Nonparametric Estimation of Crop Insurance Rates Revisited

Alan P. Ker and Barry K. Goodwin

With the crop insurance program becoming the cornerstone of U.S. agricultural policy, recovering accurate rates is of paramount interest. Lack of yield data presents, by far, the most fundamental obstacle to recovery of accurate rates. This article employs new methodology to estimate conditional yield densities and derive the insurance rates. In our application, we find the nonparametric kernel density estimator requires an additional twenty-six years of yield data to estimate the shape of the conditional yield densities as accurately as the recently developed empirical Bayes nonparametric kernel density estimator. Such methodological improvements can significantly aid in ameliorating the data problem.

Key words: empirical Bayes, insurance rating, kernel density estimation.

Recovering accurate premium rates is paramount to the actuarial soundness of the crop insurance program, a program which has gained prominence in the overall U.S. agricultural policy agenda. The 1996 Federal Agricultural Improvement and Reform (FAIR) Act signaled a new policy environment under which farmers would be subject to market forces. Under this policy regime, crop insurance remains, in some respects, one of the only government-subsidized, income stabilizing mechanisms available to agricultural producers. In addition, it appears that current political forces are in the process of fashioning crop insurance as the cornerstone of U.S. agricultural policy. Consider the 1999 State of the Union Address in which the President stated:

> As this congress knows very well, dropping prices and the loss of foreign markets have devastated too many family farmers. Last year, the Congress provided substantial assistance to help stave off a disaster in American agriculture, and I am ready to work with lawmakers of both parties to create a farm safety net that will include *crop insurance reform* and farm income assistance. (Italics added.)

A variety of crop insurance plans are currently available to farmers and a number of new pilot programs are under development. Standard crop yield insurance, termed Multiple Peril Crop Insurance (MPCI), pays an indemnity at a predetermined price to replace yield losses caused by any hazard. There exist three revenue insurance programs: Crop Revenue Coverage, Income Protection, and Revenue Assurance. These programs guarantee a minimum level of crop revenue and pay an indemnity if revenues fall beneath the guarantee. "group-risk" yield insurance, termed the Group Risk Plan (GRP), is based upon the county's yield. Insured farmers collect an indemnity whenever the county average yield falls beneath a yield guarantee, regardless of the farmers' actual yields. A variation of GRP which insures county revenues rather than yields has recently been developed and is termed Group Risk Income Plan. Over most of their existence, these all-risk crop insurance programs have been characterized by low participation and spotty actuarial performance. Consequently, the U.S. crop insurance program has been the subject of three General Accounting Office (GAO) investigations over the last five years.[1]

Alan P. Ker is an assistant professor in the Department of Agricultural and Resource Economics at the University of Arizona. Barry K. Goodwin is a professor in the Department of Agricultural Economics at North Carolina State University.

The authors thank the two anonymous referees for their insightful comments. This research was supported in part by the Federal Crop Insurance Corporation but the views expressed in this article are not necessarily endorsed by the sponsoring agency.

[1] See *Crop Insurance: Federal Program Faces Insurability and Design Problems*, GAO/RCED 93-98 (May 1993), *Crop Insurance: Opportunities Exist to Reduce Government Costs for*

For both yield and revenue insurance products, accurate estimation of conditional yield densities is crucial to minimize adverse selection. The premium rate for a contract that guarantees a percentage, denoted λ where $0 \leq \lambda \leq 1$, of the expected yield, denoted y^e, is given as

$$\text{(1)} \qquad \text{Premium Rate} = \frac{\mathcal{P}(Y < \lambda y^e)(\lambda y^e - E(Y | y < \lambda y^e))}{\lambda y^e}$$

where the expectation operator and probability measure are taken with respect to the conditional yield density $f(y|\mathcal{F}_t)$, and $\mathcal{F}_t$ is the minimal σ-algebra generated by the information known at the time of rating. If yield data were abundantly available for the given area of interest (county, farm, or farm unit), estimation of conditional yield densities would not be a particularly interesting problem. One could simply employ nonparametric kernel methods to recover a reasonably accurate estimate. Lack of historical yield data is, by far, the most fundamental obstacle in rating crop insurance contracts. Although yield data by county, farm, or farm unit tend to be extremely scarce, the number of counties, farms, or units is numerous. Presumably there is information in the extraneous yield data that may assist in estimating the conditional yield density for the particular area of interest. Unfortunately, the panel nature of the yield data is generally not exploited.[2]

Goodwin and Ker (henceforth GK) employed nonparametric kernel methods to estimate conditional yield densities and then subsequently derived GRP crop insurance rates. A concern with using nonparametric kernel methods or so-called flexible parametric forms in small samples is stability. Given this concern, GK combined innovations recovered from the temporal models of contiguous counties to estimate the conditional yield densities. We have two concerns with that approach. First, the method by which the innovations from the temporal models were combined is admittedly ad hoc. Second, innovations from contiguous counties are not likely to be independent.

In this article, a new estimator which circumvents these concerns is employed. Ker (1998) derived an empirical Bayes nonparametric kernel density estimator which uses empirical Bayes techniques on the estimated values from the kernel density estimators. This estimator may be used to exploit the panel nature of the yield data in recovering nonparametric estimates of the conditional yield densities. In addition to using this new estimator, this article makes strides in three other directions. First, we use a variable smoothing approach to estimate the conditional yield densities. Accurate estimation of lower tail probabilities is crucial to deriving accurate rates. A variable smoothing approach significantly decreases the dependency of estimated tail probabilities on the specific location of the tail realizations. Second, we restrict our estimated densities to have variance equal to the sample variance. The variance of the kernel density estimate is greater than the sample variance almost surely for an order two kernel. In small samples such as ours, the additional variance can be significant and have a profound effect on the derived rates. Third, we explicitly acknowledge that rates must be derived two years in advance. Despite popular belief, innovations may not in general be scaled up when accounting for a two period ahead forecast. This arises because the convolution may or may not be a member of the original family. We illustrate how the innovations need to be manipulated. To date, this has not been considered in the actual rating of GRP.[3] Finally, given the relative prominence of GRP with all-practice corn in Iowa, we apply the model by rating 1997 GRP Iowa all-practice corn contracts. Although we use the GRP program as our empirical case study, the proposed methods are applicable to other crop insurance products.

Spatio-temporal Process of Yields

Given that we wish to estimate the conditional yield densities for rating crop insurance

Private Sector Delivery, GAO/RCED 97-70 (April 1997), and *Crop Revenue Insurance: Problems With New Plans Need to Be Addressed*, GAO/RCED 98-111 (April 1998).

[2] See Ker (1996) for an approach that explicitly models the panel nature of the data.

[3] Recently, the temporal models used in the rating of GRP have changed from IMA(2,2) models to one-knot linear splines (Skees, Black, and Barnett). The existence of a lag structure necessitated explicit consideration of the two period versus one period prediction. This was erroneously ignored in the initial rating process. However, given that a lag structure is not accounted for in the newly adopted temporal models, one no longer needs to be concerned with the two period versus one period prediction since the conditional yield density is, under the null that the one-knot linear spline is the correct specification, invariant to a one-step or n-step ahead prediction.

contracts, this section considers the data generating process of mean yields. Yields follow a spatio-temporal process. By averaging over some spatial region (field, farm, or county) and conditioning on the temporal process, one recovers the conditional mean yield density for that given space at a point in time. The spatial region of interest for our analysis is the county as GRP is offered on a county basis. This section is separated into two parts: the spatial process of yields and the temporal process of yields.

Spatial Process of Yields

Consider that yields come from one of two distinct sub-populations: a catastrophic sub-population and a non-catastrophic sub-population. That is, in years when a catastrophic event occurs such as a drought, flood, freeze, etc., yields are drawn from the catastrophic sub-population. Conversely, in years when a catastrophic event does not occur, yields are drawn from the non-catastrophic sub-population. Thus, conditional yields may be thought of as a mixture of two unknown distributions where the secondary distribution (from catastrophic years) lives on the lower tail of the primary distribution (from non-catastrophic years) and has significantly less mass. The secondary distribution would be expected to have less mass because catastrophic events are realized with far less frequency than their complement. Also, the secondary distribution would be expected to live on the lower tail of the primary distribution because realized yields tend to be far less in catastrophic years. Given this basic structure, mean yields may have a unimodal symmetric density (mass of catastrophic distribution is negligible), a negatively skewed density (mass of catastrophic distribution is non-negligible and distribution is relatively flat), or a negatively skewed bi-modal density (mass of catastrophic distribution is non-negligible and distribution is relatively peaked).

In empirical applications, most researchers have used the beta distribution rather than a mixture distribution to accommodate negative skewness. The reader is directed to Babcock and Hennessy, Coble et al., Lee, Harwood, and Somwaru, Borges and Thurman, Kenkel, Busby, and Skees, Nelson, and Nelson and Preckel. Ker and Coble (1997) discourage the use of the beta distribution for modeling yields. Some researchers have assumed yields follow other distributional families. Gallagher used a gamma distribution while Moss and Shonkwiler used an inverse hyperbolic sine transformation to model yields. Ker (1996) used a mixture of two Gaussians.

