Cardon Research Papers
in Agricultural and Resource Economics

Research Paper 2004-04
September 2004

Rating Crop Insurance Policies with Efficient Nonparametric Estimators That Admit Mixed Data Types

Jeff Racine
Syracuse University

Alan P. Ker
University of Arizona

Department of Agricultural and Resource Economics
College of Agriculture and Life Sciences
The University of Arizona

This paper is available online at http://ag.arizona.edu/arec/pubs/workingpapers.html

RATING CROP INSURANCE POLICIES WITH EFFICIENT NONPARAMETRIC ESTIMATORS THAT ADMIT MIXED DATA TYPES

JEFF RACINE AND ALAN KER

Abstract. The identification of improved methods for characterizing crop yield densities has experienced a recent surge in activity due in part to the central role played by crop insurance in the *Agricultural Risk Protection Act of 2000* (estimates of yield densities are required for the determination of insurance premium rates). Nonparametric kernel methods have been successfully used to model yield densities, however, traditional kernel methods do not handle the presence of categorical data in a satisfactory manner and have therefore tend to be applied at the county level only. By utilizing recently developed kernel methods that admit mixed data types, we are able to model the yield density *jointly* across counties leading to substantial finite-sample efficiency gains. We find that when we allow insurance companies to strategically reinsure with the government based on this novel approach, it becomes quite clear that they accrue significant rents.

Date: September 17, 2004.
Key words and phrases. Yield distributions, insurance rating, kernel estimation, discrete data.
JEL Classification: C14 - Semiparametric and Nonparametric Methods.
The authors would like to thank but not implicate Qi Li, Thomas Knight, and Jack Rader for their thoughtful comments and suggestions. Racine gratefully acknowledges funding from the NSF, Award #3535905.

1. Introduction

Political forces have recently fashioned crop insurance as the cornerstone of U.S. agricultural policy. In 2000, Congress approved the 'Agricultural Risk Protection Act (ARPA) of 2000.' The additional cost of this legislation was estimated to be $8.2 billion over a 5-year period thereby *doubling* the federal budget on crop insurance programs to $16.1 billion. The program, only available to traditional field crops as recently as 1990, currently covers 89 different crops including such non-traditional products as cut flowers, trees and shrubs, and most specialty crops such as avocados, blackberries, etc. ARPA has mandated the expansion of crop insurance in three important dimensions: expanded product coverage including for example livestock products; expanded geographical availability for existing crops; and increasing producer demand by doubling subsidies from approximately 30% to 60% of the premium rate. Recent legislative actions indicate that crop insurance may become the policy instrument of choice to funnel resources to agricultural producers. Given the pivotal role played by crop insurance in U.S. agricultural policy and the substantial resources directed toward the support of agricultural producers, the accurate pricing of crop insurance policies along with precise risk assessment is more important than ever.

The U.S. crop insurance program is somewhat unique among insurance schemes in that three economic interests are served. The federal government through the United States Department of Agriculture's Risk Management Agency (RMA), the private insurance companies, and the farmers, all have vested interests. In 1980, the marketing of crop insurance policies, previously the domain of the RMA, was expanded to include private insurance companies in an attempt to increase farmer participation. While the pricing of the crop insurance policies remains the responsibility of the RMA, insurance companies receive compensation for administrative expenses and share, *asymmetrically*, the underwriting gains and losses of the policies.[1] The Standard Reinsurance Agreement (SRA) stipulates the terms of the

[1]Underwriting gain/loss for a set of policies is the total premium less the total indemnity payments.

sharing of these underwriting gains and losses. The structure of the SRA enables the private insurance companies to retain or cede – ex-ante and subject to constraints – varying portions of the realized underwriting gains or losses of every federally subsidized crop insurance policy it sells.[2]

The unique arrangement between the producers, private insurance companies, and RMA not only provides us with an ideal environment in which to evaluate proposed estimation methodologies but also has important policy ramifications. Under the SRA, an insurance company must decide which policies to retain and which to cede thereby requiring them to construct their own premium rate schedules. For example, consider the case where a farmer chooses to buy crop insurance from a private insurance company at the government mandated price of, say, \$100. The insurance company selling that policy must decide whether to retain or cede the premium and associated liability of that policy. Suppose the insurance company estimates the premium rate for that same policy to be \$90. In this case, a risk neutral insurance company will retain that policy because they expect a profit of \$10. Suppose instead that the insurance company estimates the premium rate for that policy to be \$105. In this case, a risk neutral insurance company will cede that policy.

