NTIFYING LONG RUN AGRICULTURAL RISKS AND EVALUATING FARMER RESPONSES TO RISK

Proceedings of a Seminar sponsored by
Southern Regional Project S-232
"Quantifying Long Run Agricultural Risks and Evaluating Farmer Reponses to Risk"
Orlando, Florida
March 22-25, 1992

Department of Agricultural and Resource Economics
The University of Arizona
Tucson, AZ 85721

September 1992

The Effects of Land Unit Size on Crop-Yield Variability

Richard Byrne
Paul Schultz
Thomas P. Zacharias
National Crop Insurance Services
Overland Park, Kansas

The relationship between crop-yield variability and acreage has been recognized since the early empirical work of Carter and Dean. Despite the recognition of this relationship, there appears to have been very little, if any, empirical work done on the problem. Lack of research could be due to data problems or lack of interest in the phenomena. The purpose of this paper is to present empirical results which attempt to capture the magnitude of the relationship and to discuss the practical implication of the relationship as it relates to the crop insurance industry. Specifically, this paper defines and presents estimates of a "variance inflation factor" for major crops and growing regions. The variance inflation factor is then used in a specific crop insurance application. Some additional empirical results are also presented.

Organization of the paper is as follows. The first section presents a brief discussion of the data used in the analysis and the construction of the data set. Section II presents the estimation of the variance inflation factor and some empirical results. Section III presents results from two crop insurance applications which provide further support for the empirical results shown in the previous section. Section IV simply concludes the paper with some final thoughts on the results and directions for future research.

Section I
Constructing County Yield Distributions from Crop Insurance Data

County yields are typically available from USDA or state agricultural departments, and the method of data collection is a questionnaire. Since individual records collected by insurance agents tend to be supported by hard-copy records, it is appropriate to use these data. Moreover, crop insurance yield data is one of the only sources of individual crop yields other than farm management record-keeping systems (public or private).
Yield history is recorded by an insurance agent for each farm unit. It consists of the last ten years of yields (bu/ac) and the corresponding acreages. Total bushels is the product of yield and acreage in any particular year. County bushels is the sum of the total bushels for all insureds in one county for one year. County acreage is the total acreage for all insureds in that same year. Finally, county yield is the ratio of county bushels to county acreage. Thus, for each of approximately ten years a county yield and a county acreage are available. Comparisons of NCIS yield data and USDA NASS yields are presented in Figures 1 -3. These figures suggest that individual insurance data aggregated to the county level corresponds well to the USDA county yield estimates. These figures also suggest that insurance participants are not noticeably below-average farmers.

CERRO GORDO COUNTY, IA
YIELD DATA

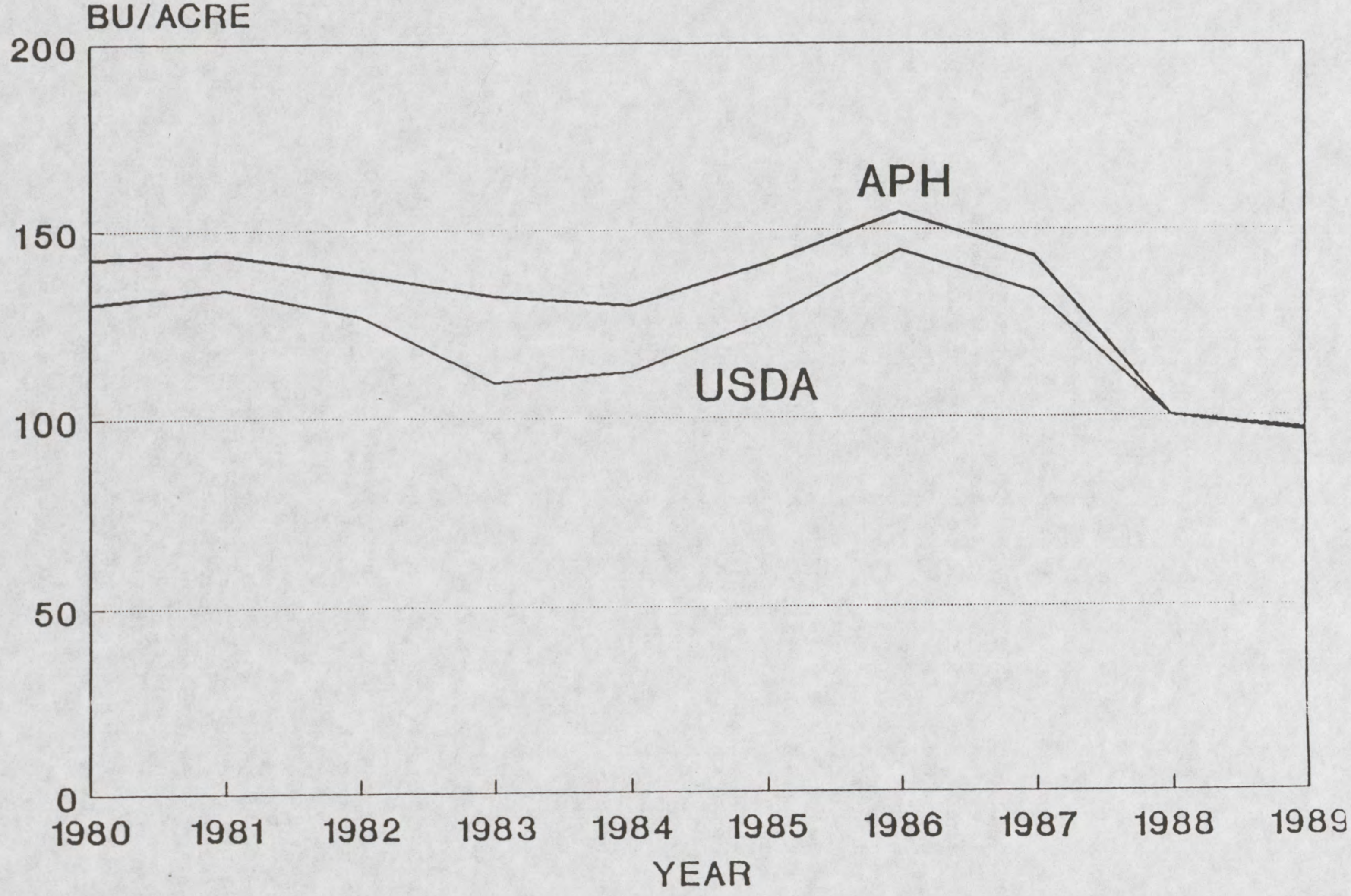
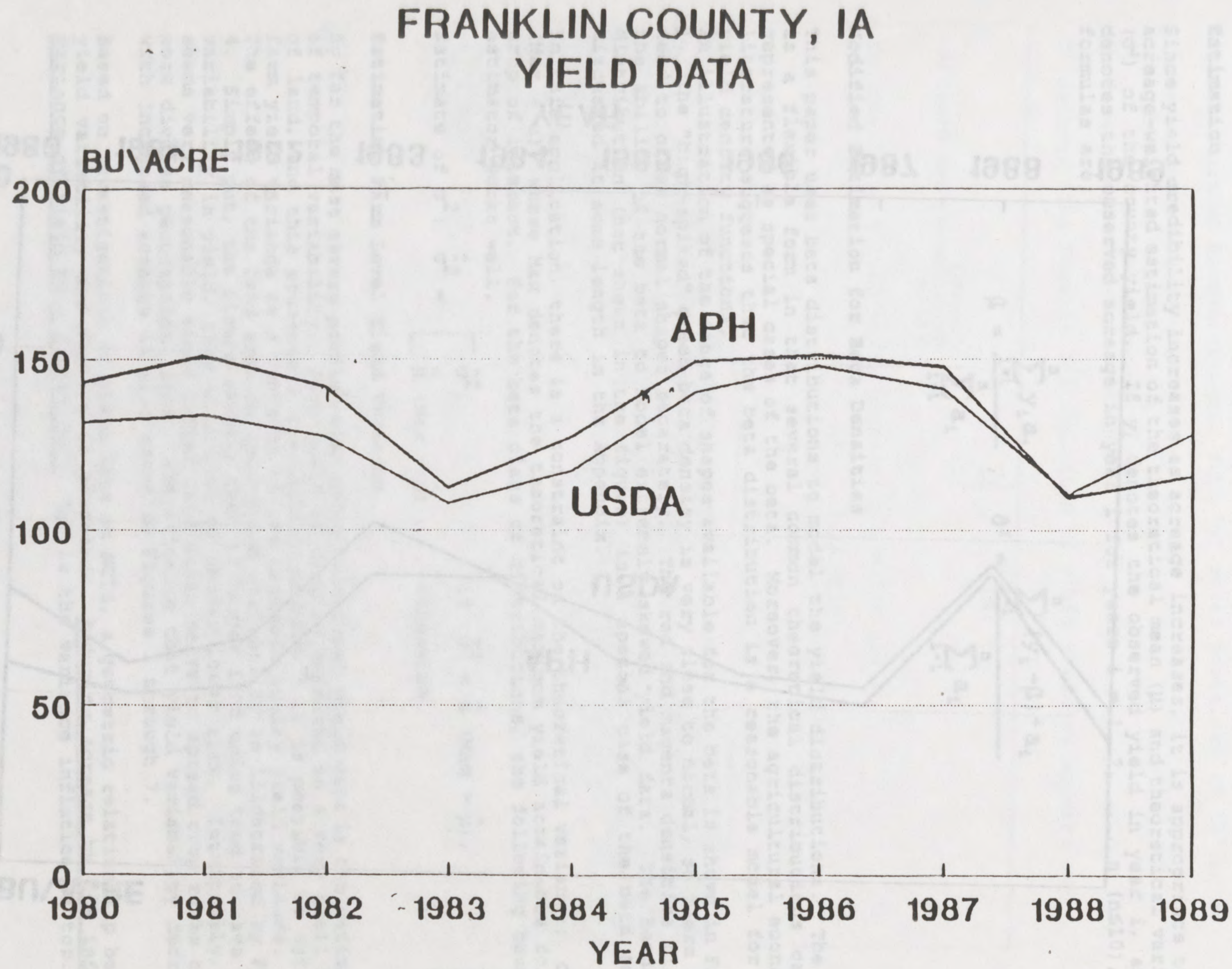