Although these parametric families can accommodate negative skewness, this by no means indicates yields can be adequately approximated by them. The unknown yield distribution may or may not be uniquely defined by its first three population moments. Theoretical distributions may be constructed such that their resulting premium rates at low coverage levels differ by an order of magnitude despite having identical first three moments. Therefore, although these *so-called* flexible parametric forms can accommodate negative skewness, we have not seen any empirical evidence or statistical theory which justifies their use in modeling yields. A second concern with these parametric families is that they do not allow bi-modality. Central Limit Theorems (CLTs) for dependent processes suggest the possibility that yields at some aggregate level may be bi-modal. GK found evidence of bi-modality. Ker (1996) represents the only parametric empirical work which allows the conditional yield density to be bi-modal; yields were modeled using semi-nonparametric maximum likelihood methods (Hermite series expansions) with a mixture of two Gaussians for the innovations. Unfortunately, we can not be assured that counties contain sufficiently large land mass such that *conditional* spatial dependence dies off and a mixture of two Gaussians would result.[4]

Given the above problems associated with parametric forms, a viable alternative is nonparametric kernel density estimators. Nonparametric methods are extremely intuitive and relatively simple to use. For use of nonparametric kernel methods in the agricultural economics literature, see GK, Ker and Coble (1997), Turvey and Zhao, and Moschini.

Temporal Process of Yields

Past work by GK, Ker and Coble (1997), and Bessler modeled yields using IMA(1, q) processes ($q = 1, 2$). The IMA(1, 1) model is represented as follows:

$$(2) \quad y_t = y_{t-1} + \theta_0 + \theta_1 e_{t-1} + e_t.$$

[4] Conditional spatial dependence refers to the spatial dependence given which sub-population is realized.

This model has economic appeal in that it is a stochastic linear trend with autocorrelated innovations. The linear trend θ_0 represents technological advances while the moving average component θ_1 suggests that the underlying factors generating the sequence of innovations have effects on future yields. Consider, for example, a drought in time t. It is obvious that yields in time t will be affected by drought conditions in time t. It is also reasonable that the soil in time $t+1$, which influences yields in time $t+1$, would have leftover effects from the drought in time t. Because weather is not a conditioning variate, these effects are represented in the innovations. Consequently, the innovations in time t would influence yields in time $t+1$, thereby suggesting the existence of an MA component. The coefficient on the MA component, θ_1, should belong to the interval $(-1, 0)$. Recall, the temporal process of yields is essentially a submartingale (without the MA component) while the effect of some variates represented by the innovations should persist but die out in subsequent periods. Therefore, the process is both trend reverting and non-overshooting and thus $\theta_1 \in (-1, 0)$. This result was found by GK, Ker and Coble (1997), and Bessler.

A problem with estimating the IMA(1, 1) model is the need to employ non-linear least squares in small samples where convergence and parameter stability become issues. To address these problems, we use the error correction form of the model by replacing the MA(1) process with its AR(∞) representation. In doing so, the model is linear. An AR(4) process is found to be sufficient to represent the MA(1) process. The ARIMA(4, 1, 0) model is represented as follows:

$$(3)\quad y_t = y_{t-1} + \beta_0 + \beta_1(y_{t-1} - y_{t-2}) + \beta_2(y_{t-2} - y_{t-3}) + \beta_3(y_{t-3} - y_{t-4}) + \beta_4(y_{t-4} - y_{t-5}) + e_t.$$

National Agricultural Statistics Service county yield data for Iowa all-practice corn during the period 1957–95 are used. Heteroskedasticity is tested and corrected for as per GK.[5] Given the estimated model, a sequence of asymptotically independent and identically distributed realizations from $f(y|\mathcal{F}_T)$ is recovered in the following manner:

$$(4)\quad \hat{y}_{T+1,\,t} = \left(\frac{\hat{e}_t}{\hat{y}_t}\right) \times \hat{y}_{T+1} + \hat{y}_{T+1} \quad \forall t = 5, \ldots, T$$

where $\hat{e}_t$ is the estimated innovation in time t, $\hat{y}_t$ is the fitted value in time t, and $\hat{y}_{T+1}$ is the prediction for time $T+1$, all from equation (3).

Empirical Bayes Nonparametric Kernel Density Estimation

Ker (1998) derived an empirical Bayes nonparametric kernel density estimator which exploits possible similarities among the set of unknown densities that are to be estimated. Rather than specifying a prior over a function space, the estimator uses empirical Bayes techniques on the estimated values from the kernel density estimators. Certainly, the strengths of the estimator result from using kernel methods as the basis for the empirical Bayes. Given the empirical Bayes is undertaken pointwise across the support for the density, Ker (1998) illustrates that the empirical Bayes estimator is uniformly continuous under standard continuity assumptions regarding the kernel function. Ker (1998) shows that the empirical Bayes nonparametric kernel density estimator converges in probability at a rate of $O_p(T^{-(4/5)})$ to the standard kernel. As such, the proposed estimator inherits the asymptotic properties of the standard kernel estimator and attains the optimal rate $O_p(T^{-(2/5)})$. Given the empirical Bayes nonparametric kernel density estimator is asymptotically equivalent to the standard kernel estimator, Ker (1998) suggests it may be considered a *competing* estimator. Before outlining the empirical Bayes nonparametric kernel density estimator, the relevant technical literature is reviewed.

Nonparametric Kernel Methods

As previously mentioned, nonparametric kernel methods have been sparingly used in the agricultural economics literature. Moschini employed nonparametric kernel regression methods to estimate cost functions and recover economies of size and scale. GK, Ker and Coble (1997, 1998), and Turvey and Zhao estimated yield densities using

[5] Heteroskedasticity is tested for using Goldfeld–Quandt parametric and nonparametric tests on both the raw and standardized (by predicted value) residuals. As with GK, the results indicate the standardized residuals be used for the analysis.

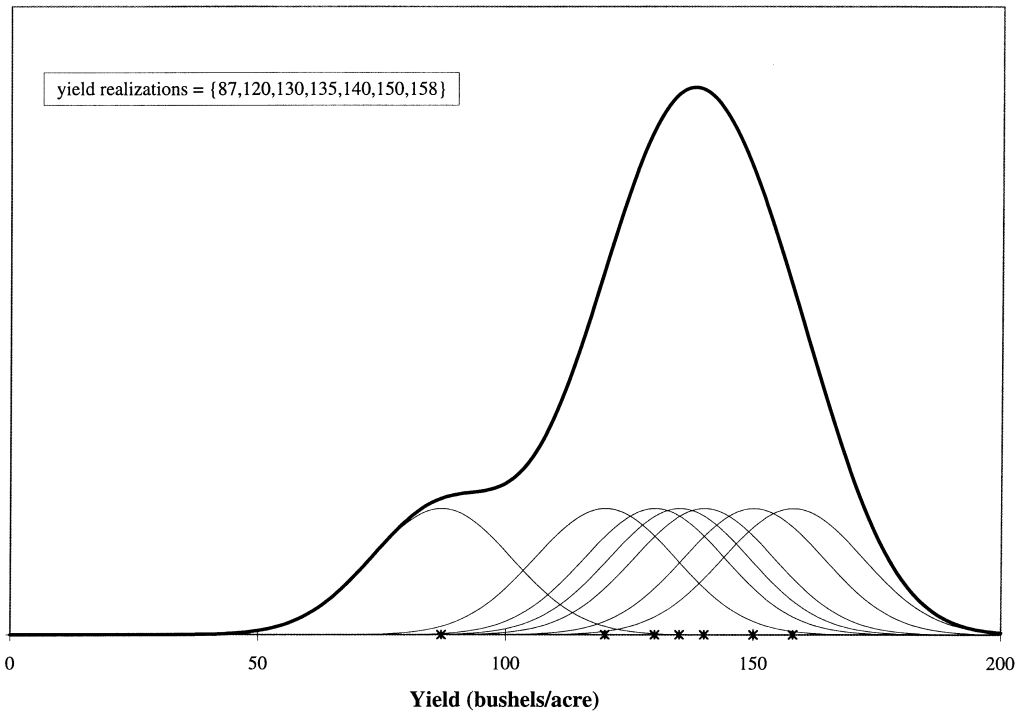

Figure 1. Standard kernel density estimator

univariate kernel density estimators. Univariate kernel density estimation is very intuitive (Silverman, Scott). A required input of the kernel density estimator is a set $(y_1, y_2, \ldots, y_T)$ of independent realizations from the unknown density of interest f_Y. Oversimplifying for the moment, the kernel density estimator places a *bump* or *individual kernel* at each realization. The estimate of the density at any given point in the support is simply the sum of the individual kernels at that point. Figure 1 illustrates this using an arbitrary sample of seven realizations. An individual kernel is centered at each of the seven points. The density estimate is just the sum of the individual kernels. Although this explanation is oversimplified, it serves to illustrate the intuitive nature of the nonparametric kernel density estimator. Conveniently ignored was any discussion regarding the shape of the individual kernels. We discuss these matters below.

