Journal of Agricultural and Resource Economics 23(2):427–444

On Choosing a Base Coverage Level for Multiple Peril Crop Insurance Contracts

Alan P. Ker and Keith H. Coble

For multiple peril crop insurance, the U.S. Department of Agriculture's Risk Management Agency estimates the premium rate for a base coverage level and then uses multiplicative adjustment factors to recover rates at other coverage levels. Given this methodology, accurate estimation of the base coverage level is critical. Currently under consideration is a change in the base coverage level from 65% to 50%. The purpose of this analysis was to provide some insight into whether such a change should or should not be carried out. Not surprisingly, our findings indicate that the higher coverage level should be maintained as the base.

Key words: base coverage levels, crop insurance, MPCI rating

Introduction

This study seeks to analyze a potential policy change by the U.S. Department of Agriculture's Risk Management Agency (RMA) in its approach to setting multiple peril crop insurance (MPCI) premium rates. MPCI enables agricultural producers to purchase insurance against realizations below target yield levels. These target yield levels, more commonly termed coverage levels, are 50%, 55%, 60%, 65%, 70%, and 75% of the expected producer yield. For example, if the expected yield for a producer of irrigated corn is 150 bushels/acre, then a contract purchased at the 50% coverage level insures against a realization below 75 bushels/acre (0.5 × 150 bushels/acre = 75 bushels/acre), while a contract at the 65% coverage level insures against a realization below 97.5 bushels/acre (0.65 × 150 bushels/acre = 97.5 bushels/acre).

RMA estimates the premium rate for a *base* coverage level and then uses multiplicative adjustment factors to recover rates at other coverage levels. Currently, the 65% coverage level is used as the *base*. Consider the following example. For a given producer, assume the estimated premium rate at the 65% coverage level is 2.00% and the multiplicative adjustment factors for the 50% and 75% coverage levels are 0.25 and 2.00, respectively. The premium rate for the 50% coverage level is derived as 2.00% × 0.25 = 0.50%. Had the estimated rate at the 65% coverage level been 1.00%, the premium rate for the 50% coverage level would be 1.00% × 0.25 = 0.25%. Similarly, the rate for the 75% coverage level for the two cases would be 2.00% × 2.00 = 4.00%, and 1.00% × 2.00 = 2.00%, respectively.

Ker is an assistant professor in the Department of Agricultural and Resource Economics, University of Arizona; Coble is an assistant professor in the Department of Agricultural Economics, Mississippi State University.

The authors thank two anonymous reviewers for their constructive and helpful comments.

The actuarially fair premium rate for a crop insurance contract that guarantees a percentage, say λ, of the expected yield, say y^e, is given as:

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

where $0 \leq \lambda \leq 1$, the expectation operator and probability measure are taken with respect to the conditional yield density $f(y \mid \mathcal{F}_t)$, and $\mathcal{F}_t$ is the minimal σ-field generated by the information known at time of rating. Note that the premium rate is expressed as the ratio of the expected indemnity to the total insurer liability. As such, it represents a percentage of total insurer liability.

Given the methodology employed by RMA in setting MPCI rates, accurate estimation of the *base* coverage level is critical.[1] Currently under consideration is a change in the *base* coverage level from the 65% level to the 50% level. The objective of this study is to provide some insight into whether such a change should or should not be carried out. To that end, we undertake the following analysis. First, the conditional yield densities are estimated for two counties, followed by an outline of the estimation approach. From the estimated conditional yield densities, we simulate sequences of independent or exchangeable yields and estimate the empirical rates at both coverage levels for each sequence. Given the empirical rate at one coverage level, we are able to derive the associated rate for the other coverage level using the multiplicative adjustment factor. This enables us to calculate the mean squared error at both coverage levels for both the estimated empirical rates and the corresponding derived rates. We then provide a discussion of the simulation approach and results, followed by the concluding section which highlights our recommendation regarding the potential policy change and summarizes the analysis.

Spatio-temporal Process of Yields

As stated above, we estimate conditional yield densities for two counties. The U.S. Department of Agriculture/National Agricultural Statistical Service (USDA/NASS) yield series for all-practice corn from Shelby County in west-central Iowa and wheat from south-central Kingman County in Kansas are chosen for the empirical example. It is felt that these counties reflect not only a diversity of crops and regions, but also the extremes of conditional yield densities—one severely skewed, the other approximately symmetric. We use county-level yield data rather than farm- or unit-level yield data because of availability. Each of the series contained observations for the 1957–96 crop years.

Basically, yields follow a spatio-temporal process. By averaging over some spatial region (field, farm, or county) and conditioning on the temporal process, the conditional mean yield density for that given space at a point in time is recovered. As noted, the spatial region of interest for our analysis is the county. The discussion in this section is presented in two parts: (*a*) the spatial process of yields, and (*b*) the temporal process of

[1] This study does not focus on the obvious problem of assuming identical scale adjustment factors for each producer-crop combination.

yields. The spatial process of yields is considered first because of its implications regarding distributional assumptions in modeling the temporal process of yields.

Spatial Process of Yields

Consider that yields come from one of two distinct subpopulations, a catastrophic subpopulation and a noncatastrophic subpopulation. That is, in years when a catastrophic event occurs (such as a drought, flood, freeze, etc.), yields are drawn from the catastrophic subpopulation. Conversely, in years when a catastrophic event does not occur, yields are drawn from the noncatastrophic subpopulation. Thus, conditional yields may best be modeled as a mixture of two unknown distributions where the secondary distribution (from catastrophic years) lives on the lower tail of the primary distribution (from noncatastrophic 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.

We conjecture that yields are highly dependent across space with respect to the subpopulation from which they are drawn. For example, if a producer experiences a drought, it is highly likely a neighboring producer also experiences a drought. On the other hand, given which subpopulation is realized, we would expect that yields are only mildly dependent across space.[2] Under this conjecture, counties may represent enough land mass such that Central Limit Theorems (CLTs) for spatially dependent processes may be appealed to.[3] That is, *conditional* spatial dependence dies off at a sufficiently quick rate such that CLTs for dependent processes may be invoked.[4] Therefore, mean yields for a county would be a mixture of two Gaussians where the secondary Gaussian (catastrophic subpopulation) lives on the lower tail of the primary Gaussian (noncatastrophic subpopulation) and has significantly less mass.