FIGURE 1



FRANKLIN COUNTY, IA
YIELD DATA
BU\ACRE
200
150
100
50
0
APH
USDA
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
YEAR

Figure 2

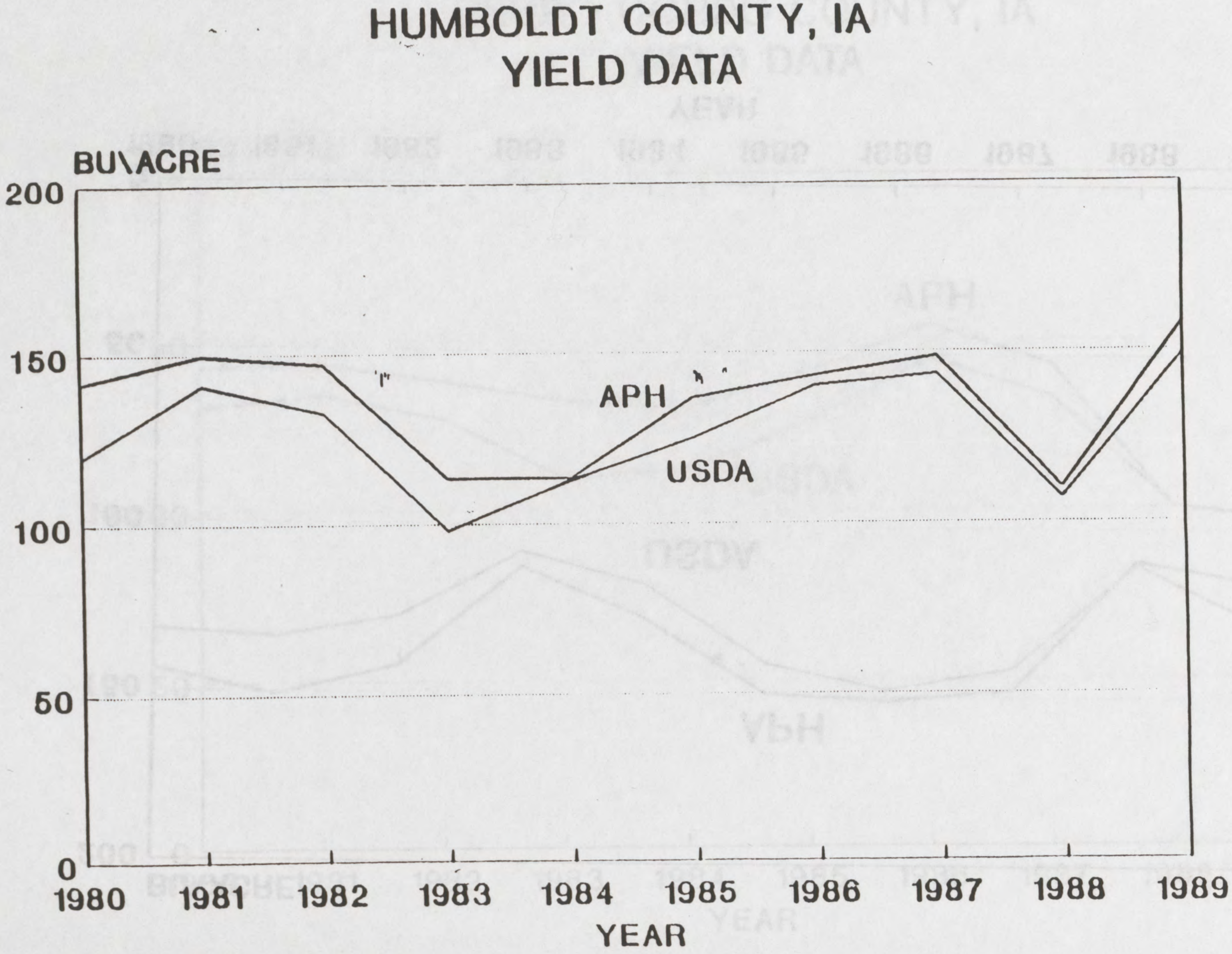

HUMBOLDT COUNTY, IA
YIELD DATA
BU\ACRE
200
150
100
50
0
APH
USDA
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
YEAR

Figure 3

Estimation

Since yield credibility increases as acreage increases, it is appropriate to use acreage-weighted estimation of the theoretical mean (μ) and theoretical variance (σ^2) of the county yield. If y_i denotes the observed yield in year i, and a_i denotes the observed acreage in year i for years i = 1, 2, ..., n (n≤10), the formulas are:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} y_i a_i}{\sum_{i=1}^{n} a_i}, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{\mu})^2 a_i}{\sum_{i=1}^{n} a_i}$$

Modified Estimation for Beta Densities

This paper uses beta distributions to model the yield distributions. The beta is a flexible form in that several common theoretical distributions can be represented as special cases of the beta. Moreover, the agricultural economics literature suggests that the beta distribution is a reasonable model for crop yield density functions.
An illustration of the range of shapes available for the beta is shown in Figure 13. The "high-spiked" green beta density is very close to normal, so there is no need to offer normal shapes separately. The red and magenta densities display the ability of the beta to model extremely skewed yield data. The Bernoulli distribution (not shown in the figure) is a special case of the beta and is discussed at some length in the Appendix.

In this application, there is a constraint on the theoretical variance: $\sigma^2 < \mu$ (Max - u), where Max denotes the theoretical maximum yield attainable for the crop of interest. For the beta class of distributions, the following boundary estimator works well.

$$\text{Estimate of } \sigma^2: \quad \hat{\hat{\sigma}}^2 = \begin{cases} \hat{\sigma}^2, & \text{if } \hat{\sigma}^2 < \hat{\mu}\,(\text{Max} - \hat{\mu}), \\ \hat{\mu}\,(\text{Max} - \hat{\mu}), & \text{otherwise.} \end{cases}$$

Estimating Farm Level Yield Variance

By far the most severe problem with using farm level yield data is the estimation of temporal variability. Farm yield is usually measured on a very small block of land, and this presents a credibility problem. It is possible to estimate farm yield variance as a function of the credible county yield variance.
The effect of the land area on crop-yield variability is illustrated by Figure 4. Simply put, the figure reveals that if larger land units tend to have lower variability in yield, they will tend to confer lower risk. Intuitively, this seems very reasonable since larger land units serve to spread crop risks over a more diverse geographical area. The evidence that yield variability decreases with increased acreage is also shown in Figures 5 through 7.