Consider estimating the unknown yield density f_Y based on a set $(y_1, y_2, \ldots, y_T)$ of independent realizations from f_Y. The nonparametric kernel estimate of f_Y at a given point, say y_0, is defined as

$$(5) \quad \hat{f}_Y(y_0) = \sum_{i=1}^{T} \frac{K\left(\frac{y_0 - y_i}{h}\right)}{Th}$$

where h is the smoothing parameter and $K(\cdot)$ is the kernel function. Thus, two decisions must be made: choice of the kernel function, and choice of the smoothing parameter. A kernel function must integrate to one but need not be everywhere nonnegative. Epanechnikov derived the optimal non-negative kernel function with respect to minimizing mean integrated squared error (MISE) of the estimated density. Subsequently, Rosenblatt showed that choice of a suboptimal kernel, such as the standard Gaussian, results in only a moderate loss in the asymptotic MISE. Following general practice, a standard Gaussian kernel is used in figure 1 and throughout the analyses.[6]

The choice of the smoothing parameter requires two distinct decisions. The first decision is the choice of the smoothing parameter itself. The second is whether this smoothing parameter should be global or local. We choose the smoothing parameter according to

[6] In practice, a truncated Gaussian must be used since the estimated density is evaluated over real closed sets with finite Lebesgue measure while the support for the Gaussian density has infinite measure. We evaluate the densities over a range of plus and minus ten standard deviations from the mean.

Silverman's rule of thumb:

$$(6)\quad \hat{h} = 0.9 \times \min\left[\text{standard deviation}, \frac{\text{interquartile range}}{1.34}\right] \times T^{-(1/5)}.$$

This has been found to yield a mean integrated square error within 10% of the optimum for t-distributions, for log-normal distributions with skewness up to about 1.8, and a Gaussian mixture with separation up to three standard deviations (Silverman). Visual inspection of the estimated densities suggests that the set of unknown true densities belongs to the above defined class.

The second decision is the choice between a local or global smoothing parameter. A global smoothing parameter smooths the data equally. It is sometimes the case where the chosen smoothing parameter will yield too much spurious detail in the tails of the density in attempts to identify detail in the main area of the density. Undersmoothing in the tail is particularly problematic in long-tailed densities such as conditional yield densities. Given the high dependence of the derived premium rates on the extreme lower tail of the conditional yield density, a global smoothing parameter is particularly problematic. Thus, in contrast to GK, *adaptive kernel* methods are employed.

Recall the kernel estimator is the sum of individual kernels centered at each realization. The adaptive kernel estimator simply allows the smoothing parameter to vary with each realization. That is, a vector of smoothing parameters with dimension equal to the data rather than a single smoothing parameter is employed. Given that we are concerned with undersmoothing in the tails, we desire our smoothing parameters to be inversely related to the denseness of the data. Thus, a tail realization would have its individual kernel significantly flatter than a non-tail realization. Figure 2 illustrates the adaptive kernel density estimator based on the same seven yield realizations as in figure 1. It is clear how the adaptive kernel estimator is composed of individual kernels, still centered on the seven realizations, but with differing variances unlike the standard kernel estimator.

Given the smoothing parameter based on Silverman's rule of thumb, we adapt or adjust it for each individual kernel. The first problem is to decide whether a realization belongs to a relatively dense or sparse region. If the true density were known, we could compare the realization to the true density and make a decision regarding the necessary smoothness for its individual kernel. Clearly, we do not know the true density. Thus, a pilot estimate of the density needs to be used. For the pilot, we use the standard kernel estimate. Denoting the pilot estimate $\check{f}$, the local scale λ_i is defined as

$$(7)\quad \lambda_i = \left(\frac{\check{f}(y_i)}{g}\right)^{-\alpha}$$

where $\log(g) = \frac{1}{T}\sum \log \check{f}(y_i)$ and $\alpha \in [0, 1]$ is the sensitivity parameter. Now consider estimating the unknown yield density f_Y based on a set $(y_1, y_2, \ldots, y_T)$ of independent realizations from f_Y with a vector of smoothing parameters. The adaptive kernel estimate of f_Y at a given point, say y_0, is defined as

$$(8)\quad \hat{f}_Y(y_0) = \sum_{i=1}^{T} \frac{K\left(\frac{y_0 - y_i}{\lambda_i h}\right)}{Th\lambda_i}$$

where $h\lambda_i$ is the smoothing parameter for realization i, and $K(\cdot)$ is the kernel function. Silverman reviews this methodology and notes that the adaptive estimate is relatively insensitive to the pilot estimate. The smoothing parameter vector depends on the power to which the pilot density is raised. The larger α, the more sensitive the method will be to variations in the pilot density and the more difference there will be between the smoothing parameters. Obviously, setting $\alpha = 0$ reduces the adaptive method to the standard kernel estimate. We set $\alpha = \frac{1}{2}$ for theoretical reasons outlined by Abramson. Although the adaptive kernel estimator increases the computational complexity of our Bayesian nonparametric kernel estimator, it is not without reward, particularly when tail estimation is crucial.

An unfortunate problem with using kernel estimators is that the estimated density does not necessarily have moments equal to the sample moments. The consistency of these estimators indicates that this is a finite sample problem. However, in the size of samples we have for estimating conditional yield densities, this can be disconcerting. Consider the first two moments of the estimated density using adaptive kernel methods. A kernel function is said to be of order p if $\int K(u)du = 1$; $k_i = \int u^i K(u)du = 0$, $i = 1, \ldots, p-1$; and $k_p = \int u^p K(u)du \neq 0$. Therefore, the standard Gaussian kernel,

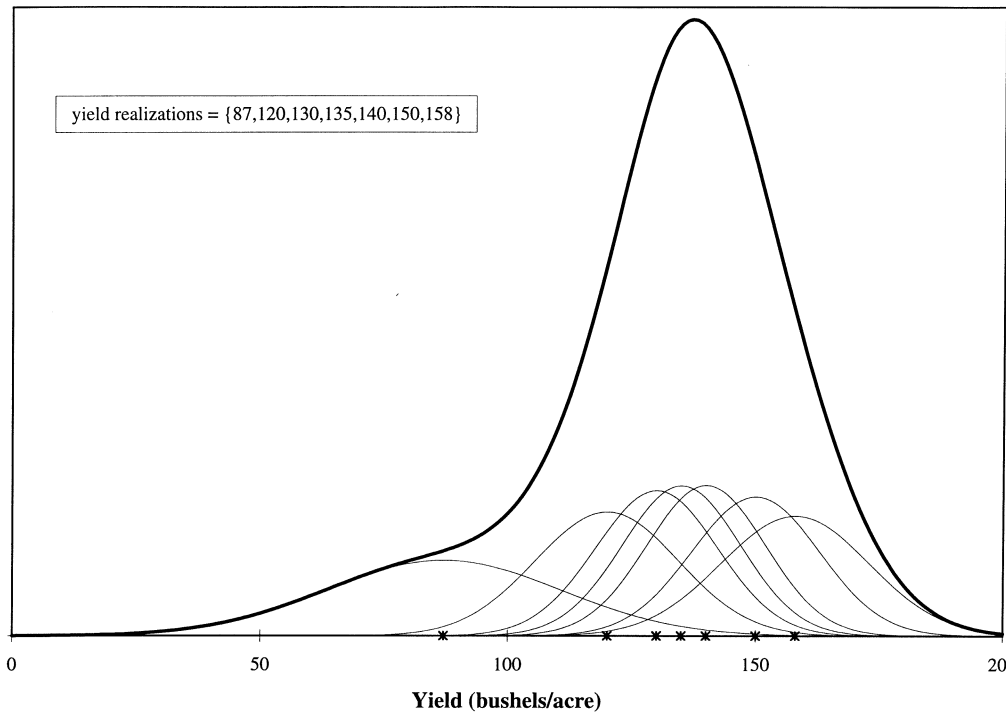

Figure 2. Adaptive kernel density estimator

which has mean 0 and variance 1, is a kernel of order 2 because $k_1 = 0$ and $k_2 = 1 \neq 0$. Now consider the first moment of the kernel density estimate. If $k_1 = 0$, that is, the kernel is symmetric about zero, the estimated density has mean equal to the sample mean almost surely. Intuitively, a symmetric kernel function about zero ensures the mean of each individual kernel is the point at which it is centered on, independent of the smoothing parameter. Given each kernel is equally weighted (they all have equal mass), the mean of the adaptive kernel estimate is just the sum of the sample divided by T. Obviously this is the sample mean. Thus, our estimated densities will have means equal to their respective sample means (see Lemma 1, Appendix 1 for proof).

