

Estimation Risk when Theory Meets Reality

Sergio H. Lence and Dermot J. Hayes

Working Paper 93-WP 107

June 1993

Center for Agricultural and Rural Development
Iowa State University
Ames, Iowa 50011-1070

Journal Paper No. J-15066 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa.
Project No. 2994.

Sergio H. Lence is a CARD postdoctoral research associate; and Dermot J. Hayes is an associate professor of economics and head of the Trade and Agricultural Policy Division of CARD.

ESTIMATION RISK WHEN THEORY MEETS REALITY

Abstract

Estimation risk occurs in the almost universal situation where parameters of importance for decision making are not known with certainty. Bayes' criterion is the procedure consistent with expected utility maximization in the presence of estimation risk. Three interrelated problems in the presence of estimation risk are analyzed: (i) the choice of the utility-maximizing decision rule in a mean-variance framework, (ii) the calculation of certainty equivalent returns, and (iii) the valuation of additional sample information.

Key words: Bayes' decision criterion, estimation risk, expected utility, risk, uncertainty.

ESTIMATION RISK WHEN THEORY MEETS REALITY

Whenever economic analysis involves incorporating estimated parameters into theoretically derived decision rules, the optimal outcome will depend on the estimation procedure. Decisions such as optimal levels for the export tax (or import subsidy), output, resource allocation, and research expenditures or the optimal portfolio are usually determined in this manner and results are provided that appear robust to several decimal places. The individual involved in estimating these parameters is, however, left with a vague sense of unease. Relatively small changes in the estimation procedure or in the number of data observations can change the magnitude and even the sign of important decision variables. This problem is called estimation risk (Bawa, Brown, and Klein) and the traditional procedure of substituting the sample parameter estimates for the true but unknown parameters is known as the plug-in approach. The incorrect use of the plug-in approach can have important consequences [Klein et al., Chalfant, Collender, and Subramanian (CCS)]. The correct procedure is to acknowledge the presence of estimation risk both when developing the theoretical model and when incorporating the estimated parameters into the derived decision rule.

Several studies have explicitly accounted for estimation risk in the financial literature (e.g., Boyle and Ananthanarayanan; Bawa, Brown, and Klein; Coles and Loewenstein; Chen and Brown; Alexander and Resnick; Jorion; Frost and Savarino; Lence and Hayes). But despite the pervasiveness and the obvious importance of estimation risk in agricultural economics, the problem has been largely ignored in this area until recently. Two exceptions are CCS and Collender.

The purpose of this study is to reexamine three interrelated analytical problems and show how the presence of estimation risk justifies changes in the way each is solved. The first problem examined is the choice of the utility-maximizing decision rule in a mean-variance framework, the second is the calculation of certainty equivalent returns, and the third is the valuation of additional sample information. First, we set up the framework of analysis more formally and introduce our notation. Then, we examine each of the three problems in turn. Finally, we summarize the findings and draw conclusions.

Decision Making in the Presence of Estimation Risk

Consider a decision maker characterized by a von Neumann-Morgenstern utility function of terminal wealth $[U(\pi), U' > 0, U'' \leq 0]$. Let wealth be a function of a vector of random variables $x \equiv y_{n+1}$ and a decision vector l [i.e., $\pi(x, l)$]. If terminal wealth were known with certainty, the optimal decision vector l^* could be easily found by maximizing utility with respect to l , i.e.,

$$(1.1) \quad l^* \equiv \operatorname{argmax}_{l \in \Lambda} U[\pi(x, l)],$$

where Λ is the set of all possible decisions. But because x is random, wealth is also random and utility is not known with certainty at the decision time; therefore, it is generally impossible to choose a decision vector that maximizes utility for all possible realizations of the random vector x . Under uncertainty, the optimal decision vector l^* is usually considered to be the one that maximizes expected utility. According to the expected utility paradigm, the decision maker's objective function is represented by

$$(1.2) \quad \max_{l \in \Lambda} E_{x|\theta}\{U[\pi(x, l)]\} = \max_{l \in \Lambda} \int_X U[\pi(x, l)] f_{x|\theta}(x|\theta) dx,$$

where $E(\cdot)$ represents the expectation operator, X is the domain of x , $f_{x|\theta}(x|\theta)$ is the probability density function (pdf) of x given θ , and θ is a known vector of parameters that characterizes the pdf. By letting $\mathcal{L}(\theta, l) \equiv -E_{x|\theta}(U)$, the objective function in (1.2) can be alternatively expressed as

$$(1.3) \quad \max_{l \in \Lambda} E_{x|\theta}\{U[\pi(x, l)]\} = \min_{l \in \Lambda} \mathcal{L}(\theta, l).$$

The function $\mathcal{L}(\theta, l)$ is called the "loss function" in statistical decision theory.

As long as the parameter vector θ is known, it is relatively straightforward to find a decision vector l^* that minimizes the loss function (or maximizes the expected utility). If θ is not known, however, the optimization as stated in (1.2) or (1.3) cannot be performed because, in general, there is no decision vector that minimizes the loss function for all possible values of θ . In this situation, there is estimation risk. Under estimation risk, the method by which an optimal decision vector can be obtained in a manner consistent with expected utility maximization is Bayes' decision criterion (Klein et al., DeGroot).