Based on investigations of yield data at NCIS, a geometric relationship between yield variability and acreage is proposed: Reducing acreage by 50% inflates variance of yield by a factor "g". "g" is the variance inflation factor.

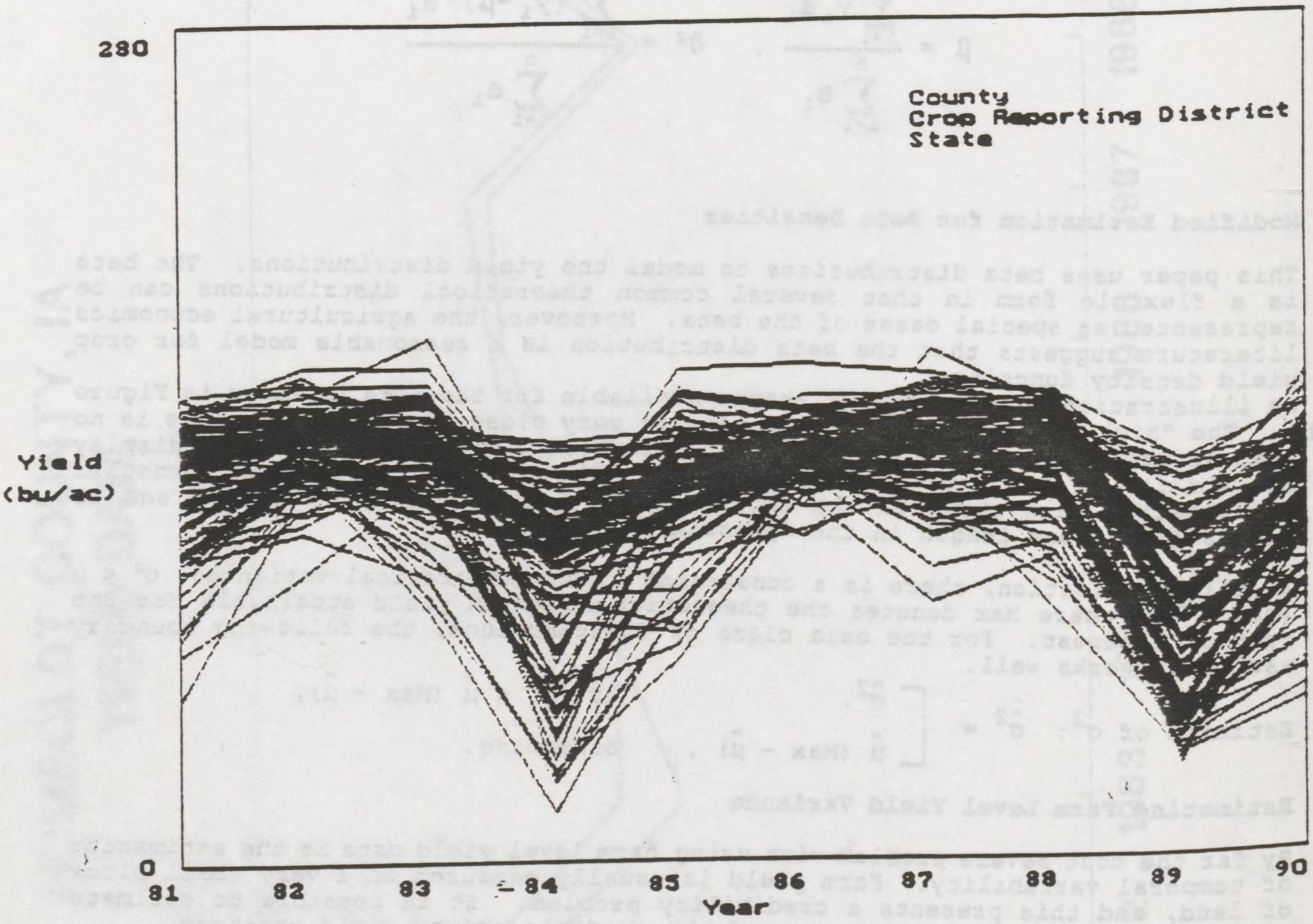


Figure 4. Yield Histories for Iowa Grain Corn

Consider nested blocks of land, with "a" acres contained in "b" acres:

acreage = b	acreage = a

In symbols, the formula for inflation of yield variance appears as:

$$\frac{\sigma_a^2}{\sigma_b^2} = g^{\log_2\left(\frac{b}{a}\right)}$$

Example 1. b = 1000 acres containing a = 500 acres:

$$\frac{\sigma_a^2}{\sigma_b^2} = g^{\log_2\left(\frac{1000}{500}\right)} = g$$

Example 2. b = 1000 acres containing a = 250 acres:

$$\frac{\sigma_a^2}{\sigma_b^2} = g^{\log_2\left(\frac{1000}{250}\right)} = g^2$$

Estimation of Variance Inflation Factor Using Yield Data

The first step is to isolate a function of g as a function of all other variables. Taking the natural log of both sides of the variance inflation formula gives

$$\ln\left(\frac{\sigma_a^2}{\sigma_b^2}\right) = h \ln\left(\frac{b}{a}\right)$$

where $h = \log_2(g)$. Letting X denote the left hand side and $Y = \ln(b/a)$, a linear model may be written: $Y_i = h X_i + e_i$, where farm acreage is a_i and the errors e_i are distributed independently with zero mean and variance δ^2/a_i, for $i=1,\ldots,n$. δ^2 is unknown. For this model there is a weighted regression estimator for h, and also an estimate of g:

$$\hat{h} = \frac{\sum_{i=1}^{n} X_i Y_i a_i}{\sum_{i=1}^{n} X_i^2 a_i}, \quad \hat{g} = 2^{\hat{h}}$$

The fraction of weighted variation in Y explained by the regression on X is

$$R^2 = \frac{\sum_{i=1}^{n} Y_i^2 a_i - \sum_{i=1}^{n} (Y_i - \hat{n} X_i)^2 a_i}{\sum_{i=1}^{n} Y_i^2 a_i}$$

The sample size n is the number of farms in all of the counties of interest, typically spanning an entire state. For each farm, the ordered pair (X,Y) is computed subject to the constraint that the farm sample variance exceed the corresponding county sample variance. This tends to discount the effects of very small farms with suspiciously constant yields. In the interest of credibility, the estimates of both farm and county variance are required to have four or more years of data.

Three entries in the following table are accompanied by graphics:
Figure 5: Iowa Corn
Figure 6: N. Dakota Wheat
Figure 7: Minnesota Soybeans

Results: Variance Inflation Factor for Several Crops and States

Crop Kind	State	$\hat{g}$	R^2	acreage
Wheat	Idaho	1.077	48%	272,860
	Minnesota	1.043	37	693,683
	Montana	1.036	41	1,173,641
	N. Dakota	1.028	32	5,157,680
	S. Dakota	1.041	28	3,653,726
Corn	Illinois	1.036	38%	1,269,882
	* Iowa	1.055	40	2,486,035
	Minnesota	1.046	34	1,169,092
	Missouri	1.042	34	231,274
	Nebraska	1.104	49	1,641,549
Soybeans	Illinois	1.063	44%	552,984
	Iowa	1.101	54	1,667,879
	Minnesota	1.085	46	1,470,519
	Missouri	1.096	45	321,525
	Nebraska	1.069	39	340,546
	** Louisiana	1.039	21	122,834

Iowa corn is one of the most predictable insurance portfolios. The table says that inflating variance 5.5% per 50% reduction in acreage would explain 40% of the adverse error in using county sample variance as the estimate of farm variance. This means the proposed NCIS system could improve rating considerably. Louisiana soybeans are of interest due to persistently high loss ratios. The table says that inflating the county sample variance 3.9% per 50% reduction in acreage would explain 21% of the adverse error in using the county sample variance as the estimate of farm variance. This suggests that variance inflation is only a small part of the problem there.