Therefore, a risk neutral insurance company will act according to the following decision rule: retain the subset of policies for which it expects a profit (insurance company premium rate is less than the RMA premium rate) and cede the subset of policies for which it expects a loss (insurance company premium rate is greater than the RMA premium rate).[3] As a result, the SRA represents an incentive for the RMA to employ the rating methodology which makes the most efficient use of the available data thereby reducing adverse selection activities

[2]In practice, the insurance company can only retain or cede varying portions of the liability and associated premium rate for any insurance policy. See Ker & McGowan (2000) for a detailed discussion of the SRA. For the current analysis, it is sufficient to assume the insurance company can retain or cede 100% of the liability and premium for any given policy. This greatly reduces the complexity of the analysis without loss of generality.

[3]While this may not be economically inefficient as it represents a simple transfer to insurance companies rather than agricultural producers, in a political economy framework this outcome may be undesired. Political rents recovered from the agricultural production sector are likely to be significantly greater than those recovered from the private insurance companies involved in agricultural crop insurance.

by the insurance companies. The appropriate context in which to evaluate any proposed methodology for rating crop insurance policies is to assume the role of an insurance company and determine if significant excess rents can be garnered using the proposed methodology to estimate the premium rates and determine which policies to retain and which to cede.

In this manuscript we investigate a method, recently introduced by Hall, Racine & Li (forthcoming), having the potential to substantially improve the accuracy of premium rate estimates. In Section 2 we briefly review the U.S. crop insurance program and the SRA, outline the construction of premium rates, and discuss the yield data used in our analysis. In Section 3 we outline the RMA rating methodology as well as the Hall et al. (forthcoming) estimator. In Section 4 we undertake an out-of-sample analysis designed to determine whether or not economically and statistically significant excess rents can be garnered using the Hall et al. (forthcoming) estimator. Policy implications and concluding remarks are found in Section 5.

2. Premium Rate Preliminaries

2.1. **The U.S. Crop Insurance Program.** Federally regulated crop insurance programs have been a prominent part of U.S. agricultural policy since the 1930s. In 2004, the estimated number of crop insurance policies exceeded 1.24 million with total liabilities exceeding $45 billion. In the past, crop insurance schemes offered farmers the opportunity to insure against yield losses resulting from nearly all risks, including such things as drought, fire, flood, hail, and pests. For example, if the farmer's expected wheat yield is 30 bushels per acre ($y^e = 30$), a policy purchased at the 70% coverage level ($\lambda = 0.7$) insures against a realization below 21 bushels per acre. If the farmer realized a yield of 16 bushels per acre, they would receive an indemnity payment for the insured value of 5 bushels per acre.

A variety of crop insurance plans and a number of new pilot programs are currently under development. Standard crop yield insurance, termed 'Multiple Peril Crop Insurance', pays an indemnity at a predetermined price to replace yield losses. 'Group-risk' yield insurance, termed 'Group Risk Plan' (GRP), is based upon the county's yield. Insured farmers collect an indemnity when their county's average yield falls below a yield guarantee, regardless of the farmers' actual yields. Three farm-level revenue insurance programs are available for a limited number of crops and regions: '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 (Goodwin & Ker (1998)). The recently developed 'Group Risk Income Plan', a variation of the Group Risk Plan, insures county revenues rather than yields (Baquet & Skees (1994)).