Bayes' decision criterion can be summarized as follows. Let the decision maker's prior beliefs regarding the parameter vector be represented by the prior pdf $p(\theta)$. Consider having available $y \equiv (y_1, \dots, y_n)$, a sample of size n generated by the same process that generates x . The optimal (Bayes') decision rule l^B is the solution to the objective function

$$(1.4) \quad \min_{l \in \Lambda} E_{p(\theta)}\{E_{y|\theta}[\mathcal{L}(\theta, l)]\} = \min_{l \in \Lambda} \int_{\Theta} \int_Y \mathcal{L}(\theta, l) f_{y|\theta}(y|\theta) p(\theta) dy d\theta,$$

where Y denotes the sample space and $f_{y|\theta}(y|\theta)$ is the pdf of y given θ . If no data are available, the objective function is simply $\min_{l \in \Lambda} E_{p(\theta)}[\mathcal{L}(\theta, l)]$. Hence, lack of sample data does not prevent the agent from being able to make an optimal choice; if sample data are available, however, the agent will use them to improve the information about the unknown true parameter vector θ and make a more informed decision.

The function $E_{p(\theta)}\{E_{y|\theta}[\mathcal{L}(\theta, l)]\}$ is called Bayes' risk of a decision rule l with respect to the prior pdf $p(\theta)$. Comparison of (1.2) and (1.4) reveals that they are entirely analogous: in (1.2), expectations are taken to eliminate the random vector x , whereas in (1.4) expectations are taken to eliminate the agent's uncertainty regarding the true but unknown parameter vector θ . We will exploit this analogy later to provide the intuition of why decisions under estimation risk that are unbiased estimators of the optimal decisions in the absence of estimation risk (l^*) are generally suboptimal according to the expected utility paradigm.

Bayes' decision rule can be calculated in an alternative way. If $\mathcal{L}(\theta, l)$ is bounded below, Fubini's theorem can be invoked to interchange the order of integration in Bayes risk and get

$$(1.5) \quad E_{p(\theta)}\{E_{y|\theta}[\mathcal{L}(\theta, l)]\} = \int_Y \int_{\Theta} \mathcal{L}(\theta, l) f_{y|\theta}(y|\theta) p(\theta) d\theta dy$$

$$(1.5') \quad = \int_Y \left[\int_{\Theta} \mathcal{L}(\theta, l) f_{p(\theta|y)}(\theta|y) d\theta \right] f_y(y) dy$$

$$= E_y\{E_{p(\theta|y)}[\mathcal{L}(\theta, l)]\},$$

where $f_{p(\theta|y)}(\theta|y)$ is the posterior pdf of θ and $f_y(y)$ is the marginal pdf of y . Bayes' theorem is used to obtain (1.5') from (1.5).¹ Because $f_y(y)$ is positive, minimizing Bayes risk with respect to l is equivalent to minimizing the term inside brackets in (1.5') with respect to l , i.e.,

$$(1.6) \quad \min_{l \in \Lambda} E_{p(\theta)}\{E_{y|\theta}[\mathcal{L}(\theta, l)]\} = \min_{l \in \Lambda} E_{p(\theta|y)}[\mathcal{L}(\theta, l)].$$

The term $E_{p(\theta|y)}[\mathcal{L}(\theta, l)]$ is called the posterior expected loss of the decision l . The left-hand side of (1.6) is the "normal form" of Bayes' criterion, whereas the right-hand side is the "extensive form" (Raiffa and Schlaifer).

When written as in (1.5'), an important sampling property of Bayes' decision rule l^B is highlighted. Expression (1.5') tells us that Bayes' rule yields, on average, the minimum loss for the prior $p(\theta)$. That is, if we were able to take an infinite number of y -type data samples and average the corresponding losses, Bayes' decision rule is such that the average loss is minimum for the prior $p(\theta)$. This result relates to the problem analyzed by CCS, i.e., the sampling properties of alternative decisions under estimation risk. Expressions (1.4) and (1.5') tell us that

¹According to Bayes' theorem, $p(a, e) = p(e) p(a|e) = p(a) p(e|a)$, and therefore $p(a|e) = p(a) p(e|a)/p(e)$, where $p(a, e)$ is the joint pdf of any pair of random variables a and e , $p(a|e)$ and $p(e|a)$ are the conditional densities, and $p(a)$ and $p(e)$ are the marginal densities.

Bayes decisions are the ones that yield, on average, the smallest loss (i.e., the greatest expected utility).

