimates show smaller variations (in terms of variance and inter-quantile-range) than their unweighted counterparts. This is consistent with the theoretical property that the EL-adjusted estimates are largely comparable with their unweighted counterparts in terms of bias but enjoy smaller variance. Intuitively, it also reflects that spatial smoothing of spatially similar densities reduces the sampling variation of functionals of estimated densities, especially when the sample sizes are small.

6 Concluding remarks

We have proposed estimators that are suitable to the estimation of spatially similar densities, each with a relatively small number of observations. Since nonparametric estimation of individual densities may be subject to considerable sampling variations, we elect to pool information from spatially proximate densities in order to improve efficiency. This is made possible by the method of empirical likelihood subject to spatially smoothed moment conditions for each density. In our investigation of crop yield distributions, a natural distance measure is readily available and the spatial similarity arises from common environmental and climate factors, weather fluctuations, farming practice, etc. More generally in some social economic investigations, the spatial similarity may be attributed to social-economic distance as is often explored in the studies of social network. We expect our methods can find useful applications in those situations as well.

We conclude this study by noting that the empirical likelihood is a member of the general empirical likelihood family, which is characterized by the minimization of certain Cressie-Read Power discrepancy criterion subject to some moment conditions. Other noted members of this family includes the exponential tilting likelihood and the Euclidean likelihood. During the course of this study, we have conducted extensive experiments with these alternative
methods and found that the overall results are essentially identical to those based on the empirical likelihood. This is not unexpected given the well known asymptotic equivalence among the members of the general empirical likelihood estimators.

References


Table 1: Estimation results of models (4.4)

<table>
<thead>
<tr>
<th></th>
<th>$\hat{\beta}_0$</th>
<th>$\hat{\beta}_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOM</td>
<td>0.0857* (0.0021)</td>
<td>0.0187* (0.0008)</td>
</tr>
<tr>
<td>CCV</td>
<td>0.0892* (0.0019)</td>
<td>0.0144* (0.0007)</td>
</tr>
</tbody>
</table>

The estimated standard errors are in parentheses; a star sign denotes the significance level $< 0.001$. 

23
Table 2: Mean squares error the estimated moments (HOM and CCV refer to homoskedastic and heteroskedastic errors)

<table>
<thead>
<tr>
<th>Panel 1: $T_0=30$</th>
<th>linear spline moment functions</th>
<th>quadratic spline moment functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOM own-sample average</td>
<td>13.19</td>
<td>13.24</td>
</tr>
<tr>
<td>spatially smoothed</td>
<td>3.69</td>
<td>3.79</td>
</tr>
<tr>
<td>CCV own-sample average</td>
<td>13.10</td>
<td>13.26</td>
</tr>
<tr>
<td>spatially smoothed</td>
<td>4.01</td>
<td>4.09</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel 2: $T_0=50$</th>
<th>linear spline moment functions</th>
<th>quadratic spline moment functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>HOM own-sample average</td>
<td>7.92</td>
<td>7.98</td>
</tr>
<tr>
<td>spatially smoothed</td>
<td>2.89</td>
<td>2.93</td>
</tr>
<tr>
<td>CCV own-sample average</td>
<td>7.92</td>
<td>7.91</td>
</tr>
<tr>
<td>spatially smoothed</td>
<td>3.12</td>
<td>3.18</td>
</tr>
</tbody>
</table>
Table 3: Simulation results (HOM and CCV refer to homoskedastic and heteroskedastic errors; EL estimators with linear and quadratic spline moment functions are denoted by (A) and (B) respectively)