Section II.A.2 of the 1998 SRA states that an insurance company *"...must offer all approved plans of insurance for all approved crops in any State in which it writes an eligible crop insurance contract and must accept and approve all applications from all eligible producers."* An eligible farmer will not be denied access to an available, federally subsidized, crop insurance product. Therefore, an insurance company wishing to conduct business in a state cannot discriminate among farmers, crops, or insurance products in that state. An unusual situation arises, however; the responsibility for pricing the crop policies lies with the RMA but the insurance company must accept some liability for each policy they write and cannot choose which policy they will or will not write.[4]

It is clear that, in the absence of additional incentive mechanisms, insurance companies are unlikely to become involved in such a risk sharing arrangement. Therefore, to elicit their participation, two mechanisms are required that, necessarily, emulate a private market from the company's perspective. First, given that insurance companies do not set premium rates, there needs to be a mechanism by which they can cede the liability, or the majority thereof, of an undesirable policy (in a private market, the insurance company would simply refuse to write any policy deemed undesirable). Second, a mechanism providing an adequate return to the insurance company's capital and a level of protection against ruin (bankruptcy) is needed. Premium rates in a private market reflect a return to capital and a loading factor

[4] A new SRA agreement will take effect in 2005.

guarding against ruin. Premium rates set by the RMA do not reflect a return to capital but do include a loading factor. The SRA provides two such mechanisms which, in effect, emulate a private market from the perspective of the insurance company. In so doing, the SRA also provides a vehicle by which an insurance company can either retain or cede most of the premium and accompanying liability of policies of its choosing.

2.2. **Determining Premium Rates for Crop Contracts.** Accurate pricing of crop insurance policies requires accurate estimation of yield densities. We define the premium rate as the probability of a loss multiplied by the expected loss given that a loss has occurred. Formally, the actuarially fair premium rate for a yield insurance contract that guarantees a percentage, say λ, of the expected yield, say y^e, is given as

$$\begin{aligned} \text{Premium Rate} &= \mathcal{P}(Y < \lambda y^e)(\lambda y^e - E(Y|y < \lambda y^e)) \\ &= \int_0^{\lambda y^e} (\lambda y^e - y) f_Y(y|\mathcal{I}_t) dy \end{aligned} \tag{1}$$

where $0 \leq \lambda \leq 1$, the expectation operator and probability measure are taken with respect to the conditional yield density $f_Y(y|\mathcal{I}_t)$, and $\mathcal{I}_t$ is the information set known at the time of rating.[5] In our forthcoming analysis, the information set contains past yields and the county in which they were recorded. The RMA premium rate is taken with respect to their predicted yield and their estimate of the conditional yield density. Conversely, the insurance company determines their premium rate for the policy by integrating their estimate of the conditional yield density over the same space, $[0, \lambda y^e]$.

2.3. **Yield Data.** In choosing suitable yield data for our application we needed a sufficiently long time series to evaluate competing estimators in terms of their out-of-sample performance. Ideally, farm-level yield data would be used but this is only available for at most 20 years and generally much less. Therefore, we use county level yield data. This in turn required us to focus our attention on the GRP crop insurance program. GRP is

[5]The premium rate defined in (1) is in terms of expected loss with units bushels per acre.

an area-yield insurance program based on National Agricultural Statistics Service county yields. We use the 87 counties with a complete yield series from 1956 to 2001 in Illinois for all-practice corn. Historically, demand for GRP has been relatively high for this region-crop combination.

3. Estimating Yield Densities

The methods which have been used to model yield distributions fall into two camps, parametric and nonparametric. In the parametric camp, one common specification is the Beta distribution; see for example Hennessy, Babcock & Hayes (1997), Babcock & Hennessy (1996), Coble, Knight, Pope & Williams (1996), Borges & Thurman (1994), Kenkel, Busby & Skees (1991), Nelson (1990), and Nelson & Preckel (1989). These authors found sufficient evidence of skewness and/or kurtosis in their yield data and opted to use the Beta distribution in lieu of the Normal distribution. Interestingly, none of these authors tested the appropriateness of the Beta distribution. Just & Weninger (1999) attempt to renew support for the Normal distribution by calling into question the use of aggregate yield data, inflexible trend modeling, and the interpretation of the Normality test results. In contrast, Atwood, Shaik & Watts (2000) attempt, using more diverse crop-region combinations, to reduce support for the Normal distribution, while Ker & Coble (2003) found empirical evidence rejecting the use of both the Normal and Beta distributions for modeling county corn yields in Illinois. In the nonparametric camp, Goodwin & Ker (1998) and Ker & Coble (2003) use univariate nonparametric kernel methods to estimate yield densities and rate crop policies. Ker & Goodwin (2000) used empirical Bayes methods pointwise across the support to shrink the univariate nonparametric kernel estimates toward the mean.