With respect to the second sample moment, the estimated density will have variance greater than or equal to the sample variance almost surely for an order two kernel.[7] The proof for the adaptive kernel estimator is given in Lemma 2, Appendix 1. The additional variance of the kernel estimate is rather intuitive. Since the mass $(\frac{1}{T})$ at each realization is being smoothed or spread out, variance must necessarily increase. This is an undesirable property given the sample variance is an *unbiased* estimator of the population variance. For our application, this is particularly problematic because the additional variance may be large in small samples while tail probabilities, and thus derived rates, are quite sensitive to changes in variance.

Appendix 1 derives the variance of the estimated density from the adaptive kernel estimator. The variance is

$$\text{(9)} \quad \text{var}(y) = \frac{h^2 \sum_{i=1}^{T} \lambda_i^2}{T} + \frac{T-1}{T} s^2 = O_p\left(T^{-(7/5)}\right) + O(1)s^2$$

where $y \sim \hat{f}_Y$ and s^2 is the sample variance of the set of independent realizations from f_Y. We adjust our adaptive kernel estimates by taking a scale transformation of the support (see Lemma 2, Appendix 1).[8] In doing this, we force the estimated density to have variance equal to the sample variance. Beirens undertakes a different approach which leads to the same result. Rather than taking a scale transformation of the initial kernel estimate,

[7] The equality holds for $T = \{1, \infty\}$.

[8] The kernel is evaluated at uniformly spaced points along an interval of ± 10 standard deviations. The transformation is taken with respect to the support points at which the kernel estimate is calculated.

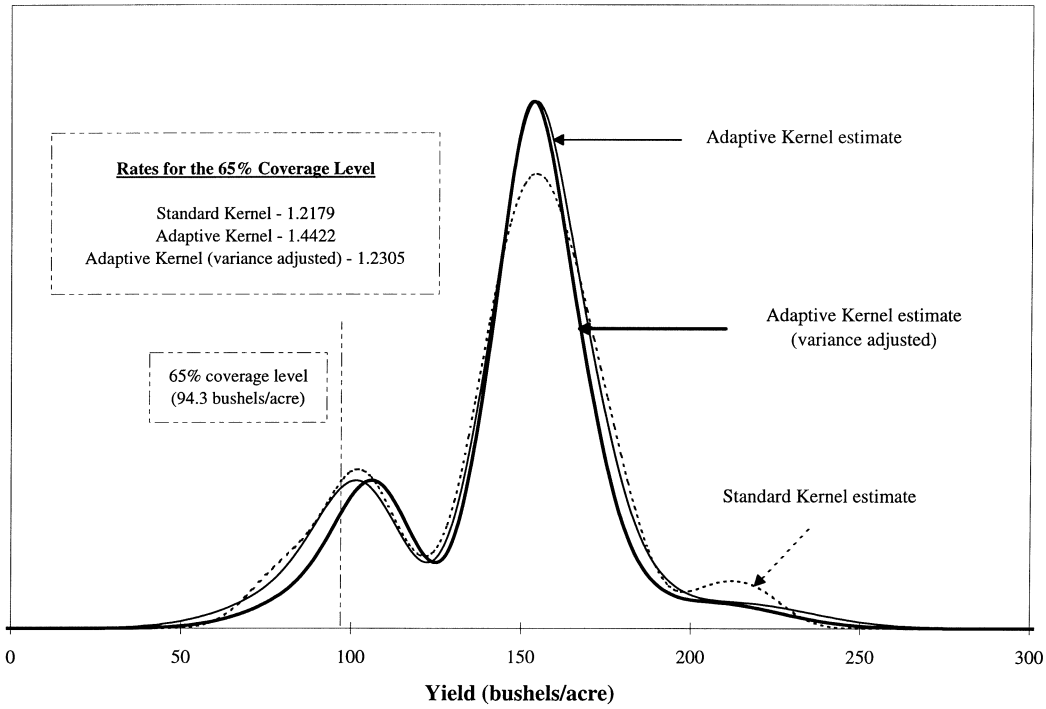

Figure 3. Adaptive versus standard kernel; Adair County, Iowa

Beirens scales the data prior to entering the kernel.[9]

Figure 3 illustrates the difference between the standard kernel estimate, the adaptive kernel estimate, and the adaptive kernel estimate adjusted to have variance equal to the sample variance for Adair County. First consider the difference between the adaptive kernel estimate and the standard kernel estimate. Clearly, the tail lumpiness with the standard kernel estimator is undesirable whereas the relative smoothness with the adaptive kernel estimator more closely resembles our prior beliefs. Note, however, that the adaptive kernel estimator will not be such that bimodality is removed when the data suggest the underlying density is bi-modal. Second, consider the difference between the adaptive kernel estimate and the adaptive kernel estimate adjusted to have variance equal to the sample variance. Their respective rates at the 65% coverage level are perhaps more different than one might have thought. As discussed, the additional variance can be non-trivial in small samples despite its quick convergence rate $O_p(T^{-(7/5)})$. The difference in the premium rate for the 65% coverage level is economically significant. If one does not correct for the inflated variance, the rates are increased by 17.2%. Not surprisingly, this result was found for most counties. Obviously, as the coverage level is increased, the effect of the additional variance on the rates decreases.

Bayesian Nonparametric Kernel Density Estimation

We wish to consider not only a single conditional yield density, but a set of conditional yield densities, one for each county. Denote the number of counties as Q and the adaptive kernel estimate at support point y_j for county i as $\hat{f}_{ij}$. Ker (1998) proposes the following hierarchical model:

$$(10)\quad \hat{f}_{ij}|\mu_{ij} \sim N(\mu_{ij}, \sigma_{ij}^2)$$
$$\mu_{ij} \sim N(\mu_j, \tau_j^2)$$

where $\mu_{ij} = f_{ij} + \beta_{ij}$, f_{ij} is the unknown density value for county i at support point y_j, β_{ij} is the bias for county i at support point y_j, σ_{ij}^2 is the variance of the adaptive kernel density estimate for county i at support y_j, μ_j is the

[9] An alternative solution would be to employ a higher-order kernel such as a fourth order kernel where $\int K(u)du = 1$, $k_1 = k_2 = k_3 = 0 \neq k_4$. In this case, $\text{var}(y) = \frac{T-1}{T} s_y^2$ a.s. since $k_2 = 0$ (see Lemma 2, Appendix 1). Unfortunately, higher-order kernels require negative mass which may lead to negative mass in the estimated yield density. Therefore, higher-order kernels were not employed.

mean value of the densities across counties at support y_j, and τ_j^2 is the variance across counties at support y_j. The intuition behind the hierarchical model is that even though the μ_{ij}'s are mutually independent for a given j, they are tied together in that there is one loss function for estimating the Q densities at support point y_j. Thus, in flavor similar to Stein's paradox, an estimator (the posterior) which is a function of the $\{\hat{f}_{1j}, \ldots, \hat{f}_{Qj}\}$ is constructed which *may* be preferable to the adaptive kernel estimate. If the bias term β_{ij} was zero or asymptotically zero, we could say that the resulting posterior dominates the adaptive kernel estimate, thus making it inadmissible. Unfortunately such a statement cannot be made and dominance, either finite or asymptotically, cannot be asserted regarding either estimator. As a result, Ker (1998) proposes a simulation approach to gauge the value of the empirical Bayes estimator in a given situation. The posterior estimate for the hierarchical is

$$(11) \quad \tilde{f}_{ij} = \hat{f}_{ij}\left(\frac{\tau_j^2}{\tau_j^2 + \sigma_{ij}^2}\right) + \mu_j\left(\frac{\sigma_{ij}^2}{\tau_j^2 + \sigma_{ij}^2}\right)$$

where the unknowns (μ_j, τ_j^2, σ_{ij}^2) must be estimated. Bootstrap methods are used to estimate the variance σ_{ij}^2. An estimate of the mean and variance across counties is obtained using the following method of moments estimators: $\hat{\mu}_j = \frac{1}{Q}\sum_{i=1}^{Q} \hat{f}_{ij}$ and $\hat{\tau}_j^2 = \hat{s}_j^2 - \frac{1}{Q}\sum_{i=1}^{Q} \hat{\sigma}_{ij}^2$ where $\hat{s}_j^2 = \frac{1}{Q-1}\sum_{i=1}^{Q}(\hat{f}_{ij} - \hat{\mu}_j)^2$ (see Lemma 3, Appendix 1). Therefore, the empirical Bayes nonparametric kernel density estimator at support y_j for county i is

$$(12) \quad \tilde{f}_{ij} = \hat{f}_{ij}\left(\frac{\hat{\tau}_j^2}{\hat{\tau}_j^2 + \hat{\sigma}_{ij}^2}\right) + \hat{\mu}_j\left(\frac{\hat{\sigma}_{ij}^2}{\hat{\tau}_j^2 + \hat{\sigma}_{ij}^2}\right).$$

The resulting posterior or empirical Bayes nonparametric kernel estimate $\tilde{f}_{ij}$ is very intuitive. As the estimated variance of the kernel estimates across counties increases ($\hat{\tau}_j^2 \uparrow$), the less the set of adaptive kernel estimates $(\hat{f}_{1j}, \hat{f}_{2j}, \ldots, \hat{f}_{Qj})$ will shrink toward the overall mean ($\hat{\mu}_j$). Conversely, the larger the estimated variance of the kernel estimate for a given county ($\hat{\sigma}_{ij}^2$), the more that given adaptive kernel estimate ($\hat{f}_{ij}$) shrinks toward the overall mean ($\hat{\mu}_j$). As expected with many shrinkage or Stein type estimators, the greater the variance within the experimental units relative to the variance across the experimental units, the greater the shrinkage and the greater the potential improvements in efficiency. Ker (1998) indicates that the empirical Bayes nonparametric kernel estimator may offer the largest efficiency gains in small samples where the variance within counties tends to be relatively high as compared to the variance across counties. Figure 4 illustrates, for Adair County, the empirical Bayes nonparametric kernel density estimate ($\tilde{f}$), along with the adaptive kernel estimate ($\hat{f}$) and the mean of the adaptive kernel estimates for the subset containing Adair County.