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Global performance</th>
<th>Crop insurance premium estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>70%</td>
</tr>
<tr>
<td>Panel 1: $T_0=30$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HOM</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KDE</td>
<td>0.1029</td>
<td>8.96</td>
</tr>
<tr>
<td>ELK (A)</td>
<td>0.0672</td>
<td>2.53</td>
</tr>
<tr>
<td>ELK (B)</td>
<td>0.0812</td>
<td><strong>2.19</strong></td>
</tr>
<tr>
<td>AKDE</td>
<td>0.1221</td>
<td>11.81</td>
</tr>
<tr>
<td>ELAK (A)</td>
<td><strong>0.0561</strong></td>
<td>4.57</td>
</tr>
<tr>
<td>ELAK (B)</td>
<td>0.0794</td>
<td>4.47</td>
</tr>
<tr>
<td><strong>CCV</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KDE</td>
<td>0.0697</td>
<td>82.32</td>
</tr>
<tr>
<td>ELK (A)</td>
<td>0.0481</td>
<td><strong>25.21</strong></td>
</tr>
<tr>
<td>ELK (B)</td>
<td>0.0575</td>
<td>26.79</td>
</tr>
<tr>
<td>AKDE</td>
<td>0.0823</td>
<td>88.66</td>
</tr>
<tr>
<td>ELAK (A)</td>
<td><strong>0.0387</strong></td>
<td>35.80</td>
</tr>
<tr>
<td>ELAK (B)</td>
<td>0.0550</td>
<td>41.18</td>
</tr>
<tr>
<td>Panel 2: $T_0=50$</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>HOM</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KDE</td>
<td>0.0674</td>
<td>5.33</td>
</tr>
<tr>
<td>ELK (A)</td>
<td>0.0468</td>
<td>1.89</td>
</tr>
<tr>
<td>ELK (B)</td>
<td>0.0560</td>
<td><strong>1.45</strong></td>
</tr>
<tr>
<td>AKDE</td>
<td>0.0790</td>
<td>7.83</td>
</tr>
<tr>
<td>ELAK (A)</td>
<td><strong>0.0405</strong></td>
<td>3.82</td>
</tr>
<tr>
<td>ELAK (B)</td>
<td>0.0556</td>
<td>3.67</td>
</tr>
<tr>
<td><strong>CCV</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KDE</td>
<td>0.0460</td>
<td>49.83</td>
</tr>
<tr>
<td>ELK (A)</td>
<td>0.0334</td>
<td><strong>18.67</strong></td>
</tr>
<tr>
<td>ELK (B)</td>
<td>0.0403</td>
<td>19.50</td>
</tr>
<tr>
<td>AKDE</td>
<td>0.0535</td>
<td>54.44</td>
</tr>
<tr>
<td>ELAK (A)</td>
<td><strong>0.0278</strong></td>
<td>27.92</td>
</tr>
<tr>
<td>ELAK (B)</td>
<td>0.0388</td>
<td>31.32</td>
</tr>
</tbody>
</table>
Table 4: Summary statistics of the estimated premium rates

<table>
<thead>
<tr>
<th>Coverage</th>
<th>mean</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>70%</td>
<td>75%</td>
<td>80%</td>
<td>85%</td>
<td>70%</td>
<td>75%</td>
<td>80%</td>
</tr>
<tr>
<td>KDE</td>
<td>1.4152</td>
<td>1.8359</td>
<td>2.3655</td>
<td>3.0704</td>
<td>1.0282</td>
<td>1.4188</td>
<td>1.9455</td>
</tr>
<tr>
<td>ELK</td>
<td>1.4028</td>
<td>1.8235</td>
<td>2.3536</td>
<td>3.0598</td>
<td>1.0248</td>
<td>1.4082</td>
<td>1.9534</td>
</tr>
<tr>
<td>AKDE</td>
<td>1.6366</td>
<td>2.0248</td>
<td>2.5113</td>
<td>3.1501</td>
<td>1.2767</td>
<td>1.6422</td>
<td>2.1128</td>
</tr>
<tr>
<td>ELAK</td>
<td>1.6277</td>
<td>2.0155</td>
<td>2.5020</td>
<td>3.1414</td>
<td>1.2623</td>
<td>1.6336</td>
<td>2.0921</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coverage</th>
<th>variance</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>70%</td>
<td>75%</td>
<td>80%</td>
<td>85%</td>
<td>70%</td>
<td>75%</td>
<td>80%</td>
</tr>
<tr>
<td>KDE</td>
<td>1.1007</td>
<td>1.4516</td>
<td>1.8648</td>
<td>2.3293</td>
<td>1.1329</td>
<td>1.3175</td>
<td>1.4917</td>
</tr>
<tr>
<td>ELK</td>
<td>0.9469</td>
<td>1.2371</td>
<td>1.5841</td>
<td>1.9834</td>
<td>0.8877</td>
<td>1.0604</td>
<td>1.3011</td>
</tr>
<tr>
<td>AKDE</td>
<td>1.1464</td>
<td>1.4505</td>
<td>1.8078</td>
<td>2.2179</td>
<td>1.1354</td>
<td>1.2769</td>
<td>1.4538</td>
</tr>
<tr>
<td>ELAK</td>
<td>0.9775</td>
<td>1.2270</td>
<td>1.5246</td>
<td>1.8742</td>
<td>0.9207</td>
<td>1.0884</td>
<td>1.3092</td>
</tr>
</tbody>
</table>