3.1. **GRP Rating Methodology.** To model the temporal process of yields, the RMA employs a one knot linear spline with once-iterated least squares while windsorizing outliers (determined based on residual estimates from the first iterations) in the second iteration to estimate the temporal process of yields (see Skees, Black & Barnett (1997)).[6] After correcting for heteroskedasticity, they estimate a normal distribution and inflate the tails. The premium rate is the higher of the empirical rate or the rate derived from the Normal with inflated tails. The interested reader is directed toward Skees et al. (1997) and the references contained therein for more detailed information on the GRP rating methodology.[7]

3.2. **Nonparametric Methodology.** Most kernel-based approaches to modeling yields apply univariate kernel methods to each county separately - a 'cell' approach where each county represents a cell. It would be appealing to *jointly* model the yield density conditional on county membership thereby resulting in more efficient estimation of the yield density relative to that arising when separate univariate densities are estimated for yields in each county. County membership is, however, discrete, and traditional nonparametric estimators do not handle mixed discrete and continuous variables in a satisfactory manner (typically the aforementioned 'cell' approach is taken). Recently developed nonparametric kernel methods allow one to model joint distributions defined over mixed data types, and we elect to use this approach herein. The theory underlying this estimator can be found in Hall et al. (forthcoming), and we briefly describe this method below. In essence, the method involves the use of generalized product kernels where the kernels used to form the product kernel differ according to the underlying data type.

Let $Y' = (Y_1, \ldots, Y_q) \in \mathbb{R}^q$ be random variables whose outcomes will be conditioned on the random variables $X' = (X_1, \ldots, X_p) \in \mathbb{R}^p$. We let q_d and q_c denote the number of categorical and continuous variables in Y respectively with $q_d + q_c = q$, and do the same for X where $p_d + p_c = p$. We arrange the data with the q_d categorical data types appearing first followed by the q_c continuous ones so that $y_i' = (y_{i1}, \ldots, y_{iq}) = (y_{i1}, \ldots, y_{iq_d}, y_{iq_d+1}, \ldots, y_{iq})$, with corresponding smoothing parameters $h_y' = (h_1^y, \ldots, h_q^y) = (h_1^y, \ldots, h_{q_d}^y, h_{q_d+1}^y, \ldots, h_q^y)$.

[6]Windsorizing involves truncating the yield such that the absolute value of the residual is bounded below some determined level.
[7]We thank Jerry Skees for providing the actual RMA GRP rating code.

We let $y' = (y_1, \ldots, y_q) = (y_1, \ldots, y_{q_d}, y_{q_d+1}, \ldots, y_q)$ denote a vector-valued point at which an object is to be estimated. Finally, let $h'_x = (h^x_1, \ldots, h^x_p) = (h^x_1, \ldots, h^x_{p_d}, h^x_{p_d+1}, \ldots, h^x_p)$ be the smoothing parameters associated with X and let $h' = (h'_y, h'_x)$.