Empirical Simulation

The empirical Bayes nonparametric kernel estimate has been proposed as a *competing* estimator to the adaptive kernel estimator. Ker (1998) proposes the following approach for choosing between the competing estimators. The data are initially used to recover estimates of the yield densities by employing the adaptive kernel estimator only. This constitutes a set of pilot density estimates from which random samples of the size of the original sample may be drawn. For each sample from the adaptive kernel estimates, or equivalently, each smoothed bootstrap sample of the original data, the adaptive kernel and the empirical Bayes nonparametric kernel estimators are recovered and compared to the pilots. If the empirical Bayes nonparametric kernel estimator mean metric value, for some appropriately chosen metric, is significantly lower than that of the adaptive kernel estimator, then the empirical Bayes nonparametric kernel estimator would be the preferable estimator if in fact the pilots were the unknown densities of interest. Conversely, if the adaptive kernel estimator mean metric value is significantly lower than that of the empirical Bayes nonparametric kernel estimator, then the adaptive kernel estimator would be preferable, if again the pilots were the unknown densities of interest. Note that this type of simulation does not indicate which estimator is preferable for the unknown yield densities, but rather which estimator is preferable if the initial estimates were the unknown yield densities. Although this may seem a reasonable way to proceed, we are comparing the competing estimators to an estimate of the unknown yield densities rather than the

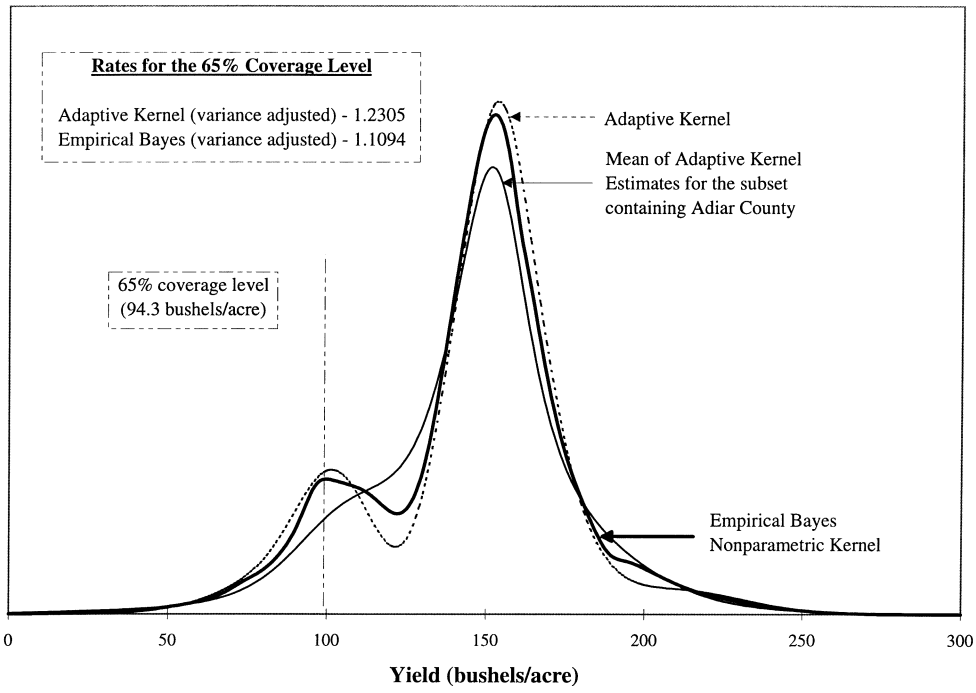

Figure 4. Adaptive kernel versus empirical Bayes nonparametric kernel; Adair County, Iowa

true unknown yield densities and thus caution must be taken.

Recall that county yield data for Iowa all-practice corn is the application. There are ninety-nine counties in Iowa, all of which produce a significant amount of corn. Rather than grouping the entire set of ninety-nine counties together for the empirical Bayes nonparametric kernel density estimator, subsets are taken for two reasons. First, yields are believed to be dependent among contiguous counties. The empirical Bayes nonparametric kernel density estimator will tend to overshrink toward the mean ($\hat{\mu}_j$) because the estimated variance ($\hat{\tau}_j^2$) will be negatively biased. This approach avoids the issue of explicitly modeling spatial correlation which would be necessary if neighboring counties were pooled. Extensions to this research may profit from a consideration of methods for pooling nonindependent observations and explicit modeling of spatial correlation. Second, the more similar the underlying densities, the greater nonparametric kernel density estimator relative to the adaptive kernel. In fact, Ker (1998) illustrates that if the unknown densities of interest are known up to an estimable transformation, the asymptotic MISE of the empirical Bayes nonparametric kernel density estimator is lower than the standard kernel density estimator.[10] Because Central Limit Theorems for dependent processes play a significant role in the shape of the conditional yield densities, counties with similar planted acres would tend to be similar in shape. Therefore, we create subsets of counties with similar mean planted acres subject to the restriction that no contiguous counties belong to the same set. This results in eleven subsets composed of nine counties each. The empirical Bayes nonparametric kernel estimator is employed separately with each subset. The adaptive kernel estimate for a given county is unaffected by the composition of the subsets.

A very nice feature of kernel estimators is that they are invariant to transforming the data, estimating the density with the transformed data, and then taking the inverse transformation on the estimated density. To ensure our density estimates have mean and variance determined by their county data only, the data are transformed to have mean zero and variance one prior to entering the density estimators. The estimated densities are then transformed back to have mean and variance equal to the sample mean and variance.

[10] The result is easily generalized to the adaptive kernel estimator.

One hundred smoothed bootstrap samples of the size of our data (35) were taken in order to compare the empirical Bayes to the adaptive kernel estimator. The smoothed bootstrap samples, after making the appropriate variance adjustment, are independent draws from the adaptive kernel estimate of the yield densities. The competing estimators are compared for each bootstrap sample to the underlying density (original adaptive kernel estimate) using both L_1 and L_2 norms. The L_1 norm is included as the tails are given more weight relative to the L_2 norm. For each simulation, the total metric value over the nine counties in each of the eleven subsets is recovered. Table 1 presents the percentage decrease, from the adaptive kernel estimator to the empirical Bayes kernel density estimator, of the mean (over the 100 simulations) of each subset's total metric value. Paired t-tests were performed for each county as well as the total for each subset. The percentage decreases in both L_1 and L_2 norms are large and statistically significant for each subset. These simulation results suggest the empirical Bayes nonparametric kernel density estimator is very likely to offer significant efficiency gains in estimating the conditional yield densities. Additionally, for subset one, smoothed bootstrap samples (100) of varying sizes were taken from the pilot estimates. The total L_2 norm over the nine counties at various sample sizes is illustrated for both estimators in figure 5. The results suggest that one would need approximately sixty-one years of yield data to estimate the shape of the pilots with adaptive kernel estimators as accurately, according to the L_2 norm, as the thirty-five years of yield data with the empirical Bayes nonparametric kernel density estimator. Given that scarcity of data remains the largest problem for rating crop insurance contracts, these simula-

Table 1. Empirical Bayes versus Adaptive Kernel: Percentage Decrease in Mean Metric

Subset	L_1 Norm	L_2 Norm
1	10.74%	19.68%
2	13.67%	19.00%
3	13.15%	13.13%
4	10.96%	12.81%
5	10.96%	9.55%
6	9.31%	7.79%
7	9.23%	24.31%
8	84.4%	9.42%
9	10.65%	16.47%
10	94.9%	9.05%
11	15.63%	13.71%