There is empirical evidence to suggest conditional yield densities may be modeled as a mixture of two Gaussians. Ker (1996) tests normality in the set of Iowa counties for all-practice corn. At a 5% significance level, normality is rejected in 31% of the counties. However, if catastrophic yield realizations are purged, normality is rejected in only 7% of the counties. Thus, conditional yields appear to be Gaussian in noncatastrophic years, suggesting land mass in most Iowa counties is sufficient for *conditional* spatial dependence to die off and CLTs for dependent processes to apply. Unfortunately, there are insufficient catastrophic yield realizations per county (≤ 5) to test normality of the secondary distribution (catastrophic subpopulation). With such limited realizations, testing normality against reasonable alternatives has very low power. Therefore, although *conditional* spatial dependence dies off sufficiently quickly in noncatastrophic years, there is no evidence to suggest this is or is not the case in catastrophic years.

In addition to the above evidence, visual inspection of kernel estimates of numerous yield densities suggests they may be represented by restrictions on the parameter space

---

[2] See Ker and Goodwin (1998a) for evidence supporting this conjecture.

[3] For a thorough review of CLTs for dependent processes including m-dependence, mixing, and Martingale differences as well as the functional CLT, the reader is directed to either Davidson or Hendry.

[4] Conditional spatial dependence refers to the spatial dependence *given* which subpopulation is realized.

of a mixture of two Gaussians. The Gaussian mixture as conjectured necessarily leads to a negatively skewed and possibly bimodal density. Thus, the Gaussian mixture is consistent with the overwhelming empirical evidence of negative skewness for all-practice corn yield data.

In empirical applications, most researchers have used the beta distribution rather than a mixture distribution (Babcock and Hennessy; Coble et al.; Lee, Harwood, and Somwaru; Borges and Thurman; Kenkel, Busby, and Skees; Nelson; Nelson and Preckel). Ker and Coble 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 employed an inverse hyperbolic sine transformation to model yields. Ker (1997) used a mixture of two Gaussians.

Temporal Process of Yields

The temporal process of yields is governed by two main factors: the current state of technology and the weather. Many analysts have used a stochastic trend to model the changing state of technology for good reason. Moss and Shonkwiler point out that technological innovations and the adoption of those technical innovations are random events. Technological innovations may be considered a Poisson process where each innovation has a distribution surrounding the magnitude of its effect on crop yields. The adoption of such technologies is neither instantaneous nor necessarily complete in the sense that many technologies are never 100% adopted because of newer innovations.

Myers and Jayne provide a very interesting regime shift-diffusion model to estimate yield trend for maize yields in Zimbabwe. Not surprisingly, it was necessary to specify a priori the shape of the diffusion path and constrain that path to be identical across innovations. Moss and Shonkwiler used a Kalman filter model which nests a deterministic trend model inside a stochastic trend model. Many others have used either IMA $(1, q)$ models or the error correction form of those models (Goodwin and Ker; Ker and Goodwin 1998b; Ker 1997; Bessler). One problem with the above models is that the estimated trend does not belong to the class of nondecreasing functions. If we are attempting to estimate the current state of technology, and technology is a nondecreasing set, then the true function would, in most situations, belong to the class of nondecreasing functions.[5]

Many of the above approaches minimized the L_2 norm.[6] Only Ker (1997), Borges and Thurman, and Moss and Shonkwiler employed likelihood methods which explicitly dealt with nonnormality. Moss and Shonkwiler used an inverse hyperbolic sine transformation to account for differences between third and fourth Gaussian population moments and the third and fourth sample moments. Although their model does not allow for bimodality, given their more aggregate data (U.S. corn yields), this is not of great concern. Spatial aggregation should be sufficient such that there is enough mixing between the catastrophic subpopulation and the noncatastrophic population.[7] However,

[5] Clearly, there exist circumstances where the function should not be restricted to the nondecreasing class. One need only consider peanut and tobacco yields.

[6] See Goodwin and Ker for justification of minimizing the L_2 norm for the insurance application despite nonnormality.

[7] As posited above, consider that yields come from one of two subpopulations. As spatial area increases, the variance of each subpopulation must necessarily decrease (assuming spatial correlation is not perfect), causing a greater tendency for bimodality. However, a secondary and more important implication of increasing spatial aggregation is mixing between the

for county yield data, bimodality is a possibility. Borges and Thurman used the beta distribution. This also does not allow for bimodality.

Finally, Ker (1997) used a mixture of two Gaussians to model county yields for all-practice corn in crop yields. In that study, yields were modeled using semi-nonparametric maximum likelihood methods (Hermite series expansions) with a Gaussian mixture for the innovations. Although this allows for bimodality, and although CLTs for spatially dependent processes suggest a Gaussian for the primary or noncatastrophic distribution, there is little evidence to suggest or condemn a Gaussian for the secondary or catastrophic distribution. Note that accurate estimation of the secondary distribution is crucial for rating contracts. Recall, an insurance contract truncates the lower tail of the conditional yield density. The premium rate for that contract is derived from the truncated tail. The majority of mass in the truncated tail is from the secondary distribution, and thus the majority of the rate is derived from the secondary distribution. Therefore, it is crucial to estimate the secondary distribution with a high level of confidence.

A major concern with the beta, gamma, and hyperbolic sine distributions is that they do not allow bimodality. Concern also arises with the Gaussian mixture in that there is little empirical evidence to either support or condemn its use. Given the above problems associated with the parametric forms, we do not feel comfortable specifying a likelihood. A viable alternative is nonparametric regression. The nonparametric approach has many advantages for our analysis. First and foremost, we do not assume the conditional yield distribution is a restriction on the parameter space of a known (with probability one) distributional family. Second, the nonparametric approach allows for the randomness of technological innovations and their adoption without constraints.[8]

Nonparametric Regression—The Isotonic Robust Super Smoother

In this section, we introduce a new approach to estimating the technology component of yields. There are many types of smoothing or nonparametric regression methods (e.g., kernel smoothing, local regression, splines, super smoothing). A kernel smoother was not used because of the nonstationarity of the yield data. Essentially, the kernel-smoothed estimate of the dependent variable at a given value for the independent variable, say X_0, is the weighted mean of all dependent realizations with weights decreasing as the distance between X_0 and the other values of the independent variable increase. Thus, if the data are nonstationary in that there exists a positive trend component, then predictions will tend to be negatively biased and lag behind. This may be fixed by smoothing with a trend line rather than the kernel smoother. Locally

---

←

subpopulations. Mixing between the two subpopulations will decrease the tendency toward bimodality. If the spatial process of yields is α-mixing, then CLTs for dependent processes suggest the mixing between the two subpopulations will begin to dominate, leading to a unimodal density and eventually a single Gaussian density. Clearly, the results of Moss and Shonkwiler suggest that the U.S. is not sufficiently large for the spatial mixing between the two subpopulations to result in a Gaussian density.