A multivariate product kernel for the random variables $(Y, X)' = (Y_1, \ldots, Y_q,\ X_1, \ldots, X_p)$ consisting of mixed categorical and continuous data types would be given by

$$
\begin{aligned}
K(y_i, x_i, y, x, h_y, h_x) = \ & \textstyle\prod_{j=1}^{q_d} \mathcal{K}(y_{ij}, y_j, h^y_j) \prod_{j=q_d+1}^{q_d+q_c} \mathcal{K}'(y_{ij}, y_j, h^y_j) \times \\
& \textstyle\prod_{j=1}^{p_d} \mathcal{K}(x_{ij}, x_j, h^x_j) \prod_{j=p_d+1}^{p_d+p_c} \mathcal{K}'(x_{ij}, x_j, h^x_j)
\end{aligned} \tag{2}
$$

where the kernel functions appearing in the first and third products are categorical and those in the second and forth products are continuous, while the product kernel for the random variables $X' = (X_1, \ldots, X_p)$ consisting of mixed categorical and continuous data types would be given by

$$
K(x_i, x, h_x) = \prod_{j=1}^{p_d} \mathcal{K}(x_{ij}, x_j, h^x_j) \prod_{j=p_d+1}^{p_d+p_c} \mathcal{K}'(x_{ij}, x_j, h^x_j), \tag{3}
$$

where the kernel functions appearing in the first product are categorical and those in the second are continuous.

For unordered categorical variables (using X by way of example) we use the kernel function of Aitchison & Aitken (1976) given by

$$
\mathcal{K}(x_{ij}, x_j, h^x_j) = \begin{cases} 1 - h^x_j & \text{if } |x_{ij} - x_j| = 0, \\ \frac{h^x_j}{c-1} & \text{if } |x_{ij} - x_j| \geq 1, \end{cases} \tag{4}
$$

where c is the number of 'categories' that the categorical variable can assume.

For continuous variables we use the Gaussian kernel function[8] given by

$$
\mathcal{K}'(x_{ij}, x_j, h^x_j) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x_{ij} - x_j}{h^x_j}\right)^2\right] \tag{5}
$$

[8]Note that we subsume the multiplicative (inverse) bandwidth $1/h^x_j$ in the definition of the kernel function itself.

Letting $K(\cdot)$ be the respective product kernel functions defined in equations (2) and (3), the kernel estimator of the *conditional* probability density function (PDF) of Y given X denoted $f(y|x)$ is given by

$$\hat{f}(y|x) = \frac{\sum_{i=1}^{n} K(y_i, x_i, y, x, h_y, h_x)}{\sum_{i=1}^{n} K(x_i, x, h_x)}, \tag{6}$$

with the same vector of smoothing parameters h_x used in both the numerator and denominator. Properties of this estimator including rates of convergence and asymptotic normality can be found in Hall et al. (forthcoming), while for a general recent treatment of a host of issues concerning nonparametric kernel estimators we direct the interested reader to Pagan & Ullah (1999).

In the current context, y_i represents county-specific yields while x_i is the county in which the yield was recorded thereby *jointly* modeling yields and counties. In contrast, the standard cell-based kernel approach conditions on a particular county and then models the yields for that county using a univariate kernel estimator, and then repeats this for all counties.

It is well-known that the judicious selection of the smoothing parameters is the most important factor underlying the estimator's performance. We elect to use a fully automatic method of smoothing parameter selection, the least-square cross-validation approach proposed by Hall et al. (forthcoming). We employ a multivariate conjugate gradient search algorithm for minimization of the cross-validation function, and this permits differing optimal bandwidths for each variable. For the estimation of unconditional distributions with mixed datatypes see Li & Racine (2003), while for local constant and local polynomial regression with mixed datatypes see Racine & Li (2004) and Li & Racine (2004) and the references therein.

4. Analysis

As discussed above, the appropriate context in which to evaluate any proposed methodology for rating crop insurance policies is to assume the role of an insurance company. We can determine whether or not significant excess rents can be garnered when using a particular methodology by estimating the premium rate schedule and then determining which policies to retain and which to cede. In this section we undertake such a simulation designed to gauge the relative performance of the Hall et al. (forthcoming) kernel estimator, the standard cell-based univariate kernel estimator, and the RMA inflated Normal parametric estimator. Our simulation has the following salient features[9]:

(1) The RMA estimates their premium rate, denoted $\hat{\pi}_{RMA}$, using the GRP rating methodology. That is, the temporal models are estimated using a robust one-knot linear spline, with premium rates being based on the maximum of the associated empirical rate or a rate derived from a Normal with inflated tails.