We now provide an intuitive argument as to why optimal decisions under risk need not be unbiased and then extend the intuition to decisions in the presence of estimation risk. Consider the utility function

$$(1.7) \quad U[\pi(x, l)] = U[pq - c(q)], \quad c' > 0, \quad c'' > 0, \quad c''' > 0, \quad c(0) = 0,$$

where p denotes price, q represents production, and $c(q)$ is the cost function. In this problem, we have $x = p$ and $l = q$. If price is known with certainty, the quantity that maximizes utility is q^c , where q^c satisfies the first-order condition $p = c'(q^c)$. As explained before, however, production cannot be chosen in this way when p is random. If expected utility theory is applied to optimize production in the presence of random prices, Sandmo has shown that the decision maker will produce q^* , where q^* is such that $E(p) \geq c'(q^*)$. In other words, because of price uncertainty, the agent will produce strictly less (if $U'' < 0$) or the same (if $U'' = 0$) than would be produced if it were known with certainty that $p = E(p)$ at the decision time (i.e., $q^* \leq q^c[E(p)]$). The magnitude of the difference between q^* and $q^c[E(p)]$ depends on the utility function. By application of Jensen's inequality (Berger, p. 40), it can be shown that $E[q^c(p)] < q^c[E(p)]$; the magnitude of this difference is independent of the utility function. Therefore, $q^* \neq E[q^c(p)]$ in general, i.e., optimal production under uncertainty (q^*) is generally a biased estimator of the optimal production under certainty ($E[q^c(p)]$). An alternative interpretation is that if the agent under uncertainty produced $q = E[q^c(p)]$ rather than $q = q^*$, he would generally be making a suboptimal choice according to the expected utility paradigm. This suboptimality arises because, according to the expected utility paradigm, the optimal decision under uncertainty (l^*) is not an estimate of the optimal decision under certainty (l^c) (Klein et al.).

Now, assume for simplicity that the loss function corresponding to (1.7) is

$$(1.8) \quad \mathcal{L}(\theta, l) = \mathcal{L}[E(p)q - c(q)], \mathcal{L}' < 0, \mathcal{L}'' \geq 0.$$

If $\theta = E(p)$ is known with certainty, the production level that minimizes the loss is q^* . If $E(p)$ is not known with certainty, however, some arbitrary optimizing decision method must be used. If (consistent with expected utility maximization) Bayes' criterion is used, we can draw inferences regarding the optimal (Bayes) production under estimation risk (q^B). Bayes production under estimation risk is such that $E_{p[E(p)|y]}[E(p)] > c'(q^B)$ and in general $q^B \neq E_{p[E(p)|y]} \{q^*[E(p)]\}$, where $E_{p[E(p)|y]} \{q^*[E(p)]\}$ is the unbiased estimator of the optimal production level in the absence of estimation risk. Hence, if the agent produced $q = E_{p[E(p)|y]} \{q^*[E(p)]\}$ under estimation risk, he would generally be making a suboptimal decision from the standpoint of Bayes' criterion.

In summary, if the optimization problem under estimation risk is solved in a manner consistent with the expected utility paradigm, decisions that are unbiased predictors of the decisions that had been taken in the absence of estimation risk are generally suboptimal. This is because, in a framework consistent with expected utility maximization, the optimal decision under estimation risk (l^B) is not an estimate of the optimal decision in the absence of estimation risk (l^*) (Klein et al.).

In the next section, we will apply the concepts just introduced to mean-variance land allocation under estimation risk and we will derive the Bayesian allocation vector. It will be seen that the Bayesian allocation is a biased predictor of the optimal allocation in the absence of estimation risk but yields a smaller expected loss than does an unbiased allocation.

Land Allocation under Estimation Risk

In a recent contribution, CCS analyzed the sampling properties of the land allocation vector obtained using the plug-in approach in a mean-variance framework. Among other results, they showed that the plug-in allocation vector is a biased estimator of the optimal allocation in the absence of estimation risk. They pursued this matter further and proposed an alternative allocation

vector, also based on the sample mean vector and covariance matrix but featuring unbiasedness. They proved that the proposed allocation yields greater expected utility than does the plug-in allocation. By explicitly incorporating estimation risk, CCS improved on the plug-in approach. However, as we discussed in the previous section, the optimal decision vector need not be unbiased. The purpose of this section is to use the tools presented earlier to derive the utility-maximizing decision vector for the land allocation problem examined by CCS.

In the notation employed in the previous section, CCS's land allocation problem can be summarized as follows:

$$(2.1) \quad U[\pi(x, l)] = -\exp[-r \pi(x, l)],$$

$$(2.2) \quad \pi(x, l) = l' x,$$

$$(2.3) \quad l = (l_1, \dots, l_k)'; \quad l_i \geq 0 \text{ for } i = 1, \dots, k; \quad l' l_k \leq L,$$

$$(2.4) \quad x \equiv y_{n+1} = (x_1, \dots, x_k)',$$

$$(2.5) \quad f_{x|\theta}(x|\theta) = N_k(x|\mu, \Sigma),$$

where l_k is a $(k \times 1)$ vector of ones and $N_k(\cdot)$ represents the k -variate normal pdf. The utility function is characterized by constant absolute risk aversion, with r denoting the Arrow-Pratt coefficient of absolute risk aversion. Terminal wealth $[\pi(x, l)]$ equals the sum of the product of returns per acre (x_i) times the corresponding number of acres planted (l_i) with each crop. The decision vector (l) comprises the land allocated to each crop. The restrictions on the decision vector are that (i) the number of acres planted with any crop cannot be negative ($l_i \geq 0$), and (ii) the total number of acres planted cannot exceed the total farm acreage ($l' l_k \leq L$). The random vector is

that of returns per acre ($x \equiv y_{n+1}$), which is assumed to follow a k -variate normal distribution with mean vector μ and covariance matrix Σ .