[8] Of course, we cannot recover the adoption cycle as do Myers and Jayne.

weighted regression smoothing estimates a different regression line at each realization (y_t, x_t). The predicted value $(\hat{y}_0)$ is recovered using weighted least squares within a neighborhood of x_0. Again, the weights are defined as a decreasing function of distance between X_0 and the other values of the explanatory variables which belong to the predetermined neighborhood. The size of the neighborhood is determined by cross-validation methods. Keeping the size of the neighborhood constant across all realizations is not optimal if the error variance or the curvature of the underlying function varies over the range of the independent variable (Cleveland). An increase in the variance requires a larger neighborhood, whereas an increase in the curvature requires a smaller neighborhood. Local cross-validation, employed to recover the size of the neighborhood, avoids this problem by choosing a different size neighborhood for each value of the independent variable x_t based on local cross-validation. Using local cross-validation to determine the neighborhood at each value of the independent variable is termed "super smoothing."

As with locally weighted least squares, the super smoother minimizes, locally, a weighted least squares criterion. However, given that yields are not believed to be Gaussian, even locally, we have not overcome our distributional problems. As such, we augment the super smoother by using robust techniques, specifically an m-estimator. Although m-estimators are susceptible to high leverage points (outliers in the independent variable), they are robust to outliers in the dependent variable. Given the independent variable is time (consequently we are assured of no outliers), m-estimators are a viable option for our application.[9] If one considers the secondary or catastrophic distribution as contaminating the primary distribution, and assumes the CLTs for dependent processes result in the primary distribution being Gaussian, then the m-estimators should model yields very well. This type of situation is what m-estimators were designed to accommodate: a Gaussian distribution with a contaminant.

Unlike standard super smoothing, using an m-estimator inside the super smoother is not a canned procedure in the S-PLUS computer program. After considering various weighting functions for m-estimators (Andrews, Huber, and bisquare), we found no reason to deviate from the default S-PLUS m-estimator for robust regression. Thus, we employ the Huber m-estimator until convergence and then perform two iterations of the bisquare. We use this m-estimator not just in the final estimates, but in the local cross-validation as well.

We wish to isotonize or constrain the smoothed estimate to belong to the class of non-decreasing functions. Therefore, we isotonize it using the pool-adjacent-violators (PAV) algorithm in Hanson, Pledger, and Wright. Hildenbrand and Hildenbrand employed this approach in nonparametric estimation of Engel curves. The isotonic robust super smoother is fully delineated in Appendix A.[10]

The resulting estimates of the temporal processes of yields for Shelby County corn and Kingman County wheat are illustrated in figures 1 and 2, respectively. The raw yield data also are plotted. Not surprisingly, the figures support the assertions of Ker and Goodwin (1998b); Goodwin and Ker; Myers and Jayne; and Moss and Shonkwiler that yield trend is erratic and lumpy.

---

[9] See Hardle for a kernel m-smoother using the Fast Fourier Transform, and Cleveland using robust techniques with locally weighted regression techniques.

[10] The SAS-IML code for the isotonic robust super smoother is available from the lead author by request.