(2) The private insurance company estimates their premium rate, denoted $\hat{\pi}_{IC}$, by using the GRP methodology to estimate the temporal models and the Hall et al. (forthcoming) estimator to estimate the conditional yield density.

(3) The private insurance company, a profit maximizer, cedes a contract if $\hat{\pi}_{RMA} < \hat{\pi}_{IC}$ because they believe the contract to be under-priced and expect a loss. Conversely, they will retain a contract if $\hat{\pi}_{RMA} > \hat{\pi}_{IC}$.

(4) One-step ahead premium rates are estimated for each of the eighteen years, indexed from 1984 to 2001, based on yield data up-to and including the preceding year. That is, when constructing the 1989 estimated premium rates, only yield data from 1956-1988 is used.

(5) The actual out-of-sample yield realizations are used to calculate the loss ratios for the set of contracts that the insurance company retains, the set of contracts the insurance company cedes (the set of contracts thereby held by the RMA), and the 'program' or entire set.

[9]For the kernel estimators, bandwidths must be recomputed for each successive one-step forecast as the estimation sample size increases. Therefore, bandwidths are not reported here but are available upon request.

If either nonparametric estimator better describes the yield distribution than the estimator used by the RMA, then the loss ratio for the contracts retained by the private insurance company would be expected to be lower than the overall loss ratio and thereby the loss-ratio for the government. Approximate randomization tests, which simulate the distribution of a desired statistic under the null, are used to ascertain statistical significance (see Kennedy (1995)). When evaluating the performance of each nonparametric method, our null is that the insurance company recover rents by strategically reinsuring with the government, i.e. that the insurance company's loss ratio is equal to the overall loss-ratio. Under the null, the insurance company estimates every policy to have zero expected gain and thus they are indifferent to retaining or ceding each and every policy. Having gauged each estimator's performance relative to the RMA, we are in a position to assess their relative performance.

To obtain a realization from the null distribution, the insurance company randomly retains a policy with probability ρ where ρ equals the fraction of policies retained in the original simulation (see table 1). We randomize over which policies are retained, not over the number of policies retained. We compare the insurance company loss ratio from the analysis, denoted τ^*, to 1000 simulated loss ratios under the null $\{\tau_1, \tau_2, \ldots, \tau_{1000}\}$. The p-value for the test equals the fraction of $\{\tau_1, \tau_2, \ldots, \tau_{1000}\}$ for which $\tau_i \leq \tau^*$.

Denote Ξ as the universe consisting of 1,566 policies (87 counties $\times$ 18 years), $\mathcal{F}$ the set of policies the insurance company retains, and $\mathcal{F}^c$ the set of policies the insurance company cedes. The loss ratio for a set, say $\mathcal{F}$, is:

$$\text{Loss Ratio}_{\mathcal{F}} = \frac{\sum_{j \in \mathcal{F}} \max(0, \lambda y_j^e - y_j)}{\sum_{j \in \mathcal{F}} \hat{\pi}_{RMA,j}} \tag{7}$$

where j is the policy, y_j is the realized yield associated with policy j, λ is the coverage level, y_j^e is the RMA expected yield associated with policy j, and $\hat{\pi}_{RMA,j}$ is the RMA premium rate for policy j. We calculate the loss ratio for the program, the insurance company, and the RMA by summing over Ξ, $\mathcal{F}$, and $\mathcal{F}^c$ respectively.

Table 1 summarizes the program, RMA, and insurance company loss ratios for all simulations at both the 75% and 85% coverage levels based on a comparison of the nonparametric and RMA premium rates.

TABLE 1. GRP Simulation Results for all Counties in Illinois for all-practice Corn. Nonparametric Versus RMA Rates.