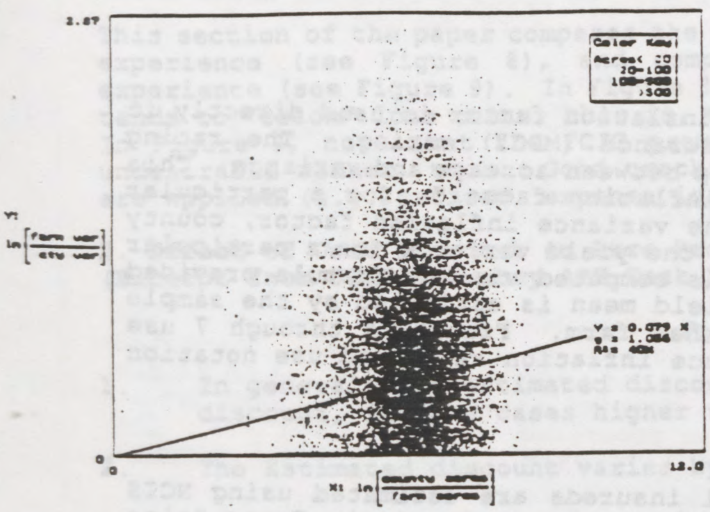


Figure 5. Farm Variance Inflation for Iowa Corn
Weighted Regression (Weight is Farm Acreage)

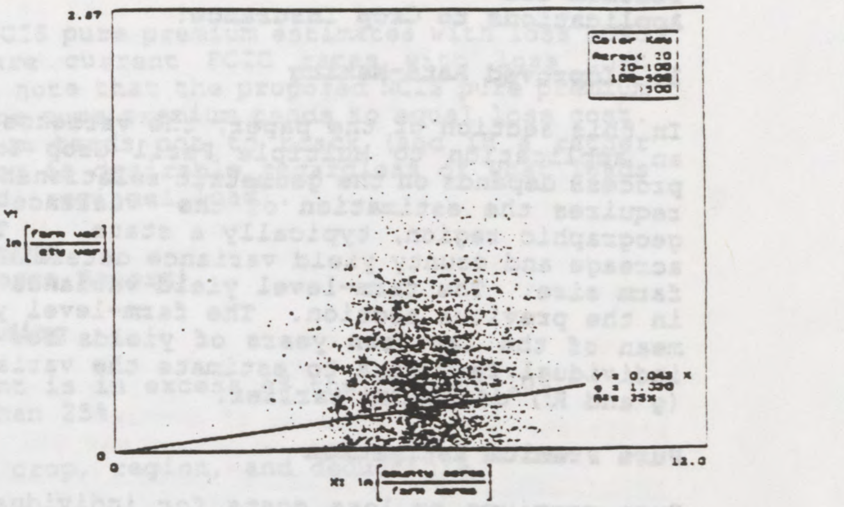


Figure 6. Farm Variance Inflation for N. Dakota Wheat
Weighted Regression (Weight is Farm Acreage)

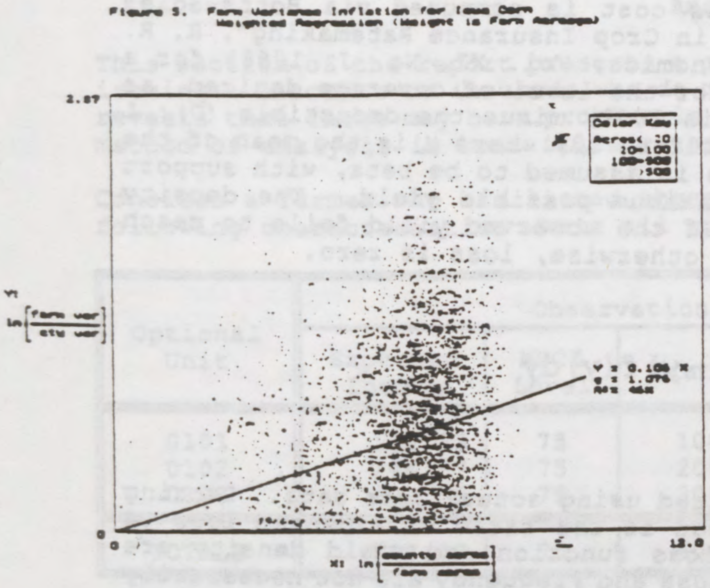


Figure 7. Farm Variance Inflation for Minnesota Soybeans
Weighted Regression (Weight is Farm Acreage)

Section III
Applications to Crop Insurance:

1. Improved Rate-Making

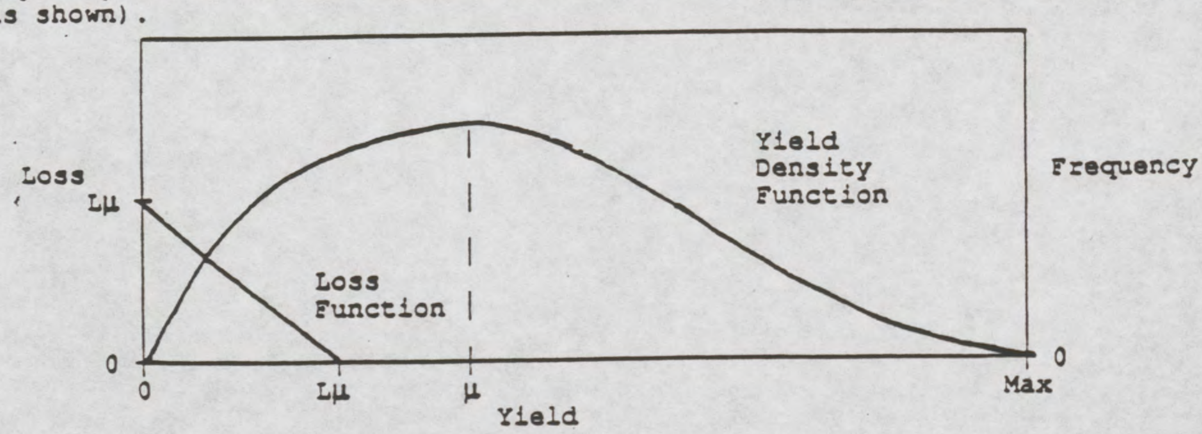
In this section of the paper, the variance inflation factor is used directly in an application to Multiple Peril Crop Insurance (MPCI) rating. The rating process depends on the geometric relationship between acreage and variance. This requires the estimation of the "variance inflation factor" for a particular geographic region, typically a state. The variance inflation factor, county acreage and county yield variance determine the yield variance for a particular farm size. The farm-level yield variance is computed via the formula provided in the previous section. The farm-level yield mean is estimated by the sample mean of the last ten years of yields for that farm. Figures 5 through 7 use individual farm data to estimate the variance inflation rate, and use notation (g and R^2) developed earlier.

Pure Premium Estimation

Pure premiums or loss costs for individual insureds are estimated using NCIS farm-level crop yield data. Expected loss cost is computed via Botts-Boles technique (see "Use of Normal Curve Theory in Crop Insurance Ratemaking", R. R. Botts and J. N. Boles, Journal of Farm Economics, Vol. XI, No. J, 1958) for a beta density with this mean and variance, for the level of coverage desired, as in Appendix D. The level of coverage (L) is 100%, minus the deductible (D): L = 1 - D. The liability (or guarantee) is written $L\mu$, where μ is the mean of the yield distribution. The yield distribution is assumed to be beta, with support from zero to Max, where Max denotes the maximum possible yield. The density function is $f(y)$, where y denotes yield. If the observed yield fails to reach the guarantee, the difference is the loss; otherwise, loss is zero.
Expected loss is defined by the integral

$$E(\text{Loss}) = \int_{y=0}^{L\mu} (L\mu - y) f(y)\,dy$$

which may be computed numerically or estimated using actual loss data. Letting L/C denote loss cost, the expected loss cost is the ratio of expected loss to liability: $E(L/C) = E(\text{Loss}) / L\mu$. The loss function and yield density are superimposed below (the vertical scales of Loss and Frequency are not necessarily as shown).