Under the stated assumptions, the loss function has the closed-form solution

$$(2.6) \quad \mathcal{L}(\theta, l) = \exp(-r l' \mu + \frac{1}{2} r^2 l' \Sigma l).$$

If the mean vector μ and the covariance matrix Σ were known by the decision maker, minimization of the loss function (2.6) with respect to the feasible decision vector l would yield the optimal land allocation under uncertainty l^* . Noticing that this loss function is monotonically decreasing in $(l' \mu - r l' \Sigma l/2)$, the solution can be easily found by maximization of the Lagrangian function

$$(2.7) \quad \mathcal{L} = l' \mu - \frac{1}{2} r l' \Sigma l + \lambda (L - l' l_k),$$

where λ is the Lagrangian multiplier corresponding to the total acreage restriction. Assuming an interior solution and that land is constraining (i.e., $\lambda > 0$), the first-order conditions for the maximization of (2.7) can be expressed in matrix form as

$$(2.8) \quad \begin{bmatrix} 0 & l_k' \\ l_k & r \Sigma \end{bmatrix} \begin{bmatrix} \lambda \\ l^* \end{bmatrix} = \begin{bmatrix} L \\ \mu \end{bmatrix}.$$

The premultiplying $[(k+1) \times (k+1)]$ matrix in the left-hand side of (2.8) is symmetric and can be inverted by application of Theorem A.3.3 in Anderson (p. 594) to yield

$$(2.9) \quad \begin{bmatrix} 0 & l_k' \\ l_k & r \Sigma \end{bmatrix}^{-1} = \begin{bmatrix} B & C' \\ C & 1/r D \end{bmatrix},$$

where: $B = -r (l_k' \Sigma^{-1} l_k)^{-1}$,
 $C = \Sigma^{-1} l_k (l_k' \Sigma^{-1} l_k)^{-1}$,
 $D = \Sigma^{-1} - \Sigma^{-1} l_k l_k' \Sigma^{-1} (l_k' \Sigma^{-1} l_k)^{-1}$.

By employing the result in (2.9), the optimal land allocation in the absence of estimation risk can be shown to equal

$$(2.10) \quad l^* = \frac{1}{r} D \mu + C L.$$

But in the presence of estimation risk regarding μ and Σ , the land allocation problem cannot be solved by direct minimization of (2.6). We will therefore apply Bayes' decision criterion to calculate the land allocation vector under estimation risk l^B . To do so, it is necessary to postulate a prior pdf for the unknown parameters. To simplify the exposition and to avoid criticisms regarding the reasonability and subjectivity of any particular informative prior, we will postulate a "non-informative" or "diffuse" prior for μ and Σ . This is equivalent to hypothesizing that the decision maker has no information about the parameters other than that provided by $y \equiv (y_1, \dots, y_n)'$, a $(k \times n)$ matrix consisting of n past observations on the vector of crop returns. Under the combined assumptions of a non-informative prior and a k -variate normal distribution of returns, the posterior pdf of the mean vector and the covariance matrix is k -variate normal-Wishart of the form (DeGroot)

$$(2.11) \quad f_{p(\theta|y)}(\theta|y) = \text{NW}_k(\mu, \Sigma^{-1}|\hat{\mu}, \hat{\Sigma}^{-1}, n, n-1) \\ = N_k(\mu|\hat{\mu}, \Sigma/n) W_k[\Sigma^{-1}|(n-1)^{-1} \hat{\Sigma}^{-1}, n-1],$$

where: $\hat{\mu} = y \iota_n / n$,

$$\hat{\Sigma} = (y - \hat{\mu} \iota_n)' (y - \hat{\mu} \iota_n) / (n-1),$$

and $W_k(\cdot)$ denotes the k -variate Wishart pdf. Given the posterior pdf (2.11), the posterior expectations of the mean vector μ and the covariance matrix Σ are $E_{p(\theta|y)}(\mu) = \hat{\mu}$ and $E_{p(\theta|y)}(\Sigma) = (n-1)/(n-k-2) \hat{\Sigma}$, respectively (Anderson, p. 270).

An approximate solution to the posterior expectation of the loss function (2.6) can be obtained by adopting a relatively mild assumption.² This assumption consists of approximating the k -variate Student- t pdf $S_k(x|\hat{\mu}, \Sigma, n-k)$ with the k -variate normal distribution $N_k[x|\hat{\mu}, (n-k)/(n-k-2) \Sigma]$ (Shimizu, p. 199). To this end, it is more helpful to analyze the problem in terms of the predictive pdf as follows. The posterior expected loss function can be rewritten as

$$\begin{aligned}
 (2.12) \quad E_{p(\theta|y)}[\mathcal{L}(\theta, I)] &= \int_{\Theta} \mathcal{L}(\theta, I) f_{p(\theta|y)}(\theta|y) d\theta \\
 &= \int_{\Theta} \left\{ - \int_X U[\pi(x, I)] f_{x|\theta}(x|\theta) dx \right\} f_{p(\theta|y)}(\theta|y) d\theta \\
 &= - \int_X U[\pi(x, I)] \left[\int_{\Theta} f_{x|\theta}(x|\theta) f_{p(\theta|y)}(\theta|y) d\theta \right] dx \\
 &= - \int_X U[\pi(x, I)] f_{x|y}(x|y) dx \\
 (2.12') \quad &= - E_{x|y} \{ U[\pi(x, I)] \},
 \end{aligned}$$

where $f_{x|y}(x|y)$ is the predictive pdf of x given y (Aitchison and Dunsmore). Note that the predictive pdf does not depend on the unknown parameter vector θ .