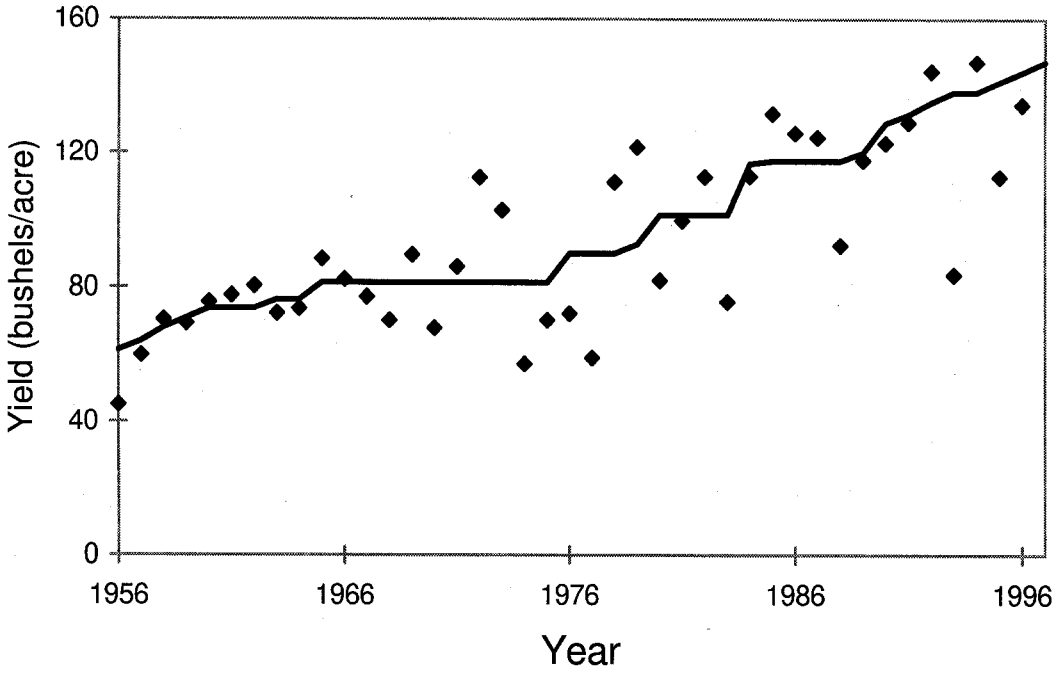

Figure 1. Isotonic robust super smoother: Shelby County, Iowa, all-practice corn

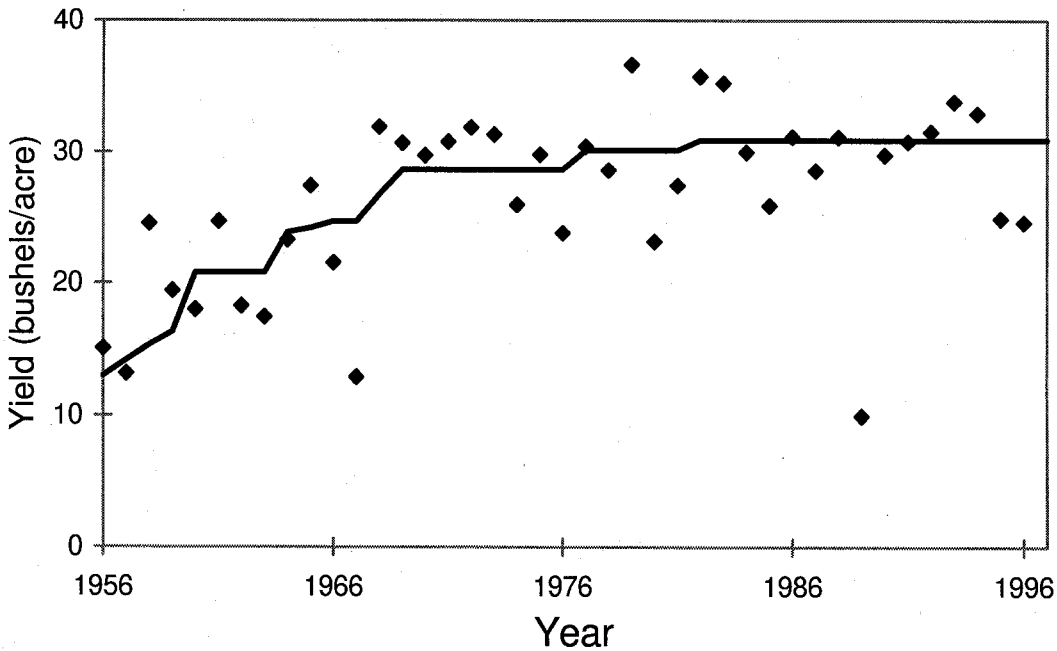

Figure 2. Isotonic robust super smoother: Kingman County, Kansas, wheat

Heteroskedasticity Considerations

Using time-series data to estimate the conditional yield densities is complicated by the possibility that crop yield variance has changed over time. For example, yield variance may increase with yield levels, suggesting yields have a constant coefficient of variation. Many tests for heteroskedasticity are available. However, most assume normality. Thus, we employ the Goldfeld-Quandt nonparametric peak test for heteroskedasticity. Since rank or count type tests only exploit the ordinal properties of the data, they are more generally applicable but less powerful.

The Goldfeld-Quandt nonparametric peak test is no exception. Under the null of homoskedasticity, the p-value for the unstandardized innovations ($\hat{e}$) for Shelby corn is 16.67%, while the p-value for the standardized innovations ($\hat{e}/\hat{y}$) is 50%. For Kingman wheat, the p-value for the unstandardized innovations is 4.167%, while the p-value for the standardized innovations is 50%. As a result, we construct a sequence of identically distributed exchangeable yields from the unknown density $f(y|\mathcal{F}_t)$ for each county using the standardized innovations rather than the unstandardized innovations. Therefore, the sequence is:

$$\hat{y}_{97,t} = \left(\frac{\hat{e}_t}{\hat{y}_t}\right) \times \hat{y}_{97} + \hat{y}_{97} \quad \forall\, t = 1, 2, \ldots, 40, \tag{2}$$

where $\hat{y}_{97}$ is the predicted value for 1997, $\hat{e}_t$ is the unstandardized innovation at realization t, and $\hat{y}_t$ is the fitted value at realization t. The predicted value $\hat{y}_{97}$ is recovered using the parameter estimates from the robust super smoother at the final realization. Consequently, we obtain an exchangeable realization from $f(y|\mathcal{F}_t)$ for each standardized innovation.

Estimation of Conditional Yield Distributions

We employ nonparametric kernel methods to estimate the conditional yield densities. Nonparametric kernel methods have been sparingly used in the agricultural economics literature (see Ker and Goodwin 1998b; Goodwin and Ker; and Turvey and Zhao). Univariate kernel density estimation is very intuitive. A required input of the kernel density estimator is a set $(y_1, y_2, \ldots, y_T)$ of exchangeable or independent realizations from the unknown density of interest f_Y. Oversimplifying, 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. We use an adaptive kernel estimator to recover the conditional yield densities. Appendix B details the methodology.

Figures 3 and 4 illustrate the estimates of the conditional yield densities for Shelby corn and Kingman wheat, respectively. The conditional yield density for Shelby corn appears to be a negatively skewed, possibly bimodal density. In contrast, the conditional yield density for Kingman wheat is mildly skewed and unimodal.