Comparison of Proposed NCIS Pure Premium with FCIC Rates and Loss Cost

This section of the paper compares the NCIS pure premium estimates with loss cost experience (see Figure 8), and compare current FCIC rates with loss cost experience (see Figure 9). In Figure 8, note that the proposed NCIS pure premium tends to "track" loss costs, that is, the pure premium tends to equal loss cost. In Figure 9, note that the FCIC premium tends not to track (and in a rather undesirable manner, too). Good tracking is desirable regardless of what loads are applied (e.g. catload, expense load, regional load).

2. Effect of Unit Discount on Pure Premium (Excerpt from FCIC/Industry APH Task Force Report)

Finding

1. In general, the estimated discount is in excess of the current 10% discount, in some cases higher than 25%.

2. The Estimated discount varies by crop, region, and deductible.

Analysis

This section of the report presents a preliminary study on the reasonableness of issuing a discount to insureds who do not use optional units. The investigation reveals that there may be a problem in the current pricing of the discount. The method of analysis is best illustrated by an example.

Consider a farmer who has listed three optional units with the following characteristics where all numbers are in dollars.

Optional Unit	Observations				Estimated Loss		
	Expected Income	MPCI Level	Liability	Loss	Pess.	Mid.	Opt.
0101	133.33	75	100.00	5.00	5.00	5.00	5.00
0102	266.67	75	200.00	0.00	0.00	-33.33	-66.67
0103	400.00	75	300.00	80.00	80.00	80.00	80.00
TOTALS	800.00	75	600.00	NA	85.00	51.67	18.33

Optional unit 0102 is tricky because it did not sustain a loss. This <u>censors</u> the income information. Three methods, ranging from optimistic to pessimistic, are used to reconstruct the loss information on each optional unit, so it can be summed into a loss on the entire unit. The middle method is a compromise, halfway between optimistic and pessimistic, and is probably the best answer.

The optimistic or middle methods may show negative losses. Positive and negative losses may cancel out, when summing to get the unit total loss. For each crop/state/level combination with substantial benefit due to this loss cancellation phenomenon, the unit discount should be just as substantial.

The pessimistic loss estimate presumes the yield was just high enough to prevent a loss on the optional unit. The optimistic loss estimate presumes the optional unit's yield attained the average APH yield. This is optimistic if a loss is reported on another optional unit, and it is irrelevant to unit computations if no loss is reported on any optional unit, so it is just called optimistic. For the example, the middle estimate of loss is $51.67 and the liability is $600.00, so the loss cost for the entire unit is 8.6%.

Considering the three optional units separately, total loss is $85.00 and liability is $600.00, so the loss cost is 14.2%. By pooling optional units into one unit, the loss cost decreases by 39%. This is an estimate of the unit discount, using the one unit in the example.

In this study, all optional units were selected for the states on the accompanying maps, and premium reduction factors were produced for each individual level of coverage as well as for all levels aggregated. An examination of the maps indicates that the discount (for all levels aggregated) is in the neighborhood of ten percent only for soybeans in Arkansas, Louisiana and Mississippi. Further examination reveals that in each case the discount decreases as the level of coverage increases.

If the Estimated discount is greater than the FCIC-regulated discount, then losses on optional units are larger than losses on single units. Expected indemnities on multiple units will be greater than expected indemnities on single units for the same amount of premium. Thus, farmers have a greater incentive to use optional units and excess losses will be generated in this situation. The net effect is an increased loss ratio due to excess losses if the Estimated discount is greater than the FCIC-regulated discount.

If the Estimated discount is less than the FCIC-regulated discount, then farmers are given too much of a credit for combining units. If the Estimated discount is 5% and the FCIC-regulated discount is 10%, farmers are given a net excess credit of 5% and premiums collected are less than need be had the appropriate discount been applied. The net effect is an increased loss ratio due to insufficient premium collection if the Estimated discount is less than the FCIC-regulated discount.

Section IV

Evidence has been presented in this paper which indicates that larger farm unit size results in sizeable reductions in crop-yield variability. The qualitative nature of these results should come as no great surprise, rather, they are as expected. What should be of interest is the estimation procedure, the data used, and implications of the results for agricultural risk management.

It is hoped that this paper will stimulate future research efforts in improving estimation, data, and risk management related to the land size phenomenon. The estimation procedure presented here was done as a private sector initiative, but should be evaluated on its academic and technical merits. Moreover, the data utilized is not readily accessible to agricultural researchers, so replication of these findings is not likely. Although one could compare NASS variance inflation factors at the county and crop reporting district level with results from insurance data. Lastly, the implications for risk management contained in this paper would hopefully impact both the rating and underwriting of the MPCI program.