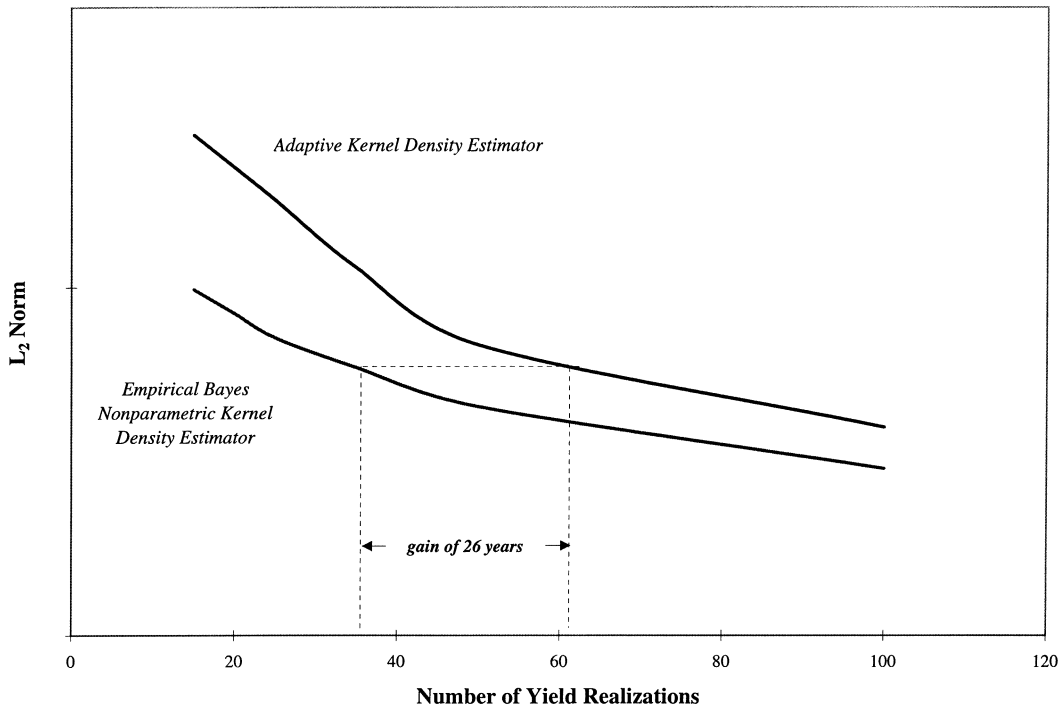

Figure 5. Empirical Bayes versus adaptive kernel

tion results are noteworthy. They illustrate that potentially large gains in estimation efficiency may be captured with the use of the empirical Bayes nonparametric kernel density estimator.

Rating Crop Insurance Contracts: Empirical Results

In this section we derive and contrast estimates of the 1997 GRP rates for Iowa all-practice corn associated with the various conditional yield density estimates. Specifically, we contrast the empirical Bayes nonparametric kernel rates with existing GRP rates, GK rates, standard kernel rates, standard kernel rates adjusted for variance inflation, adaptive kernel rates, and, adaptive kernel rates adjusted for variance inflation. Comparisons are made for the 65% and 85% coverage levels.

Recall the temporal models were found to be ARIMA(4, 1, 0) and thus

$$(13)\quad y_t = y_{t-1}+\beta_0+\beta_1(y_{t-1}-y_{t-2}) +\beta_2(y_{t-2}-y_{t-3})+\beta_3(y_{t-3}-y_{t-4}) +\beta_4(y_{t-4}-y_{t-5})+e_t$$

and

$$(14)\quad y_{T+2} = \hat{y}_{T+2} + \nu_{T+2}$$

where $\nu_{T+2} = (1+\beta_1)e_{T+1} + e_{T+2}$ and $\hat{y}_{T+2}$ is the two-step ahead prediction.[11]

The set of transformed innovations represents a set of asymptotically independent realizations from $f(e)$. Given that GRP rates *must* be set two years in advance, we require the density $f(\nu)$. In general, innovations are inflated or scaled by $\sqrt{1+(1+\beta_1)^2}$ to account for the increased variance as y_{t+1} is unknown. This procedure is valid if, and only if, the innovation distribution has two special properties: the family is a scale family distribution and the convolutions of two independent draws belong to the same family. The Gaussian family is rare in that it maintains these two properties. Unfortunately, conditional yield densities $f(y_{T+1}|\mathcal{F}_T)$ are not Gaussian.[12] However, estimates of the density $\hat{f}(\nu)$ may be numerically recovered from the nonparametric estimates of the density $\hat{f}(e)$ (see Appendix 2 for the derivations). Figure 6 illustrates, for Adair County, the effect of considering the two-step ahead prediction using the empirical Bayes nonparametric kernel density estimator. Not surprisingly, there is very little difference since $\hat{\beta}_1 = -0.887$ for Adair County. Since $\nu_{T+2} = (1+\beta_1)e_{T+1} + e_{T+2}$, as $\beta_1 \to -1$ then $\nu_{T+2} \to e_{T+2}$ and $f(\nu) \to f(e)$.

Table 2 illustrates the rates derived from the various methods averaged over the ninety-nine counties.[13] The results are as one would expect. First, the standard and adaptive kernel rates are significantly higher than their respective counterparts that adjust the densities for the inflated variance. For the standard kernel estimator, the rates are 16% and 10% smaller, on average, for the 65% and 85% coverage levels, respectively. At the 65% coverage level, the percentage decrease, by county, ranges from 2% to as high as 63% lower when the standard kernel density estimate is transformed to have variance equal to the sample variance. Similarly, the percentage decrease, by county, ranges from 1% to 20% lower at the 85% coverage level. With respect to the adaptive kernel estimator, the rates are 19% and 12% smaller, on average, for the 65% and 85% coverage levels, respectively. The percentage decrease, by county, ranges from 4% to 58% for the 65% coverage level and from 3% to 27% for the 85% coverage level. Second, the adaptive kernel rates tend to be higher than the standard kernel rates at the lower coverage level given mass in the tails of the estimated densities is more dispersed. Third, the empirical Bayes rates are in line with the variance adjusted adaptive kernel rates. This is expected given the empirical Bayes nonparametric kernel density estimator is based on the variance adjusted adaptive kernel estimator. Finally, the empirical Bayes rates, adjusted for the two-step ahead forecast, are 3% higher than the unadjusted rates for both coverage levels. Although the increase was small for most counties, the percentage adjustment, by county, ranged from approximately 0% to 20% at the 65% coverage level and 0% to 11% for the 85% coverage level.

[11] The two-step ahead prediction is defined as $\hat{y}_{T+2} = (1+\beta_1)y_T + (2+\beta_1)\beta_0 + (\beta_1+\beta_1^2+\beta_2)(y_T - y_{T-1}) + (\beta_2+\beta_1\beta_2+\beta_3)(y_{T-1}-y_{T-2}) + (\beta_3+\beta_1\beta_3\beta_4)(y_{T-2}-y_{T-3}) + (\beta_4+\beta_1\beta_4)(y_{T-3}-y_{T-4})$.

[12] Ker (1996) illustrates that if the innovations are a mixture of two Gaussians, which not surprisingly happens to be a location-scale family, then the convolution is a mixture of four Gaussians. Thus, variances may not simply be scaled.

[13] Rates by method and county are available upon request from the lead author.

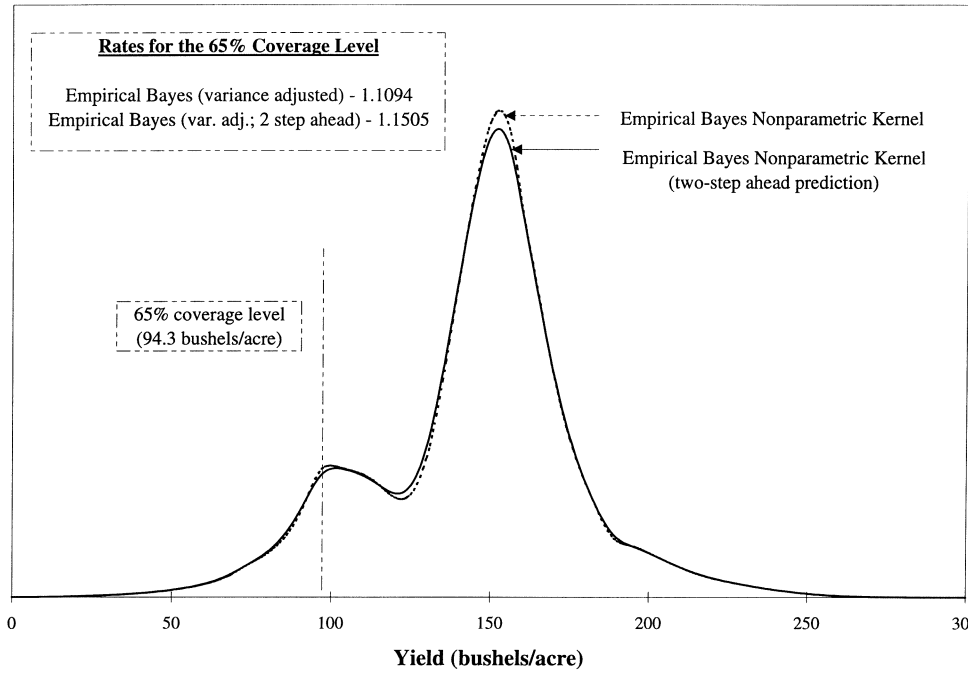

Figure 6. Empirical Bayes versus empirical Bayes two-step ahead prediction; Adair County, Iowa

Conclusions

This manuscript revisits the use of nonparametric kernel methods used in GK for estimating the conditional yield densities and the subsequent derivation of premium rates. We feel this is particularly topical given the increased prominence of the crop insurance program in the U.S. agricultural policy agenda.