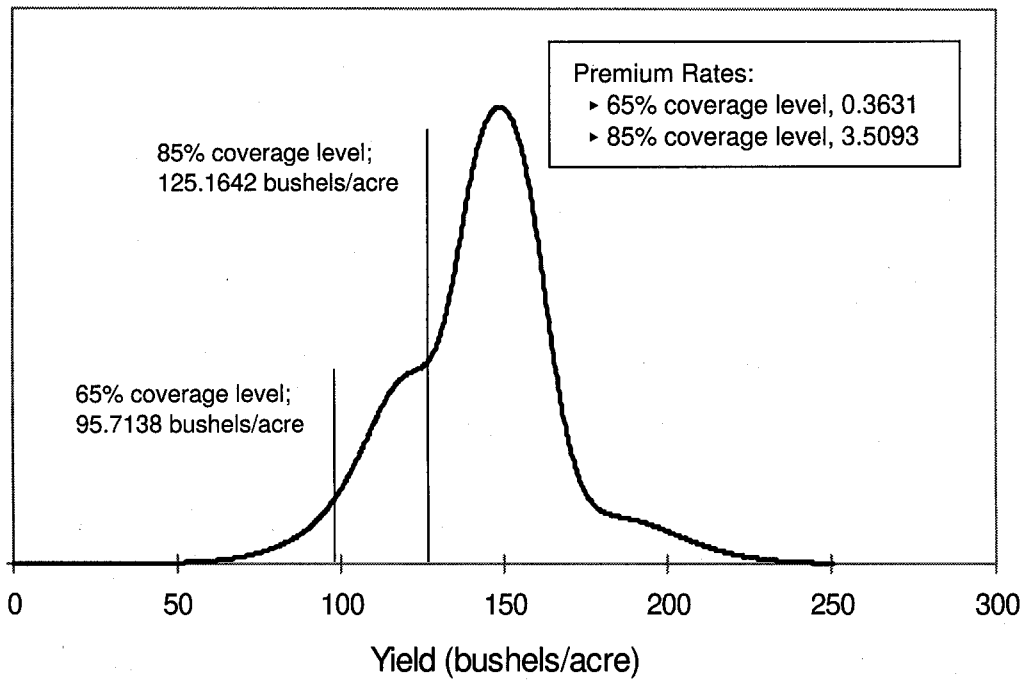

Figure 3. Adaptive kernel estimate of the conditional yield density: All-practice corn, Shelby County, Iowa

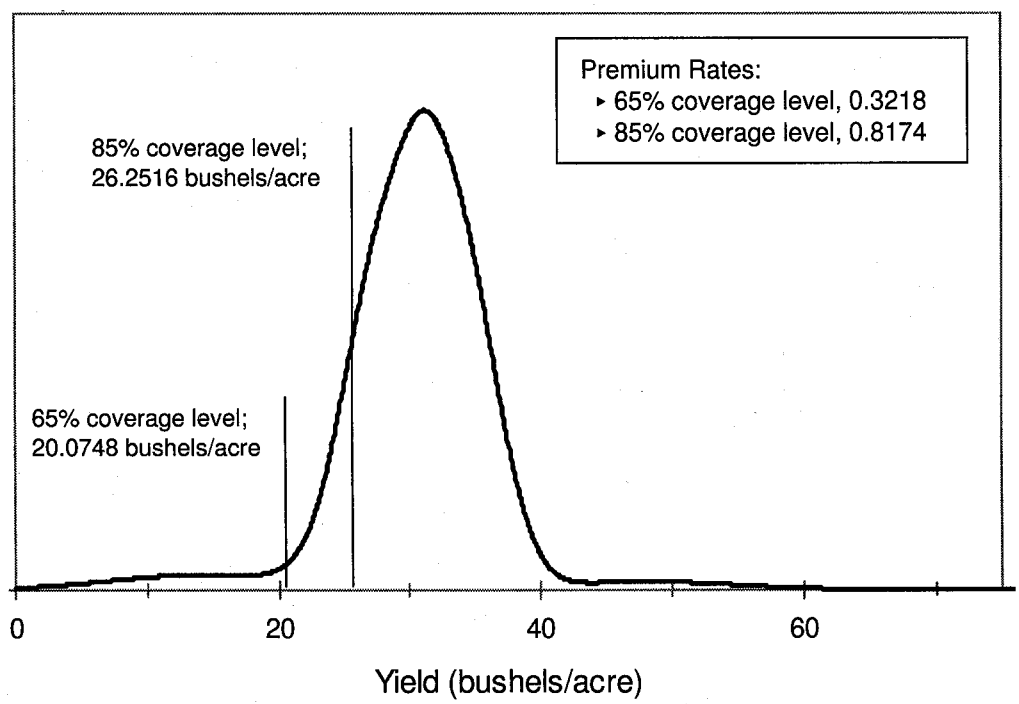

Figure 4. Adaptive kernel estimate of the conditional yield density: Wheat, Kingman County, Kansas

Simulation Analysis

The objective of this section is to evaluate the two strategies currently under consideration by the RMA. Recall that the agency currently is considering changing the *base* coverage from the 65% level to the 50% level. We evaluate the two strategies by simulating sequences of exchangeable yields from the estimated conditional yield densities for Shelby County corn and Kingman County wheat. As noted earlier, data availability made it necessary to use the less variable county yield data rather than farm or unit yield data. Therefore, we use 65% and 85% county-level premium rates to represent the 50% and 65% farm-level rates.[11] For each sequence, we estimate the empirical rates at the 65% and 85% coverage levels. In turn, we derive rates for one coverage level assuming the other is the *base*. Finally, we consider the mean squared error of the estimated and derived rates at both coverage levels.

By simulating from the estimated conditional yield densities, we are treating these as the population or true densities. As such, we are able to numerically integrate and recover the true premium rates. For Shelby corn, the true rate at the 65% coverage level is 0.3631, while for the 85% coverage level it is 3.5093. Therefore, when the 65% coverage level is used as the *base*, the derived rate at the 85% coverage level is $9.6649 \times er65$, where $er65$ is the estimated empirical rate at the 65% coverage level, and the scale factor is calculated as the ratio of the true rate at the 85% coverage level to the true rate at the 65% coverage level ($9.6649 = 3.5903 \div 0.3631$). Conversely, when the 85% coverage level is used as the *base*, the derived rate at the 65% coverage level is $0.1035 \times er85$, where $er85$ is the estimated empirical rate at the 85% coverage level, and the scale factor is calculated as the ratio of the true rate at the 65% coverage level to the true rate at the 85% coverage level ($0.1035 = 0.3631 \div 3.5903$).

For Kingman wheat, the true rate at the 65% coverage level is 0.3218, while for the 85% coverage level it is 0.8174. Therefore, when the 65% coverage level is used as the *base*, the derived rate at the 85% coverage level is $2.5401 \times er65$, where $er65$ is the estimated empirical rate at the 65% coverage level, and the scale factor is calculated as the ratio of the true rate at the 85% coverage level to the true rate at the 65% coverage level ($2.5401 = 0.8174 \div 0.3218$). Conversely, when the 85% coverage level is used as the *base*, the derived rate at the 65% coverage level is $0.3937 \times er85$, where $er85$ is the estimated empirical rate at the 85% coverage level, and the scale factor is calculated as the ratio of the true rate at the 65% coverage level to the true rate at the 85% coverage level ($0.3937 = 0.3218 \div 0.8174$).