Figure 8 — Iowa Non-Irrigated Grain Corn

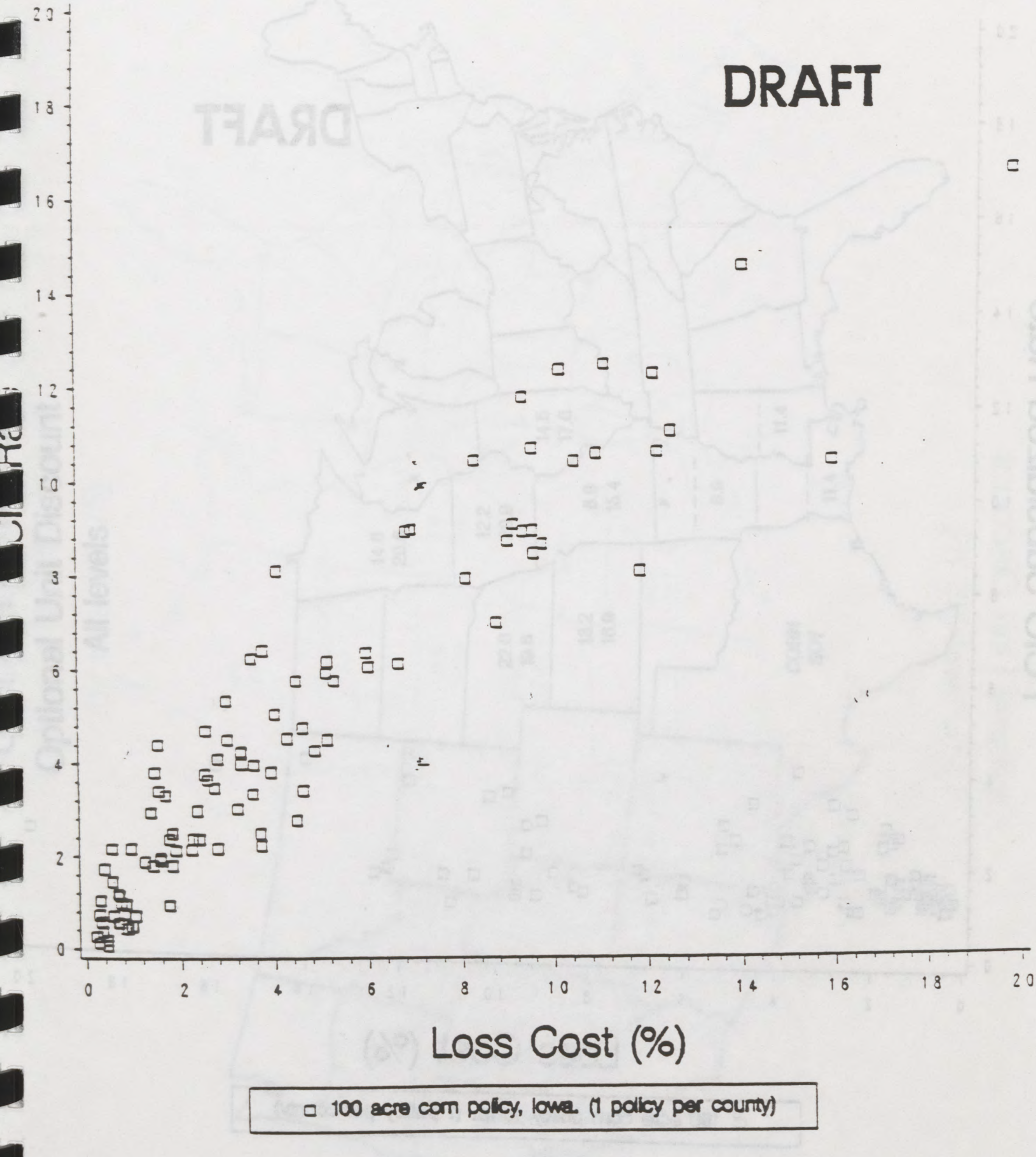
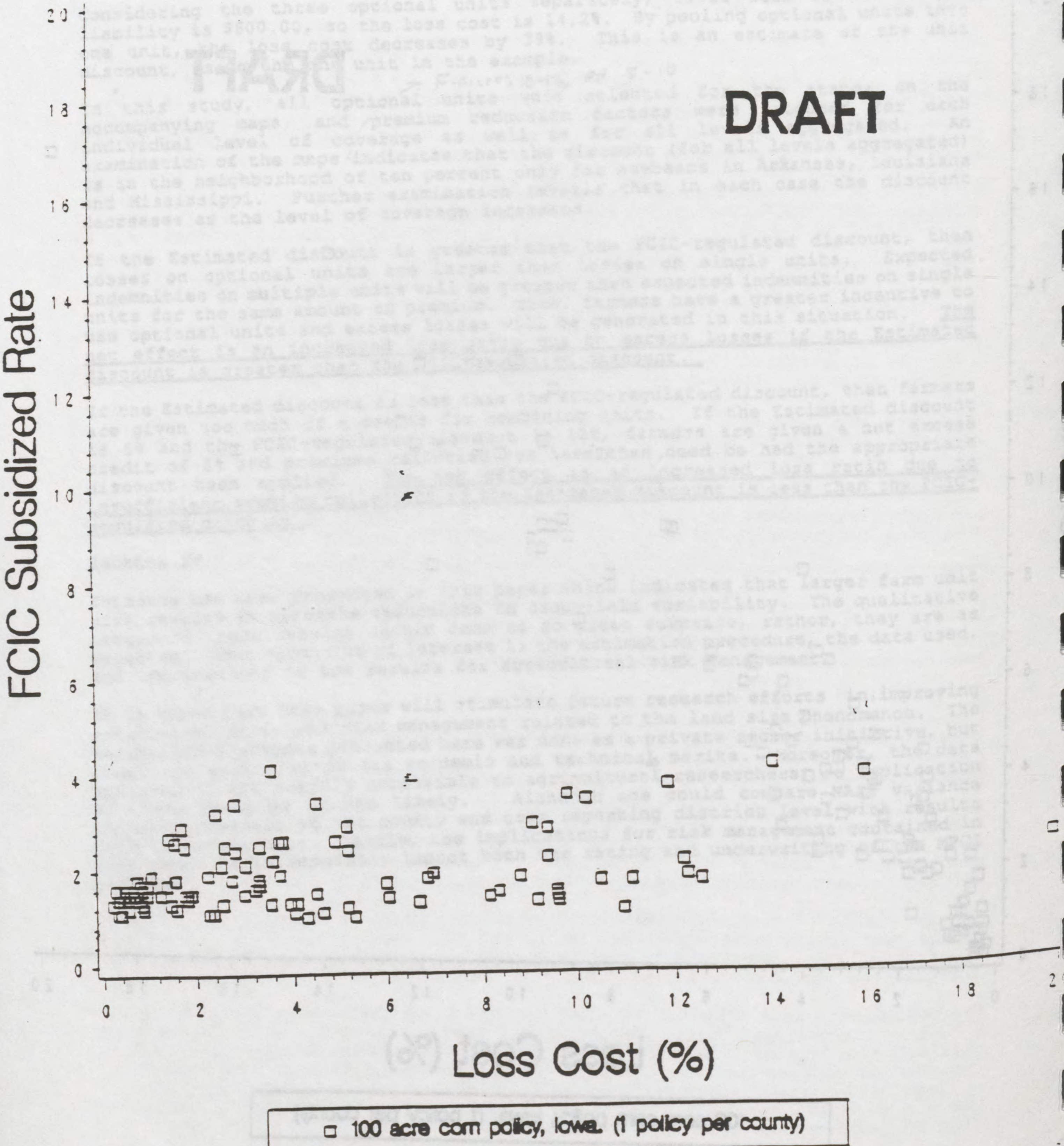


Figure 9 – Iowa Non–Irrigated Grain Corn

DRAFT



Optional Unit Discount
All levels

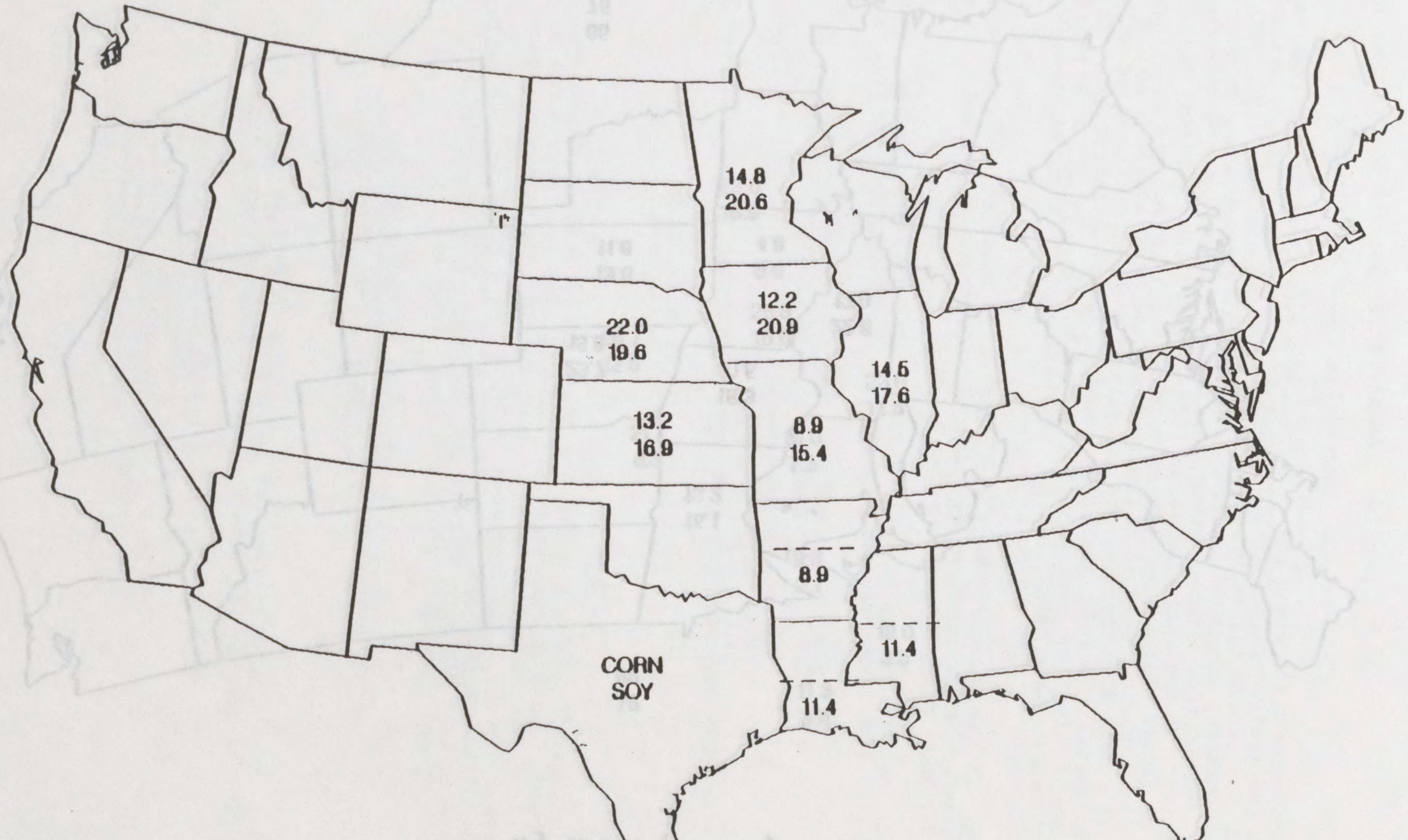


Figure 10

CORN
Optional Unit Discount
by level (65 75)