A concern with using nonparametric kernel methods or so-called flexible parametric forms in small samples is stability. Given this concern, GK combined innovations recovered from the temporal models of contiguous counties to estimate the conditional yield densities. We have two concerns with that approach. First, the method by which the innovations from the temporal models were combined was admittedly ad hoc. Second, innovations from contiguous counties are not likely to be independent. Our objective in this manuscript has been to employ both an equally flexible method for estimating conditional yield densities as well as one that exploits the similarities among the county yield densities. The empirical Bayes nonparametric kernel density estimator, which uses

Table 2. Mean 1997 GRP Rates for Iowa All-Practice Corn

Methodology[a]	Mean Rate 65% Coverage Level	Mean Rate 85% Coverage Level
Actual GRP rates	1.4609	3.5118
Goodwin and Ker rates	1.6440	4.0379
Standard kernel rates	1.6147	4.3944
Standard kernel rates[b]	1.3581	3.9724
Adaptive kernel rates	1.8468	4.3945
Adaptive kernel rates[b]	1.5017	3.8500
Empirical Bayes nonparametric kernel rates[b]	1.5066	3.8645
Empirical Bayes nonparametric kernel rates[b, c]	1.5496	3.9653

[a] Rates by county and methodology are available from the lead author.
[b] Densities are adjusted to have variance equal to sample variance.
[c] Accounts for two-step ahead prediction.

empirical Bayes techniques on the estimated values from the kernel density estimator, was used. Simulations suggest that this estimator may provide very significant efficiency gains in estimating conditional yield densities. In our simulations, sixty-one years of data were required for the adaptive kernel to estimate the shape of the conditional yield densities (up to a location-scale transformation), as accurately, on average and with respect to L_2 norm, as the empirical Bayes nonparametric kernel density estimator given only thirty-five years of data. Finally, the empirical Bayes nonparametric kernel estimator may be employed in rating other crop insurance products as well.

In addition to employing a new estimator, this manuscript makes three additional strides. First, we employ an adaptive kernel method. Accurate estimation of lower tail probabilities is crucial to deriving accurate rates. A variable smoothing approach significantly decreases the dependency of estimated tail probabilities on the specific location of the tail realizations. Second, we force the estimated densities to have variance equal to the sample variance. The variances of the standard and adaptive kernel density estimators are greater than the sample variance almost surely when a kernel of order two, such as the standard Gaussian, is used. Although it was shown that this additional variance is $O(T^{-(7/5)})$, it was found to have a significant effect on the derived rates. Finally, we have explicitly recognized that rates need to be calculated two years in advance. Given the AR(1) parameters are close to -1 for most counties, the resulting rates do not increase significantly by considering a two-step ahead forecast.

[Received November 1997; accepted August 1999.]

References

Abramson, I.S. "On Bandwidth Variation in Kernel Estimates—a Square Root Law." *Annals of Statist*. 10(December 1982):1217–23.

Babcock, B.A., and D.A. Hennessy. "Input Demand under Yield and Revenue Insurance." *Amer. J. Agr. Econ*. 78(May 1996): 416–27.

Bierens, H.J. "Sample Moments Integrating Normal Kernel Estimators of Multivariate Density and Regression Functions." *Indian J. Statist. Ser. B* 45(August 1983):160–92.

Bessler, D. "Aggregated Personalistic Beliefs on Yields of Selected Crops Estimated Using ARIMA Processes." *Amer. J. Agr. Econ*. 62(November 1980):666–74.

Borges, R., and W. Thurman. "Marketing Quotas and Subsidies in Peanuts." *Amer. J. Agr. Econ*. 76(May 1994):801–09.

Coble, K.H., T.O. Knight, R.D. Pope, and J.R. Williams. "Modeling Farm-Level Crop Insurance Demand with Panel Data." *Amer. J. Agr. Econ*. 78(May 1996):439–47.

Epanechnikov, V.A. "Nonparametric Estimation of a Multidimensional Probability Density." *Theory Probability Applic*. 14(1969):153–58.

Gallagher, P. "U.S. Soybean Yields: Estimation and Forecasting with Nonsymmetric Disturbances." *Amer. J. Agr. Econ*. 69(November 1987):798–803.

Goldfeld, S.M., and R.E. Quandt. "Some Tests for Heteroskedasticity." *J. Amer. Statist. Assoc*. 60(June 1965):539–47.

Goodwin, B.K., and A.P. Ker, "Nonparametric Estimation of Crop Yield Distributions: Implications for Rating Group-Risk (GRP) Crop Insurance Contracts." *Amer. J. Agr. Econ*. 80(February 1998):139–53.

Kenkel, P.I., J.C. Busby, and J.R. Skees. "A Comparison of Candidate Probability Distributions for Historical Yield Distributions." Paper presented at SAEA meetings, Fort Worth TX, February 1991.

Ker, A.P. "Empirical Bayes Nonparametric Kernel Density Estimation." Working paper, Dept. of Agr. and Resource Econ., University of Arizona, 1998.

——. "Using SNP Maximum Likelihood Techniques to Recover Conditional Densities: A New Approach to Recovering Premium Rates." Working paper, Dept. of Agr. and Resource Econ., University of Arizona, 1996.

Ker, A.P., and K. Coble. "On The Use of the Beta Distribution in Modeling Yields." Working paper, Dept. of Agr. and Resource Econ., University of Arizona, 1997.

——. "On Choosing a Base Coverage Level for Multiple Peril Crop Insurance Contracts." *J. Agr. Resource Econ*. 23(1998):427–44.

Lee, H., J. Harwood, and A. Somwaru. "Implications of Disaster Assistance Reform for Noninsurable Crops." Paper presented at AAEA Meetings, Indianapolis IN, August 1995.

Moschini, G. "Nonparametric and Semiparametric Estimation: An Analysis of Multiprod-

uct Returns to Scale." *Amer. J. Agr. Econ.* 72(August 1990):589–96.

Moss, C.B., and J.S. Shonkwiler. "Estimating Yield Distributions with a Stochastic Trend and Nonnormal Errors." *Amer. J. Agr. Econ.* 75(November 1993):1056–72.

Nelson, C.H. "The Influence of Distribution Assumptions on the Calculation of Crop Insurance Premia." *N. Central J. Agr. Econ.* 12(January 1990):71–78.

Nelson, C.H., and P.V. Preckel. "The Conditional beta Distribution As a Stochastic Production Function." *Amer. J. Agr. Econ.* 71(May 1989): 370–78.

Rosenblatt, M. "Curve Estimation." *Ann. Math. Statist.* 42(December 1971):1815–42.

Scott, D.W. *Multivariate Density Estimation: Theory, Practice, and Visualization*. New York: John Wiley and Sons Inc, 1994.

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

Skees, J.R., J.R. Black, and B.J. Barnett. "Designing and Rating an Area Yield Crop Insurance Contract." *Amer. J. Agr. Econ.* 79(May 1997): 430–38.

Turvey, C.G., and C. Zhao. "Parametric and Nonparametric Crop Yield Distributions and their Effects on All-Risk Crop Insurance Premiums." Working paper, Dept. of Agr. Econ. and Business, University of Guelph, May 1993.

Appendix 1

This appendix contains the technical derivations for the adaptive kernel estimator and the empirical Bayes kernel density estimator. Although we have not seen the below proofs in the statistical literature, most likely because of their trivial nature, we are not advancing these as contributions. Rather, we include them because of their pertinence to our analysis. For Lemmas 1 and 2 it is necessary to define second order kernels. A kernel is of order two if $k_1 = \int uK(u)du = 0$, $k_2 = \int u^2K(u)du \neq 0$, and $\int K(u)du = 1$.