From each density, we simulate 10,000 sequences of exchangeable yields. Each sequence is of length 10. At first glance, this may appear to be an extremely small number of yields sampled for each simulation. However, 10 years of yield data more closely represent the maximum rather than average number of yields from which MPCI rates are estimated. Rather than use the 10 simulated yields to estimate a parametric or nonparametric density from which a rate may be recovered, we follow MPCI procedures and calculate the empirical rate at the 65% and 85% coverage levels. The simulation results are summarized in table 1.

---

[11] Since the county density is significantly less disperse than the individual farm density, the empirical rate at the 50% coverage level for the county is equal to zero with probability very close to one. This is not true at the 50% coverage level for the individual farm. Therefore, we used the 65% and 85% county-level premium rates to represent the 50% and 65% farm-level rates.

Table 1. Simulation Results for Conditional Yield Densities

Description	MPCI Base Premium Rate	
	65% Coverage Level	85% Coverage Level
Shelby County, Iowa—CORN		
True Rate	0.3631	3.5093
Empirical Rate:		
▸ Mean	0.1659	3.7883
▸ Standard Deviation	0.3310	2.6432
▸ Mean Squared Error	0.1485	7.0646
Derived Rate:		
▸ Mean	0.3920	1.6037
▸ Standard Deviation	0.2735	3.1991
▸ Mean Squared Error	0.0756	13.8656
Kingman County, Kansas—WHEAT		
True Rate	0.3218	0.8174
Empirical Rate:		
▸ Mean	0.3437	0.7893
▸ Standard Deviation	0.5234	0.9280
▸ Mean Squared Error	0.2744	0.8620
Derived Rate:		
▸ Mean	0.3107	0.8730
▸ Standard Deviation	0.3654	1.3295
▸ Mean Squared Error	0.1336	1.7706

For Shelby corn, the estimated rates at the 65% coverage level have a mean squared error of 0.1485. Note, however, if the 85% coverage level is used as a *base* to derive the rates at the 65% coverage level, the mean squared error is reduced to 0.0756. Conversely, the estimated rates at the 85% coverage level have a mean squared error of 7.0646, while the rates derived using the 65% coverage level as the *base* inflate the mean squared error to 13.8656. Clearly, this suggests that the higher rather than lower coverage level should be used as the *base*.

The results for Kingman wheat parallel those for Shelby corn. The estimated rates at the 65% coverage level have a mean squared error of 0.2744, whereas if the 85% coverage level is used as a *base* to derive the rates at the 65% coverage level, the mean squared error is reduced to 0.1336. In contrast, the estimated rates at the 85% coverage level have a mean squared error of 0.8620, while the rates derived using the 65% coverage level as the *base* inflate the mean squared error to 1.7706. Simulations from both conditional yield densities strongly suggest that the higher rather than lower coverage level should be used as the *base*.

The intuition is relatively simple if one considers a nonparametric estimator. As the coverage level is increased ($\lambda \uparrow$), the truncation or conditioning point (λy^e) moves away from the lower tail of the conditional density $f(y \mid \mathcal{F}_t)$ toward its center (y^e). Thus, more of the realizations are used to estimate both the probability of a loss and the density

conditional on a loss occurring. As a result, the estimated premium rate at the higher coverage level is relatively more accurate. Although less intuitive, this holds true for most parametric distributions; tail probabilities tend to be relatively more sensitive to changes in the parameters indexing the distribution than non-tail probabilities.

Conclusions

Given the methodology employed by RMA in setting MPCI rates, accurate estimation of the *base* coverage level is critical. Currently under consideration by the RMA is a change in the *base* coverage level for MPCI from the 65% coverage level to the 50% coverage level. Our objective was to provide some insight into whether such a change should or should not be carried out. To that end, we undertook the following simulation analysis.

From two estimated conditional yield densities, we simulated sequences of exchangeable yields and estimated the empirical rates at both coverage levels for each sequence. Given the empirical rate at one coverage level, we derived the associated rate for the other coverage level using the initial level as the *base*. This enabled us to calculate the mean squared error at both coverage levels for both the estimated empirical rates and the corresponding derived rates. Not surprisingly, our findings suggest that the higher coverage level should be maintained as the *base*. Therefore, we recommend that the USDA Risk Management Agency not change its *base* coverage level from the 65% to the 50% coverage level.

[*Received October 1997; final revision received August 1998.*]

References

Abramson, I. S. "On Bandwidth Variation in Kernel Estimates—A Square Root Law." *Ann. Statis.* 10(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.

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

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

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

Cleveland, W. S. "Robust Locally Weighted Regression and Smoothing Scatter Plots." *J. Amer. Statis. Assoc.* 74(1979):829–36.

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.

Davidson, J. *Stochastic Limit Theory.* New York: Oxford University Press, 1994.

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(1987):798–803.

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

Goodwin, B. K., and A. P. Ker. "Rating Group Risk Crop Insurance Premiums with Nonparametric Yield Densities." *Amer. J. Agr. Econ.* 80(1998):139–53.

Hanson, D. L., G. Pledger, and F. T. Wright. "On Consistency in Monotonic Regression." *Ann. Statis.* 1(1973):401–21.

Hardle, W. "Resistant Smoothing Using the Fast Fourier Transform." *Appl. Statis.* 36(1987):104–11.

Hendry, D. F. *Dynamic Econometrics.* New York: Oxford University Press, 1995.

Hildenbrand, K., and W. Hildenbrand. "On the Mean Income Effect: A Data Analysis of the U.K. Family Expenditure Survey." In *Contributions to Mathematical Economics*, eds., W. Hildenbrand and A. Mas-Colell. New York: North-Holland, 1986.

Huber, P. J. "Robust Smoothing." In *Robustness in Statistics*, eds., E. Launer and G. Wilkinson, pp. 33–47. New York: Wiley, 1979.

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

Ker, A. P. "Rating and Yield Predicting Procedures for the Group Risk Federal Crop Insurance Program: A Nonparametric Approach." Unpub. Ph.D. diss., Dept. of Econ. and Dept. of Statis., North Carolina State University, Raleigh, 1996.

———. "Using SNP Maximum Likelihood Techniques to Recover Conditional Densities: A New Approach to Recovering Premium Rates." Work. pap., Dept. of Agr. and Resour. Econ., University of Arizona, Tucson, 1997.