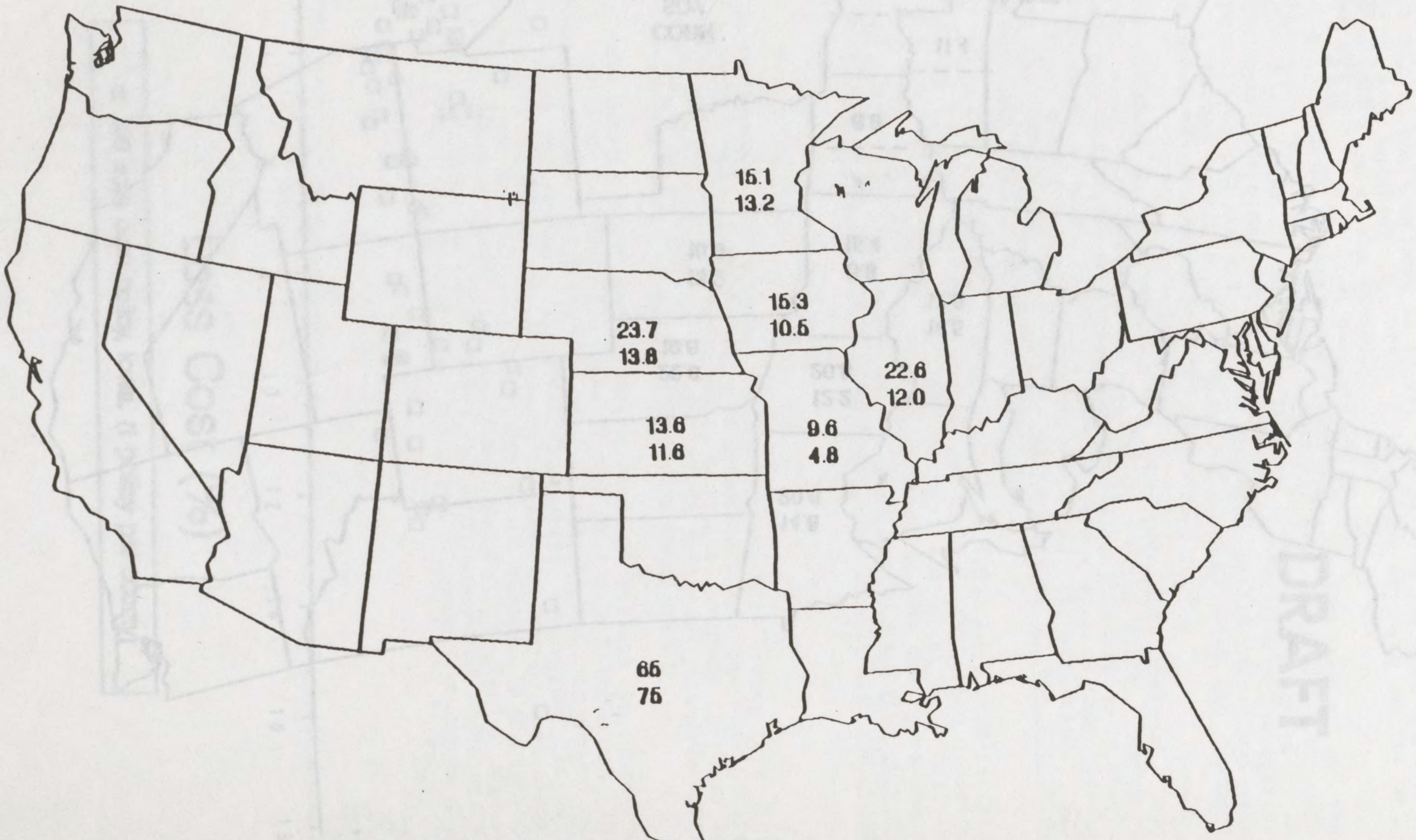


Figure 11

SOYBEANS

Optional Unit Discount

by level (65 75)

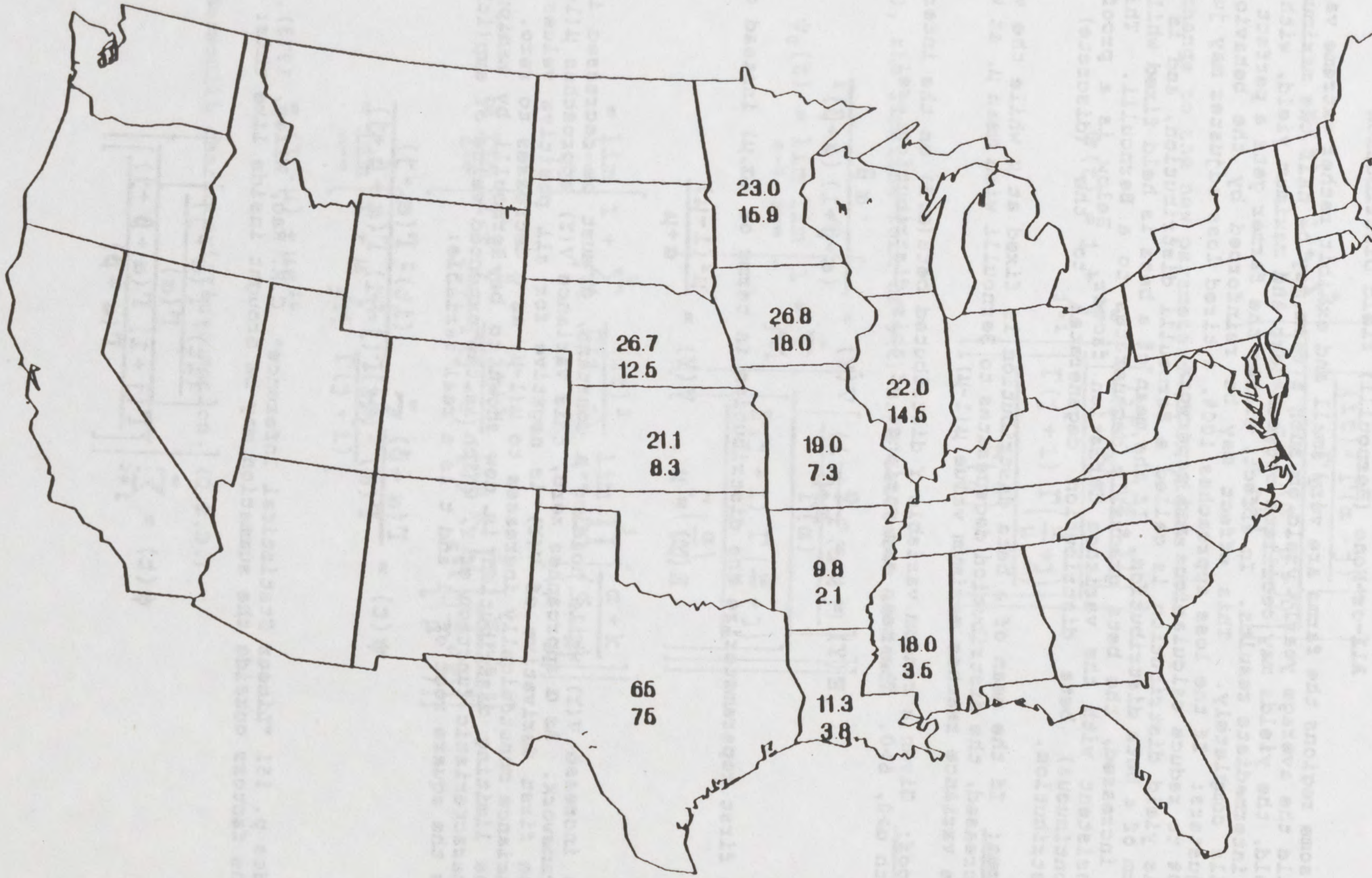


Figure 12

Appendix

All-or-None (Bernoulli) Yield Distribution

In some regions the farms are very small and exhibit rather extreme variability. While the average yearly yield on such a farm may be half the maximum possible yield, the yields may oscillate between zero and maximum yield, with virtually no intermediate results. In effect, either the farmer gets a perfect crop or he fails completely. This effect may be reinforced by the behavior of loss adjusters: If the loss approaches 100%, a tired loss adjuster may just call it 100% to reduce calculations and paperwork.

This yield distribution is called a Bernoulli distribution, and is a limiting form of a beta distribution. If the mean of a beta is held fixed while variance is increased, the beta gradually degenerates to a Bernoulli. This idea is consistent with the variance inflation factor. Below is a proof that the (continuous) beta distribution degenerates to the (discrete) Bernoulli distribution.