From (2.12') and (1.6), it follows that, in general, Bayes' decision vector can be alternatively obtained by maximizing the predictive expected utility with respect to the feasible decision vector. For the scenario being analyzed, the predictive pdf is k -variate Student- t of the

²Strictly speaking, the posterior expectation of the loss function (2.6) is not a real number when the pdf is given by expression (2.11).

form $S_k(x|\hat{\mu}, \Sigma'', n-k)$, where $\Sigma'' = (1 + 1/n) (n-1)/(n-k) \hat{\Sigma}$. The closed-form approximation (2.13) can be obtained by approximating the Student- t pdf with the normal pdf as indicated above.

$$(2.13) \quad E_{x|y}\{U[\pi(x, l)]\} \equiv -\exp[-r l' \hat{\mu} + \frac{r^2 (1 + 1/n) (n-1)}{2 (n-k-2)} l' \hat{\Sigma} l].$$

Application of the standard optimization results to the right-hand side of (2.13) yields the (approximate) Bayes allocation

$$(2.14) \quad l^B = \frac{(n-k-2)}{r (1 + 1/n) (n-1)} \hat{D} \hat{\mu} + \hat{C} L,$$

$$\text{where: } \hat{C} = \hat{\Sigma}^{-1} l_k (l_k' \hat{\Sigma}^{-1} l_k)^{-1},$$

$$\hat{D} = \hat{\Sigma}^{-1} - \hat{\Sigma}^{-1} l_k l_k' \hat{\Sigma}^{-1} (l_k' \hat{\Sigma}^{-1} l_k)^{-1}.$$

For the sake of comparison, the plug-in land allocation (\hat{l}) and the land allocation advocated by CCS (\tilde{l}) are reported below as expressions (2.15) and (2.16), respectively.³

³Interestingly, a decision vector almost identical to CCS's land allocation (\tilde{l}) is obtained by minimizing the posterior expectation of the loss function (2.17)

$$(2.17) \quad \mathcal{E}(\theta, l) = -r l' \mu + \frac{1}{2} r^2 l' \Sigma l.$$

This loss function is equal to the exponent of the problem's actual loss (2.6). Minimization of the posterior expectation of (2.17), i.e.,

$$(2.18) \quad \min_{l \in \Lambda} E_{x|y}(-r l' \mu + \frac{1}{2} r^2 l' \Sigma l) = \min_{l \in \Lambda} [-r l' \hat{\mu} + \frac{r^2 (n-1)}{2 (n-k-2)} l' \hat{\Sigma} l],$$

yields the land allocation

$$(2.19) \quad \tilde{l} = \frac{(n-k-2)}{r (n-1)} \hat{D} \hat{\mu} + \hat{C} L.$$

It can be seen that the only difference between \tilde{l} and \hat{l} is that the former contains the factor $(n-k-2)$, which replaces $(n-k-1)$ in the latter. This subtle difference is due to the land constraint ($l' l_k \leq L$); because of the way \tilde{l} is constructed, \tilde{l} would be identical to \hat{l} if total acreage were not constraining. This result indicates that \tilde{l} is approximately (exactly) Bayesian for the loss (2.17) under (no) land constraint. But the loss function for the

$$(2.15) \hat{l} = \frac{1}{r} \hat{D} \hat{\mu} + \hat{C} L.$$

$$(2.16) \tilde{l} = \frac{(n - k - 1)}{r(n - 1)} \hat{D} \hat{\mu} + \hat{C} L.$$

CCS have shown that \hat{l} is a biased estimator of l^* and that \tilde{l} is unbiased. By applying their technique, it is straightforward to show that l^B is a biased estimator of l^* . But l^B yields the maximum predictive expected utility by construction; therefore, l^B is to be preferred to either \hat{l} or \tilde{l} on the basis of the expected utility paradigm.

Table 1 exemplifies how different the allocations obtained by means of (2.14) through (2.16) may be.⁴ The sample mean vector and the covariance matrix used to build Table 1 are those reported in the classical article by Freund, that is,

$$(2.20) \hat{\mu} = (100, 100, 100, 100)',$$

$$(2.21) \hat{\Sigma} = \begin{bmatrix} 7304.69 & 903.89 & -688.73 & -1862.05 \\ & 620.16 & -471.14 & 110.43 \\ & & 1124.64 & 750.69 \\ & & & 3689.53 \end{bmatrix}.$$

The mean vector (2.20) and the covariance matrix (2.21) are expressed in unit levels; activity 1 (potatoes) requires 1.199 acres per unit, activity 2 (corn) requires 1.382 acres per unit, activity 3

problem initially posed is (2.6), not (2.17). Moreover, the loss function (2.17) does not represent the same preferences as does the loss (2.6) because, under expected utility theory, the utility function is unique only up to a positive linear transformation (Hey, p. 36) and (2.17) is not a linear transformation of (2.6). Hence, we can conclude that neither \hat{l} nor \tilde{l} can be Bayes solutions for the land allocation problem initially posed.