Ker, A. P., and K. Coble. "On the Use of the Beta Distribution in Modeling Yields." Work. pap., Dept. of Agr. and Resour. Econ., University of Arizona, Tucson, 1997.

Ker, A. P., and B. K. Goodwin. "On the Correlation Between Yields and Prices." Work. pap., Dept. of Agr. and Resour. Econ., University of Arizona, Tucson, 1998a.

———."Nonparametric Estimation of Crop Insurance Rates Revisited." Work. pap., Dept. of Agr. and Resour. Econ., University of Arizona, Tucson, 1998b.

Lee, H., J. Harwood, and A. Somwaru. "Implications of Disaster Assistance Reform for Noninsurable Crops." Paper presented at annual meetings of the American Agricultural Economics Association, Indianapolis IN, August 1995.

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

Myers, R. J., and T. Jayne. "Regime Shifts and Technology Diffusion in Crop Yield Growth Paths with an Application to Maize Yields in Zimbabwe." *Austral. J. Agr. and Resour. Econ.* 41(1997):285–303.

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

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

Rosenblatt, M. "Curve Estimates." *Ann. Math. and Statis.* 42(1971):1815–42.

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

Turvey, C. G., and C. Zhao. "Parametric and Nonparametric Crop Yield Distributions and Their Effects on All-Risk Crop Insurance Premiums." Work. pap., Dept. of Agr. Econ. and Bus., University of Guelph, Guelph, Ontario, 1993.

U.S. Department of Agriculture, National Agricultural Statistics Service. Various published and unpublished crop yield series data for Shelby County, Iowa, and Kingman County, Kansas. USDA/NASS, Washington DC, 1957–96.

Appendix A: The Isotonic Robust Super Smoother

For completeness, we start by outlining locally weighted regression smoothing. For any predictor value, say x_0, locally weighted regression smoothing is calculated as follows:

i. For any given point, say x_0, find the k nearest neighbors of x_0. We define the set of k nearest neighbors of x_0 as a neighborhood $N_k(x_0)$. The span is defined as the ratio of the cardinality of the neighborhood set to the cardinality of the entire sample. For our purposes, it is simpler just to consider the span as the number of points in the neighborhood, k.

ii. Define $d\max(x_0) = \max_{N_k(x_0)}(|x_0 - x_i|)$, and assign the following weights,

$$W\left(\frac{|x_0 - x_i|}{d\max(x_0)}\right),$$

where

$$W(u) = \begin{cases} (1 - u^3)^3 & \text{for } 0 \le u < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{A1}$$

iii. Calculate the weighted least squares for the neighborhood $N_k(X_0)$ only.

Generally k, or the span, is determined using cross-validation methods. Cross-validation methods, using the "leave-one-out" approach, choose the span which minimizes the following sum:

$$\sum_{i=1}^{n}\left[y_i - \hat{y}_{(i)}^{k}\right]^2, \tag{A2}$$

where $\hat{y}_{(i)}^{k}$ is the weighted least squares estimate of y_i after excluding (y_i, x_i) using span k. As noted in the main text, a constant span across the sample is not optimal if the error variance or the curvature of the underlying function varies over the range of x. An increase in the variance requires an increase in the span, whereas an increase in the curvature requires a decrease. Local cross-validation avoids this problem by choosing a span for the predictor values x_0 based only on the neighborhood $N_\tau(x_0)$. That is, for each predictor value x_0, the span is chosen according to the following sum:

$$\sum_{N_\tau(x_0)}\left[y_i - \hat{y}_{(i)}^{k}\right]^2, \tag{A3}$$

where $\hat{y}_{(i)}^{k}$ is the weighted least squares estimate of y_i after excluding (y_i, x_i) with span k. Note that the sum is only over $N_\tau(x_0)$ rather than over the entire sample. This is repeated for each realized value of the explanatory variable (x_i), thus leading to a separate span (k_i) for each. Note that an overall span must be specified to define the $N_\tau(X_0)$ to undertake the local cross-validation. Also note that the local span need not be bounded above by the overall span. The overall span simply dictates the neighborhood for the summation so as to recover the local cross-validation sum of squares.

As stated in the main text, the super smoother minimizes, locally, a weighted least squares criterion. However, given yields are not believed to be Gaussian, even locally, we have not overcome our distributional problems. Thus we augment the super smoother by incorporating robust techniques, specifically an m-estimator. After considering various weighting functions for m-estimators (Andrews, Huber, and bisquare), we found no reason to deviate from the default S-PLUS m-estimator for robust regression—i.e., the Huber m-estimator iterated until convergence and then followed by two iterations of the bisquare. We use the m-estimator not in just the final estimates, but in the local cross-validation as well. The procedure for the m-estimator, which uses iterative reweighted least squares, is as outlined below:

i. Given W, $N_\tau(x_0)$, and k, estimate $\hat{\beta}$ using weighted least squares.

ii. Given the estimated residuals, calculate the mean absolute deviation (MAD) estimator and divide the vector of residuals by the MAD. Define u_i as the absolute value of residual i divided by MAD.

iii. Recover the Huber weights, Ω, where Ω is defined as:

$$\Omega(u) = \begin{cases} 1 & \text{for } u < 1.345, \\ \dfrac{1.345}{u} & \text{otherwise.} \end{cases} \tag{A4}$$

iv. Go to step (i) using new weights ΩW until convergence is obtained.

v. After convergence, use the estimated residuals to define the bisquare weights, Ψ, where Ψ is defined as:

$$\Psi(u) = \begin{cases} \left(1 - \left(\dfrac{u}{4.685}\right)^2\right)^2 & \text{for } u < 4.685, \\ 0 & \text{otherwise.} \end{cases} \tag{A5}$$

vi. Using weights ΨW, calculate weighted least squares estimates. Iterate step (v) once.

Note that the above procedure must be undertaken at each point for all possible spans. The overall span which defines the neighborhood for the local cross-validations must be chosen. In our case, we use an overall span of 15 observations. We use the pool-adjacent-violators (PAV) algorithm to isotonize the robust super smoother. The PAV algorithm is outlined as follows:

i. Starting with $\hat{y}_i$, move right until $(\hat{y}_i, \hat{y}_{i+1})$ violates the monotonicity constraint. Pool $(\hat{y}_i, \hat{y}_{i+1})$ and replace with their average, $\hat{y}_i^* = \hat{y}_{i+1}^* = (\hat{y}_i + \hat{y}_{i+1})/2$.

ii. Check that $\hat{y}_{i-1} \leq \hat{y}_i^*$. If not, pool $(\hat{y}_{i-1}, \hat{y}_i, \hat{y}_{i+1})$ and average. Continue to the left until monotonicity is satisfied. Then proceed to the right.