Lemma: If the mean of a beta distribution is fixed at μ while the variance is increased, the distribution degenerates to Bernoulli with mean μ, at which point the variance reaches maximum value $\mu(1-\mu)$.

Proof: Given a random variable Y distributed beta(α,β) on the interval (0,1), with $\alpha>0$, $\beta>0$. The mean and variance of this distribution are:

$$E(Y) = \mu = \frac{\alpha}{\alpha+\beta}, \quad V(Y) = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$$

We first reparameterize the distribution in terms of (α,μ) instead of (α,β):

$$E(Y) = \mu, \quad V(Y) = \frac{\mu^2(1-\mu)}{\alpha+\mu}$$

To increase V(Y) while holding μ constant, α must be decreased in this new framework. As α approaches zero, this variance V(Y) approaches $\mu(1-\mu)$. Since the first derivative of V(Y) is negative for all positive values of α, the variance monotonically increases to $\mu(1-\mu)$ as α decreases to zero.

The limiting distribution is now shown to be Bernoulli by manipulating the characteristic function of Y, which is the expected value of exp(itY), where i is the square root of -1 and t is a real variable:

$$\psi(t) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)} \sum_{j=0}^{\infty} \frac{(it)^j\,\Gamma(\alpha+j)}{\Gamma(j+1)\,\Gamma(\alpha+\beta+j)}$$

(See p. 151 "Linear Statistical Inference", C.R. Rao, Wiley, 1973).

The factors outside the summation may be brought inside like this:

$$\psi(t) = \sum_{j=0}^{\infty} \left[\frac{\left[\dfrac{(it)^j\,\Gamma(\alpha+j)}{\Gamma(\alpha)}\right]}{\left[\dfrac{\Gamma(j+1)\,\Gamma(\alpha+\beta+j)}{\Gamma(\alpha+\beta)}\right]} \right]$$

Removing the zeroth term for special treatment, we have:

$$\psi(t) = 1 + \sum_{j=1}^{\infty} \left[\frac{\left[\dfrac{(it)^j \, \Gamma(\alpha + j)}{\Gamma(\alpha)} \right]}{\left[\dfrac{\Gamma(j+1) \; \Gamma(\alpha + \beta + j)}{\Gamma(\alpha + \beta)} \right]} \right]$$

Making the change to the new parameters (α, μ) yields this:

$$\psi(t) = 1 + \sum_{j=1}^{\infty} \left[\frac{\left[\dfrac{(it)^j \, \Gamma(\alpha + j)}{\Gamma(\alpha)} \right]}{\left[\dfrac{\Gamma(j+1) \; \Gamma\left(\dfrac{\alpha}{\mu} + j\right)}{\Gamma\left(\dfrac{\alpha}{\mu}\right)} \right]} \right]$$

We seek $\psi_0(t)$, the limiting form of $\psi(t)$ as α decreases to zero:

$$\psi_0(t) = \lim_{\alpha \to 0} \lim_{n \to \infty} \left[1 + \sum_{j=1}^{\infty} \left[\frac{\left[\dfrac{(it)^j \, \Gamma(\alpha + j)}{\Gamma(\alpha)} \right]}{\left[\dfrac{\Gamma(j+1) \; \Gamma\left(\dfrac{\alpha}{\mu} + j\right)}{\Gamma\left(\dfrac{\alpha}{\mu}\right)} \right]} \right] \right]$$

$$= \lim_{n \to \infty} \left[1 + \sum_{j=1}^{n} \frac{(it)^j}{\Gamma(j+1)} \lim_{\alpha \to 0} \prod_{k=0}^{j} \left[\frac{\alpha + k}{\left(\dfrac{\alpha}{\mu} + k \right)} \right] \right]$$

$$= \lim_{n \to \infty} \left[1 + \sum_{j=1}^{n} \frac{(it)^j}{\Gamma(j+1)} \lim_{\alpha \to 0} \mu \prod_{k=1}^{j} \left[\frac{\alpha + k}{\left(\dfrac{\alpha}{\mu} + k \right)} \right] \right]$$

$$= \lim_{n \to \infty} \left[1 + \mu \sum_{j=1}^{n} \frac{(it)^j}{\Gamma(j+1)} \right]$$

$$= (1 - \mu) + \mu e^{it}$$

This is the Bernoulli characteristic function. (Q.E.D.)

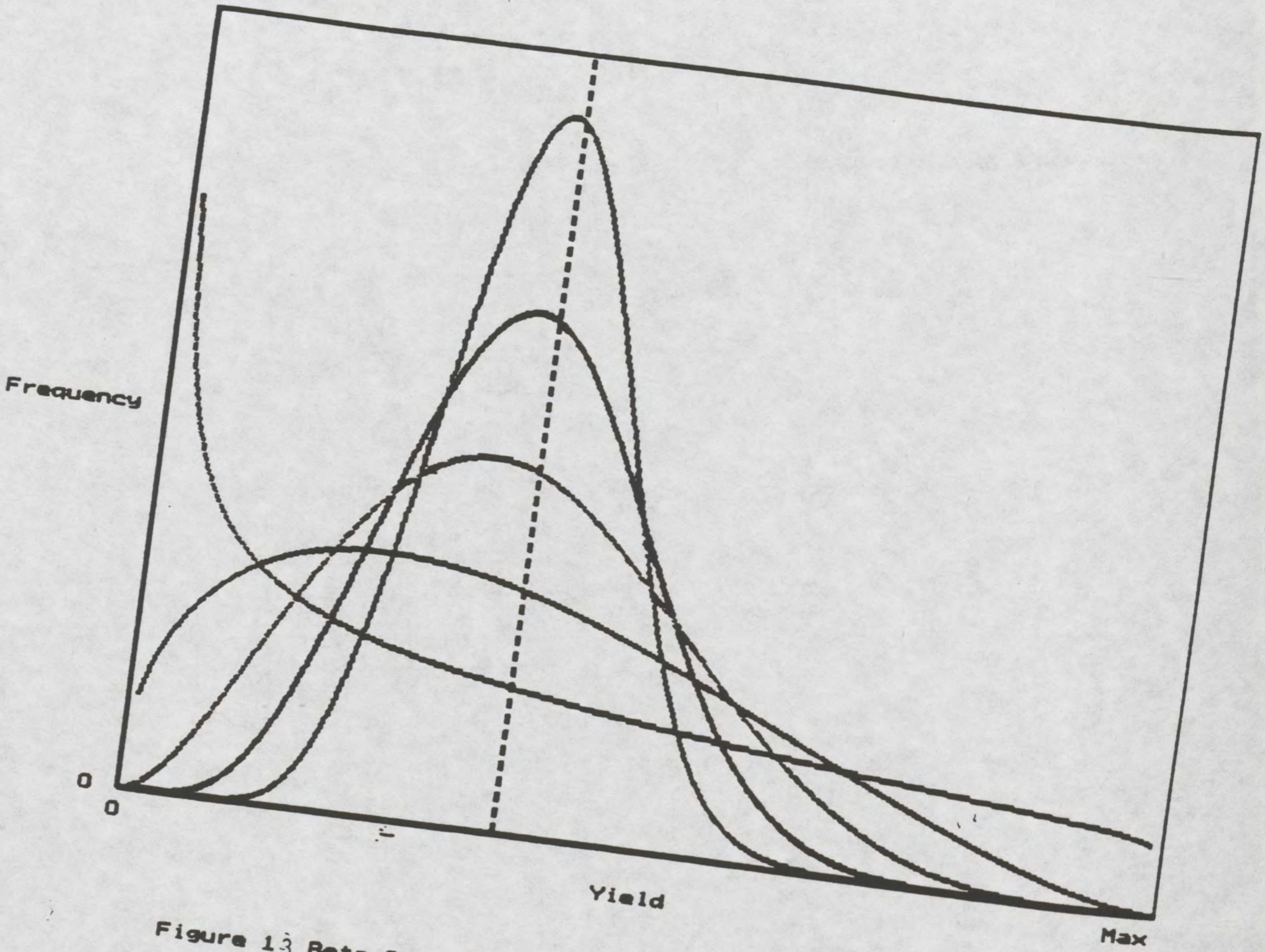


Figure 13 Beta Densities with Common Mean (Dotted Line)