⁴The negative plug-in and CCS land allocations are corner solutions under the nonnegativity restriction. We preferred to report the unrestricted plug-in and CCS solutions because they are directly obtained by means of expressions (2.15) and (2.16), respectively.

(beef) requires 1.400 acres per unit, and activity 4 (fall cabbage) requires 0.482 acres per unit.⁵ The values of the coefficient of absolute risk aversion that we employed reflect moderately low to moderately high risk aversion. The number of observations chosen for the simulations ($n = 7$) is low but not uncommon.⁶

The differences among the three alternative allocations are smaller in absolute value as the degree of absolute risk aversion is greater. Also, the differences are smaller in relative terms when the total acreage is greater. It is also evident that CCS allocations are more similar to Bayes allocations than to plug-in allocations. But the most salient characteristic of the figures reported in Table 1 is the noticeable differences among the allocations obtained by means of expressions (2.14) through (2.16). The intuition here can best be grasped by comparing allocations for potatoes and cabbage on one hand and for beef and corn on the other. Potatoes and cabbage have a relatively larger per acre expected return but also are considerably more risky than are beef and corn. In the traditional plug-in approach, it is assumed that the producer uses seven years of data to derive the exact mean and variance of the individual returns, i.e., he "plugs in" the sample estimates into the first-order conditions as if they were the true parameters. In the Bayes solution, the producer realizes that with only seven years of data the estimators of the mean, variances, and covariances themselves are quite uncertain. This additional source of uncertainty causes the producer to grow much more corn and beef and much less cabbage and potatoes than the levels prescribed by the plug-in approach. The message of Table 1 is that, unless the number of observations is sufficiently large (and the number of activities small), it cannot be taken for granted that the allocations obtained by employing (2.14) through (2.16) will be similar. Therefore, given the expected utility-maximizing properties of Bayes' allocation and the fact that (2.14) is no more difficult to calculate than either (2.15) or (2.16), Bayes' allocation is the one that should be used.

⁵The original requirement of 2.776 acres per unit of beef was too high relative to the requirements of the other activities, which led to corner Bayesian solutions for reasonable values of the coefficient of absolute risk aversion and total acreage. Because we wanted to avoid corner solutions (at least for the Bayesian allocation), we substituted 1.4 for Freund's 2.776.

⁶We do not report results of simulations performed for a greater number of observations, say $n = 30$, because they led to corner solutions for Bayes allocation. For $n = 30$ and allowing for negative land allocations, the differences among the three alternatives are smaller than for $n = 7$ but still quite noticeable.

Certainty Equivalent Returns

A value often computed in studies regarding uncertainty is the certainty equivalent return (CER). The CER of a risky investment is the return on a risk-free investment that leaves the decision maker indifferent between selecting the risky choice and accepting the riskless CER. CER bears a monotonically increasing relationship with expected utility, i.e., the risky investment I has a greater CER than the risky investment II if and only if I yields greater expected utility than does II. Hence, analyzing the CERs of alternative risky prospects allows us to draw inferences about the expected utility of these prospects.

In the absence of estimation risk, CER is obtained as the root of equality (3.1), which we will call CER^n (the superscript n standing for "no" estimation risk).

$$(3.1) \quad U(CER^n) = E_{x|\theta} \{U[\pi(x, l)]\}.$$

For the land allocation example analyzed in the previous section, we have

$$(3.2) \quad CER^n(l^*) = l^* \mu - \frac{1}{2} r l^* \Sigma l^*$$

as the CER corresponding to the optimal decision vector in the absence of estimation risk (i.e., l^*).

In the presence of estimation risk, however, CER cannot be calculated from an expression like (3.1) because such an expression depends on the true but unknown parameter vector θ . If CER^n were the certainty equivalent under estimation risk, then $E_{x|\theta} \{U[\pi(x, l)]\}$ would be known to the decision maker, thus contradicting the definition of estimation risk. CER in the presence of estimation risk (CER^e) is the root of the equality (3.3),

$$(3.3) \quad U(CER^e) = E_{x|y} \{U[\pi(x, l)]\},$$

as opposed to (3.1). For the land allocation example, CER^e of Bayes' decision vector is

$$(3.4) \quad CER^e(l^B) = l^B, \hat{\mu} - \frac{r(1+1/n)(n-1)}{2(n-k-2)} l^B, \hat{\Sigma} l^B.$$

From (3.1) and (3.3), it is clear that CER^n is not directly comparable to CER^e . CER^n is a function of the true parameter vector θ (which is known in the absence of estimation risk), whereas CER^e is a function of the particular sample information available, y , and the prior pdf, $p(\theta)$. If $CER^n(l^*)$ and $CER^e(l^B)$ are to be compared, the most reasonable way of doing so is in an *ex ante* fashion, i.e.,

$$(3.5) \quad CER^n(l^*) \text{ versus } \{E_y[CER^e(l^B)] = E_y[l^B, \hat{\mu} - \frac{r(1+1/n)(n-1)}{2(n-k-2)} l^B, \hat{\Sigma} l^B]\}.$$

The right-hand side of (3.5) is the value that would be obtained if infinite y -type data samples were taken, $CER^e(l^B)$ for each sample were calculated, and then these $CER^e(l^B)$ s were averaged.