	Univariate Kernel Estimator vrs RMA Rating Methodology	Hall, Racine & Li Kernel Estimator vrs RMA Rating Methodology
75% Coverage Level		
Program Loss Ratio	0.935	0.935
Insurance Company Loss Ratio	1.297	0.719
RMA Loss Ratio	0.925	1.182
Percent of Policies Retained	4.5%	26.4%
p-value	0.799	0.036
85% Coverage Level		
Program Loss Ratio	1.097	1.097
Insurance Company Loss Ratio	0.000	0.751
RMA Loss Ratio	1.102	1.213
Percent of Policies Retained	0.4%	18.0%
p-value	0.287	0.017

At the 75% coverage level, the insurance company using the Hall et al. (forthcoming) estimator would retain approximately 26% of the policies while ceding 74% of the policies to the RMA suggesting that the current RMA rating methodology may underestimate premium rates as noted in Section 1. More importantly, the insurance company's loss ratio based on the 26% of contracts it retains is reduced to 0.72 while the RMA loss ratio increases to 1.18. For a multi-billion dollar program, the marked reduction in the insurance company's loss ratio is significant. The p-value of 0.036 strongly indicates that the company is doing better than if it were to randomly select policies. By way of comparison, using the standard cell-based kernel estimator applied to each county individually the insurance company loss ratio actually increased to 1.297. The percent of contracts retained is, not surprisingly, very small.[10] Note that the univariate kernel estimator performs quite poorly in that if one just randomly chooses 4.5% of the contracts, the resulting loss ratio will tend to be less than 1.297 with roughly 80% probability.

At the 85% coverage level, the insurance company using the Hall et al. (forthcoming) estimator retain approximately 18% of the policies while ceding 82% of the policies to the RMA suggesting that the current RMA rating methodology appears to underestimate the rates more significantly than at the 75% coverage level. This finding is consistent with the overall loss ratio being greater than 1. Interestingly, the insurance companies' loss ratio based on the 18% of contracts it retains using the Hall et al. (forthcoming) estimator is reduced to 0.75 while the RMA loss ratio rises to 1.21. The p-value of 0.017 suggests that the company is doing significantly better than if it were to randomly select policies. By way of comparison, using the standard cell-based univariate kernel estimator applied to each county individually, the insurance company only retains seven out of the 1,566 policies. Although the loss ratio for those seven policies is zero, if we randomly chose seven policies to retain, the insurance company would realize zero loss ratio with roughly 29% probability. Therefore, while the loss ratio has decreased, it does not differ statistically from the overall program loss-ratio.

It is apparent that, by using a more efficient nonparametric estimator of yield densities than the traditional cell-based estimator, a private insurer could successfully adverse select against the government via their choice of which policies to retain.

5. Conclusions and Policy Implications

Given the increasing interest in crop insurance and agricultural risk arising in part due to the *Agricultural Risk Protection Act*, there has been a recent surge in interest in the identification of improved methods for characterizing yield distributions. There is mounting evidence indicating that common parametric yield distribution models may be inappropriate

[10]The premium rate based on the univariate kernel will necessarily be higher than the empirical rate and as such, the univariate kernel rate will tend to be higher than the RMA premium rate resulting in very few policies retained.

for characterizing the underlying data generating process, hence some have turned instead to nonparametric estimation methods. In this article we continue this trend and investigate the application of a new nonparametric conditional distribution estimator proposed by Hall et al. (forthcoming). This estimator has the potential to improve the accuracy of yield density estimates and their attendant insurance rates through the joint modeling of *continuous* data (yield) and *discrete* data (county in which the yield was recorded) using generalized product kernels.

We investigate the behavior of the Hall et al. (forthcoming) estimator by focusing on economic implications of estimation error. Competing parametric and nonparametric estimators are used to estimate a set of yield densities and to derive the associated premium rates. We evaluate the competing estimators by calculating out-of-sample loss ratios based on decision rules for retaining or ceding GRP crop insurance contracts. This simulation is of interest from an economic and policy perspective because the SRA enables the private insurance companies to retain or cede, ex-ante and subject to constraints, varying portions of the realized underwriting gains or losses of every federally subsidized crop insurance contract it sells. Our results indicate that, for this set of counties, the current RMA rating methodology may underestimate the premium rates as evidenced by the few policies retained in our simulation. However, of the policies retained, the loss ratio suggests that the Hall et al. (forthcoming) estimator was successful at significantly increasing rents to insurance companies. In addition, the new estimator is found to significantly outperform standard univariate nonparametric kernel estimators.