Lemma 1. *The mean of the estimated density from the adaptive kernel estimator is equal to the sample mean almost surely:*

$$\begin{aligned}
\text{(A.15)}\quad E(y) &= \int y\hat{f}(y)dy \\
&= \int y \frac{1}{Th} \sum_{i=1}^{T} \left(\frac{K\left(\frac{y-y_i}{\lambda_i h}\right)}{\lambda_i} \right) dy \\
&= \frac{1}{T} \sum_{i=1}^{T} \int y \frac{1}{\lambda_i h} K\left(\frac{y-y_i}{\lambda_i h}\right) dy \\
&= \frac{1}{T} \sum_{i=1}^{T} \int \frac{u\lambda_i h + y_i}{\lambda_i h} \\
&\quad \times K(u)\lambda_i h\, du \\
&\quad \text{where } u = \frac{y-y_i}{\lambda_i h} \\
&= \frac{1}{T} \sum_{i=1}^{T} \left(\lambda_i h \int uK(u)\, du \right. \\
&\quad \left. + y_i \int K(u)\, du \right) \\
&= \frac{1}{T} \sum_{i=1}^{T} (0 + y_i) \\
&= \bar{y} \quad \text{a.s.}
\end{aligned}$$

Lemma 2. *The variance of the estimated density from the adaptive kernel estimator is greater than the sample variance almost surely. Denote* $\int u^2K(u)\, du = k_2$. *For the standard Gaussian kernel,* $k_2 = 1$,

$$\begin{aligned}
\text{(A.16)}\quad E(y^2) &= \int y^2 \hat{f}(y)\, dy \\
&= \int y^2 \frac{1}{Th} \sum_{i=1}^{T} \left(\frac{K\left(\frac{y-y_i}{\lambda_i h}\right)}{\lambda_i} \right) dy \\
&= \frac{1}{T} \sum_{i=1}^{T} \int y^2 \frac{1}{\lambda_i h} K\left(\frac{y-y_i}{\lambda_i h}\right) dy \\
&= \frac{1}{T} \sum_{i=1}^{T} \int \frac{u^2\lambda_i^2 h^2 + 2y_i u h \lambda_i + y_i^2}{\lambda_i h} \\
&\quad \times K(u)\lambda_i h\, du \\
&\quad \text{where } u = \frac{y-y_i}{\lambda_i h} \\
&= \frac{1}{T} \sum_{i=1}^{T} \left(\lambda_i^2 h^2 \int u^2 K(u)\, du \right. \\
&\quad + 2y_i h \lambda_i \int uK(u)\, du \\
&\quad \left. + y_i^2 \int K(u)\, du \right) \\
&= \frac{1}{T} \sum_{i=1}^{T} (h^2\lambda_i^2 k_2 + 0 + y_i^2) \\
&= \frac{1}{T} \left(k_2 h^2 \sum_{i=1}^{T} \lambda_i^2 + \sum_{i=1}^{T} y_i^2 \right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\text{(A.17)}\quad \text{Var}(y) &= E(y^2) - E(y)^2 \\
&= \frac{1}{T} \left(h^2 k_2 \sum_{i=1}^{T} \lambda_i^2 + \sum_{i=1}^{T} y_i^2 \right) - \bar{y}^2 \\
&= \frac{h^2 k_2 \sum_{i=1}^{T} \lambda_i^2}{T} + \frac{T-1}{T} s_y^2 \quad \text{a.s.}
\end{aligned}$$

Lemma 3. $E[\hat{s}_j^2] = (\sum_{i=1}^{Q} \sigma_{ij}^2/Q) + \tau_j^2$ *where* $\hat{s}_j^2 = \frac{1}{Q-1} \sum_{i=1}^{Q} (\hat{f}_{ij} - \hat{\mu}_j)^2$ *and* $\hat{\mu}_j = \frac{1}{Q} \sum_{i=1}^{Q} \hat{f}_{ij}$.

If $\hat{s}_j^2 = \frac{1}{Q-1}\sum_{i=1}^{Q}(\hat{f}_{ij} - \hat{\mu}_j)^2$, *then*

$$E[s^2] = \frac{1}{Q-1}E\left[\sum_{i=1}^{Q}\hat{f}_{ij}^2 - Q\hat{\mu}_j^2\right] = \frac{1}{Q-1}\left[\sum_{i=1}^{Q}\left[E[\hat{f}_{ij}^2]\right] - QE[\hat{\mu}_j^2]\right]. \tag{A.18}$$

Since $\hat{f}_{ij}|\mu_{ij} \sim N(\mu_{ij}, \sigma_{ij}^2)$ *where* $\mu_{ij} \sim N(\mu_j, \tau_j^2)$ *then* $\hat{f}_{ij} \sim N(\mu_j, \sigma_{ij}^2 + \tau_j^2)$ *and* $E[\hat{f}_{ij}^2] = \sigma_{ij}^2 + \tau_j^2 - \mu_j^2$. *Similarly, under independence across counties* $\sum_{i=1}^{Q}\hat{f}_{ij} \sim N(Q\mu_j, \sum_{i=1}^{Q}\sigma_{ij}^2 + Q\tau_j^2)$ *and thus* $\hat{\mu}_j = \sum_{i=1}^{Q}(\hat{f}_{ij}/Q) \sim N\left(\mu_j, \left(\sum_{i=1}^{Q}\sigma_{ij}^2/Q^2\right) + (\tau_j^2/Q)\right)$ *and* $E[\hat{\mu}_j^2] = \left(\sum_{i=1}^{Q}\sigma_{ij}^2/Q^2\right) + (\tau_j^2/Q) - \mu_j^2$. *Hence,*

$$E[\hat{s}_j^2] = \frac{1}{Q-1}\left[\sum_{i=1}^{Q}\left[E[\hat{f}_{ij}^2]\right] - QE[\hat{\mu}_j^2]\right] \tag{A.19}$$

$$= \frac{1}{Q-1}\left[\sum_{i=1}^{Q}\sigma_{ij}^2 + Q\tau_j^2 - Q\mu_j^2 - \frac{\sum_{i-1}^{Q}\sigma_{ij}^2}{Q} - \tau_j^2 + Q\mu_j^2\right] \tag{A.20}$$

$$= \frac{1}{Q-1}\left[\frac{\sum_{i-1}^{Q}\sigma_{ij}^2}{Q}(Q-1) + \tau_j^2(Q-1)\right]$$

$$= \frac{\sum_{i=1}^{Q}\sigma_{ij}^2}{Q} + \tau_j^2.$$

Thus, τ_j^2 *is estimated by* $\hat{\tau}_j^2 = \hat{s}_j^2 - (\sum_{i=1}^{Q}\hat{\sigma}_{ij}^2/Q)$. *Note, if* $\hat{\sigma}_{ij}^2 = \hat{\sigma}_j^2 \ \forall\ i = 1, \ldots, Q$ *then the common estimator* $\hat{\tau}_j^2 = \hat{s}_j^2 - \hat{\sigma}_j^2$ *would result.*

Appendix 2

The temporal models were found to be ARIMA(4, 1, 0) and thus

$$\begin{aligned} y_t = y_{t-1} &+ \beta_0 + \beta_1(y_{t-1} - y_{t-2}) \\ &+ \beta_2(y_{t-2} - y_{t-3}) \\ &+ \beta_3(y_{t-3} - y_{t-4}) \\ &+ \beta_4(y_{t-4} - y_{t-5}) + e_t \end{aligned} \tag{A.21}$$

and

$$y_{T+2} = \hat{y}_{T+2} + \nu_{T+2} \tag{A.22}$$

where $\nu_{T+2} = (1 + \beta_1)e_{T+1} + e_{T+2}$. As mentioned in the main text, estimates of the density $\hat{f}(\nu)$ may be *numerically* recovered from the nonparametric estimates of the density $\hat{f}(e)$.

If $\omega = (1 + \beta_1)e$ (hence $\frac{de}{d\omega} = 1/(1 + \beta_1)$), then by the well-known change-of-variables formula

$$f_\Omega(\omega) = f_E(e)\left|\frac{de}{d\omega}\right| = f_E(e)\frac{1}{1+\beta_1}. \tag{A.23}$$

Now consider the joint density of e and ω. We know they are independent because e_t and e_{t-1} are independent. Hence, $f_{E,\Omega}(e, \omega) = f_E(e) \times f_\Omega(\omega)$. Now let $\nu = e + \omega$ and consider the joint density $f_{E,V}(e, \nu)$. Obviously both transformations are monotone and thus one to one and onto. Thus,

$$f_{E,V}(e, \nu) = f_E(e)\, f_\Omega(\nu - e) \times 1 \tag{A.24}$$

since the determinant of the Jacobian is 1. Therefore, the marginal density $f_V(\nu)$ is defined as

$$f_V(\nu) = \int_{-\infty}^{\infty} f_{E,V}(e, \nu)de = \int_{-\infty}^{\infty} f_E(e) \times f_\Omega(\nu - e)de \tag{A.25}$$

which we must calculate numerically. Recall, however, that we desire to have the mean and variance of the estimated conditional yield densities to be determined by the county data only. The variance of the estimated density, $\hat{f}_E(e)$, is

$$\frac{h^2 k_2 \sum_{i=1}^{T}\lambda_i^2}{T} + \frac{T-1}{T}s_E^2.$$

Therefore, we transform the estimated density by the scalar

$$\sqrt{\frac{s_E^2}{\frac{h^2 k_2 \sum_{i=1}^{T}\lambda_i^2}{T} + \frac{T-1}{T}s_E^2}}$$

to get $\tilde{f}_E(e)$ with mean zero and variance s_E^2 and calculate various estimates of $f_V(\nu)$ numerically from our various estimates of $f_E(e)$.