Criterion (3.5) differs from that employed by CCS, who propose using

$$(3.6) \quad CER^n(l^*) \text{ versus } \{E_y[CER^n(l^B)] = E_y(l^B, \mu - \frac{1}{2} r l^B, \Sigma l^B)\}$$

as the *ex ante* comparison of CERs. In the right-hand side of (3.6), the optimal decision under estimation risk (l^B) substitutes the optimal decision in the absence of estimation risk (l^*) in CER^n . But the certainty equivalent return for an individual facing estimation risk is CER^e , not CER^n . The measure CER^e solely depends on available (sample and prior) information and takes into account the additional uncertainty caused by the lack of perfect information about parameters. CER^n is the correct measure in the absence of estimation risk because it depends on the true parameter values and does not incorporate the additional uncertainty due to estimation risk. Thus, the decision rule l^B is best paired with CER^e , and the decision rule l^* is best paired with CER^n .

Bayes allocation is obtained by maximizing the predictive expected utility function.

Therefore, it is the allocation that yields the maximum certainty equivalent return in the presence of estimation risk. In particular, it must be the case that $CER^e(l^B)$ is greater than or equal to $CER^e(\hat{l})$ and $CER^e(\tilde{l})$. Indeed, it can be shown that⁷

$$(3.7) \quad CER^e(l^B) - CER^e(\tilde{l}) = \frac{(2n - k - 1)^2}{2rn(n^2 - 1)(n - k - 2)} \hat{\mu}' \hat{D} \hat{\mu} \geq 0,$$

$$(3.8) \quad CER^e(\tilde{l}) - CER^e(\hat{l}) = \frac{k(4n + nk - k - 2)}{2rn(n - 1)(n - k - 2)} \hat{\mu}' \hat{D} \hat{\mu} \geq 0.$$

Inequalities (3.7) and (3.8) hold because the ratios in the middle terms are strictly positive and the quadratic term $\hat{\mu}' \hat{D} \hat{\mu}$ is nonnegative.⁸ The interpretation of the results in (3.7) and (3.8) is that the decision maker will be indifferent among l^B , \hat{l} , and \tilde{l} if and only if the extremely unlikely event $\hat{\mu}' \hat{D} \hat{\mu} = 0$ occurs. Otherwise, the agent will strictly prefer l^B over \tilde{l} , and \tilde{l} over \hat{l} .

In Table 2 we illustrate the CER^e 's corresponding to the simulations reported in Table 1.⁹ It can be observed that Bayes allocation yields the greatest and the plug-in the smallest CER^e . As expected, $CER^e(l^B)$ increases with total acreage and decreases with the degree of absolute risk aversion. Interestingly, both $CER^e(\tilde{l})$ and $CER^e(\hat{l})$ increase with the degree of absolute risk aversion. This result can occur only because both decision vectors are suboptimal.

In line with the differences in allocations observed in Table 1, $CER^e(\tilde{l})$ is more similar to $CER^e(l^B)$ than to $CER^e(\hat{l})$. The difference between $CER^e(\tilde{l})$ and $CER^e(l^B)$ is smaller in absolute

⁷It can also be shown that

$$(3.9) \quad CER^e(l^B) - CER^e(\tilde{l}) = \frac{(n - k - 2)}{2rn(n^2 - 1)} \hat{\mu}' \hat{D} \hat{\mu} \geq 0,$$

$$(3.10) \quad CER^e(\tilde{l}) - CER^e(\hat{l}) = \frac{(1.5n - k - 1.5)}{rn(n - 1)(n - k - 2)} \hat{\mu}' \hat{D} \hat{\mu} \geq 0.$$

⁸ $\hat{\mu}' \hat{D} \hat{\mu}$ is nonnegative because the matrix \hat{D} is positive semidefinite.

⁹Note that $CER^e(\hat{l})$ and $CER^e(\tilde{l})$ under the nonnegativity restriction had been even smaller in the examples with corner solutions.

terms when the coefficient of absolute risk aversion is greater, and smallest in relative terms in the scenario with high risk aversion and large total acreage. But even in the latter scenario, $CER^e(\tilde{l})$ is 10 percent smaller than $CER^e(l^B)$, indicating that the differences in CER^e s are not negligible. Consistent with the concluding remarks of the previous section, the potential differences in CER^e s among the alternative allocation approaches are large enough to advocate using Bayes' criterion, particularly given the simplicity of its calculation relative to the plug-in or CCS methods.

The Value of Additional Sample Information in the Presence of Estimation Risk

Sample information in general (and additional sample information in particular) has no value to the decision maker in the absence of estimation risk because the agent already knows whatever information sample data can provide him. However, most common situations are characterized by the presence of estimation risk.

The value of additional sample information is an issue of potential importance in the presence of estimation risk. The analysis of this topic is closely related to that of calculating the optimal sample size when buying additional sample information. The solutions to both problems are conceptually (if not operationally) simple; they are obtained by employing backward induction in a Bayesian framework. Because both solutions apply similar concepts, for pedagogical reasons we will solve the second problem first and then extend the analysis to find the value of additional sample information of a particular size.