References

Aitchison, J. & Aitken, C. G. G. (1976), 'Multivariate binary discrimination by the kernel method', *Biometrika* **63**(3), 413–420.

Atwood, J., Shaik, S. & Watts, M. (2000), 'Normality of crop yields', *Working Paper, Department of Agricultural Economics and Economics, Montana State University* .

Babcock, B. & Hennessy, D. (1996), 'Input demand under yield and revenue insurance', *American Journal of Agricultural Economics* **78**, 416–427.

Baquet, A. & Skees, J. (1994), 'Group risk plan insurance: An alternative management tool for farmers', *Choices* .

Borges, R. & Thurman, W. (1994), 'Marketing quotas and subsidies in peanuts', *American Journal of Agricultural Economics* **76**, 801–809.

Coble, K., Knight, T., Pope, R. & Williams, J. (1996), 'Modeling farm-level crop insurance demand with panel data', *American Journal of Agricultural Economics* **78**, 439–447.

Goodwin, B. & Ker, A. (1998), 'Nonparametric estimation of crop yield distributions: Implications for rating group-risk (grp) crop insurance contracts', *American Journal of Agricultural Economics* **80**, 139–153.

Hall, P., Racine, J. & Li, Q. (forthcoming), 'Cross-validation and the estimation of conditional probability densities', *Journal of the American Statistical Association* .

Hennessy, D., Babcock, B. & Hayes, D. (1997), 'Budgetary and producer welfare effects of revenue insurance', *American Journal of Agricultural Economics* **79**, 1024–34.

Just, R. & Weninger, Q. (1999), 'Are crop yields normally distributed?', *American Journal of Agricultural Economics* **81**, 287–304.

Kenkel, P., Busby, J. & Skees, J. (1991), 'A comparison of candidate probability distributions for historical yield distributions', *Presented Paper at SAEA meetings, Fort Worth TX* .

Kennedy, P. E. (1995), 'Randomization tests in econometrics', *Journal of Business and Economic Statistics* **13**, 85–94.

Ker, A. & Coble, K. (2003), 'Modeling conditional yield distributions', *American Journal of Agricultural Economics* **85**, 291:304.

Ker, A. & Goodwin, B. (2000), 'Nonparametric estimation of crop insurance rates revisited', *American Journal of Agricultural Economics* **83**, 463:478.

Ker, A. & McGowan, P. (2000), 'Weather based adverse selection: The private insurance company perspective', *Journal of Agricultural and Resource Economics* **25**, 386–410.

Li, Q. & Racine, J. (2003), 'Nonparametric estimation of distributions with categorical and continuous data', *Journal of Multivariate Analysis* **86**, 266–292.

Li, Q. & Racine, J. (2004), 'Cross-validated local linear nonparametric regression', *Statistica Sinica* **14**(2), 485–512.

Nelson, C. (1990), 'The influence of distribution assumptions on the calculation of crop insurance premia', *North Central Journal of Agricultural Economics* **12**, 71–78.

Nelson, C. & Preckel, P. (1989), 'The conditional beta distribution as a stochastic production function', *American Journal of Agricultural Economics* **71**, 370–378.

Pagan, A. & Ullah, A. (1999), *Nonparametric Econometrics*, Cambridge University Press.

Racine, J. & Li, Q. (2004), 'Nonparametric estimation of regression functions with both categorical and continuous data', *Journal of Econometrics* **119**(1), 99–130.

Skees, J., Black, J. & Barnett, B. (1997), 'Designing and rating an area yield crop insurance contract', *American Journal of Agricultural Economics* pp. 430–438.

Jeff Racine, Department of Economics and Center for Policy Research, Syracuse University, Syracuse, NY USA 13244, 315-443-1630, jracine@maxwell.syr.edu.

Alan P. Ker, Department of Agricultural and Resource Economics, University of Arizona, Economics Building, Tucson, AZ 85704, 520-621-6265, alan_ker@arizona.edu.