The estimated isotonized robust super smoothers for Shelby County corn and Kingman County wheat are graphed in text figures 1 and 2, respectively.

Appendix B: Kernel Density Estimation Methodology

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

$$\hat{f}_Y(y_0) = \sum_{i=1}^{T} \frac{K\left(\dfrac{y_0 - y_i}{h}\right)}{Th}, \tag{A6}$$

where h is the smoothing parameter and $K(\cdot)$ is the kernel function. Thus, we have two decisions that must be made: (*a*) choice of the kernel function, and (*b*) choice of the smoothing parameter. The logical criterion on which to base these choices is MISE, which is the function space analogue to MSE and is defined as:

(A7)
$$\begin{aligned} MISE(\hat{f}) &= E\int_{-\infty}^{\infty}(\hat{f}(y) - f(y))^2 dy \\ &= \int_{-\infty}^{\infty}(\hat{f}(y) - f(y))^2 dy \\ &= \int_{-\infty}^{\infty}(E\hat{f}(y) - f(y))^2 dy + \int_{-\infty}^{\infty}\mathrm{var}\hat{f}(y)\,dy. \end{aligned}$$

Thus, MISE is the sum of the integrated squared bias and the integrated variance.

We first consider the kernel function and then proceed with a discussion of the smoothing parameter. Any function that integrates to one may be used as the kernel (nonnegativity is not a necessary condition). Epanechnikov derived the optimal nonnegative kernel function with respect to minimizing 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. In practice, a standard Gaussian kernel is generally used. For our analysis, we employ the standard Gaussian kernel. In practice, however, 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 is the entire real line. We evaluate the densities over a range of plus and minus 10 standard deviations from the mean.

The selection 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. Optimally, the chosen smoothing parameter would minimize MISE. Recall, MISE is composed of two parts: the integrated squared bias and the integrated variance. With manipulation,

(A8)
$$MISE(\hat{f}) = \frac{1}{4}h^4 k_2^2 \int_{-\infty}^{\infty} f''(y)^2 dy + \frac{1}{Th} f(y) \int_{-\infty}^{\infty} K(t)^2 dt,$$

where $k_2 = \int_{-\infty}^{\infty} t^2 K(t)\,dt \neq 0$. With the standard Gaussian kernel, $k_2 = 1$. Note that the smoothing parameter h is inversely related to the variance but directly related to the bias. If one attempts to reduce bias by choosing a small h, the variance increases. Conversely, if a large h is chosen to reduce variance, bias increases. Unfortunately, the optimal h is a function of the unknown density. Thus, we choose the smoothing parameter according to Silverman's rule of thumb:

(A9)
$$\hat{h} = 0.9 \times \left[\text{standard deviation}, \frac{\text{interquartile range}}{1.34}\right] \times T^{-1/5}.$$

This has been found to yield a mean integrated squared 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. Visual inspection of the estimated densities suggests that our densities belong 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 distribution in attempts to identify detail in the main area of the distribution. 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, we employ *adaptive kernel* methods.

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, we use a vector of smoothing parameters with dimension equal to the data rather than a single smoothing parameter. 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.

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 a 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 employ the simple kernel estimate. Denoting the pilot estimate $\breve{f}$, the local scale λ_i is defined as:

$$\lambda_i = \left(\frac{\breve{f}(y_i)}{g} \right)^{-\alpha}, \tag{A10}$$

where $\log(g) = 1/T \sum \log \breve{f}(y_i)$, 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 exchangeable or 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:

$$\hat{f}(y_0) = \sum_{i=1}^{T} \frac{K\left(\frac{y_0 - y_i}{\lambda_i h} \right)}{T \lambda_i h}, \tag{A11}$$

where $\lambda_i h$ 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 of the pilot density. The larger α is, 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 simple 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 so when tail estimation is crucial.

An unfortunate problem with using kernel estimators is that the estimated density does not necessarily have its moments equal to the sample moments. Clearly, the consistency of these estimators indicates this is a finite sample problem. However, in our sample sizes for estimating conditional yield densities, this can be disconcerting. Consider the first two moments of the estimated density using adaptive kernel methods. With respect to the first sample moment, if the kernel is symmetric, the estimated density has a mean equal to the sample mean. Intuitively, the symmetric kernel ensures that the mean of the individual kernel, independent of its smoothing parameter, is the point on which the individual kernel is centered. Given each kernel is equally weighted (they all have equal mass), the mean of the adaptive kernel estimate is equal to the sample mean. Thus, our estimated densities will have means equal to their respective sample means.

With respect to the second sample moment, the estimated density will have variance greater than or equal to the sample variance almost surely. The additional variance of the kernel estimate is rather intuitive. Since the mass $(1/T)$ 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 is large in small samples, while tail probabilities, and thus derived rates, are quite sensitive to changes in variance. In fact, the additional variance is $(O_p(T^{-7/5}))$, which may be nontrivial in samples of our size $(T = 40)$.

Ker and Goodwin (1998b) derive the variance of the estimated density for the adaptive kernel estimator. The variance is:

$$\text{var}(y) = \frac{h^2 \sum_{i=1}^{T} \lambda_i^2}{T} + \frac{T - 1}{T} s^2, \tag{A12}$$

where $y \sim \hat{f}_Y$, and s^2 is the sample variance of the set of exchangeable or independent realizations from f_Y. We adjust our adaptive kernel estimates by taking a scale transformation of the support. That is, we multiply the support by the following:

(A13)
$$\sqrt{\frac{s^2}{\frac{h^2 \sum_{i=1}^{T} \lambda_i^2}{T} + \frac{T-1}{T} s^2}}.$$

In doing this, we force the estimated density to have variance equal to the sample variance. Ker and Goodwin (1998b) prove the transformation. Beirens undertakes a different approach which yields the same result. Rather than taking a scale transformation of the resulting kernel estimate, the data are scaled prior to entering the kernel.