Consider a decision maker who has the possibility of acquiring additional (still unobserved) sample information v_{n_v} of size n_v at the cost $C(n_v)$ before selecting l . The decision maker's problem is to determine the optimal size of the additional information (n_v^B), based on the already observed sample y .¹⁰ Let $z_{n_v} = y + v_{n_v}$, i.e., z_{n_v} is the total sample information after having

¹⁰For simplicity, we will assume that there is a single-shot opportunity to buy additional sample information. The solution for the case in which the agent may buy additional information after having bought the initially optimal n_v^B but before selecting l can be obtained by applying the same principles, but it is too cumbersome and provides little additional insight.

acquired the additional sample information v_{n_v} . If the decision maker had the sample information z_{n_v} available at the time of making the decision about l , we know from (2.12) and (1.6) that the optimal Bayes decision vector would be given by expression (4.1):¹¹

$$(4.1) \quad \hat{l}^B(z_{n_v}) \equiv \operatorname{argmax}_{l \in \Lambda} E_{x|z_{n_v}} \{U[l'x - C(n_v)]\},$$

where the decision vector $\hat{l}^B(z_{n_v})$ is identified by z_{n_v} to stress that the decision depends on the total sample information (z_{n_v}). But the additional sample information (v_{n_v}) has not been observed by the agent at the time he must select the optimal size n_v^B ; therefore, $\hat{l}^B(z_{n_v})$ is unknown when choosing n_v . Hence, the problem reduces to maximizing the predictive expected utility from having the additional sample information v_{n_v} with respect to the size of the additional sample n_v , i.e.,

$$(4.2) \quad n_v^B = \operatorname{argmax}_{n_v \geq 0} E_{x, v_{n_v}|y} \{U[\hat{l}^B(y + v_{n_v})'x - C(n_v)]\}.$$

The solution to (4.2) will generally be obtained by means of numerical methods. The optimal size of the additional sample (n_v^B) will depend on the particular sample information available at the decision time (y), the agent's prior $[p(\theta)]$, and the cost of the additional sample information $[C(n_v)]$.

The concepts applied to obtain the optimal size of additional sample information can be employed to calculate how much the decision maker is willing to pay for an additional sample w_{n_w} of size n_w , which we will denote by W^B . W^B is the root of the equality

$$(4.3) \quad E_{x|y} \{U[\hat{l}^B(y)'x]\} = E_{x, w_{n_w}|y} \{U[\hat{l}^B(y + w_{n_w})'x - W]\},$$

¹¹In the land allocation example considered in the previous sections, $\hat{l}^B(z_{n_v})$ is independent of $C(n_v)$ because the utility function is negative exponential. In general, however, Bayes' rule will depend on $C(n_v)$.

where: $l^B(y + w_{n_w}) \equiv \operatorname{argmax}_{l \in \Lambda} E_{x|y+w_{n_w}} [U(l'x - W)]$.

The term in the left-hand side is the predictive expected utility in the *status quo*, i.e., with no additional sample information. The term in the right-hand side represents the predictive expected utility from acquiring the additional sample information w_{n_w} of size n_w at the price W . In particular, it should be noted that the right-hand side is the predictive expected utility *before* observing the additional sample because, in general, we will have

$$(4.4) \quad E_{x, w_{n_w} | y} \{U[l^B(y + w_{n_w})'x - W]\} \neq E_{x|y+w_{n_w}} \{U[l^B(y + w_{n_w})'x - W]\}.$$

If the price of the additional sample information is low ($W < W^B$), the agent will prefer to buy the additional sample. Conversely, when the price of the additional sample information is high ($W > W^B$), the agent will prefer not to buy the additional sample. Hence, the agent will be indifferent between the two alternatives only when $W = W^B$.

In Table 3 we report an example of the willingness to pay for an additional sample of size n_w . For simplicity, it is assumed that there is a single asset ($k = 1$) that can be held with no restrictions (i.e., the decision maker can be short or long the desired amount). The returns on the asset are normally distributed, and the decision maker has observed that the sample mean $\hat{\mu}$ equals $3 \cdot 10^4$ and the sample variance $\hat{\sigma}^2$ equals $2.25 \cdot 10^8$ with a sample of size n_y . It is further assumed that the coefficient of absolute risk aversion r equals $1/10000$, which reflects moderate risk aversion. The values reported in Table 3 were obtained by means of Monte Carlo simulations. It can be seen that the value of an additional sample decreases as the size of the observed sample (n_y) increases. This result is to be expected because the decision maker is more confident about the sample information when the sample available is large than when it is small. It can also be seen that, for a large observed sample ($n_y = 100$), the value of an additional small sample is negligible. Table 3 also reveals that the value of an additional sample increases with its size (n_w); however, the marginal value of each additional observation decreases very rapidly with n_w .

It is interesting to note from Table 3 that CER^e equals \$15,105 after observing a sample of size 12 with mean $3 \cdot 10^4$ and variance $2.25 \cdot 10^8$, whereas CER^e equals \$16,718 after observing a sample of size 18 with the same mean and variance. This increase of \$1,613 in CER^e by going from a sample of size 12 to a sample of size 18 grossly overestimates the willingness to pay for 6 additional observations by an agent who has observed a size-12 sample (compare \$1,613 with about \$270). The explanation for this result is that the agent who has only observed a sample of size 12 does not know what the outcome of the additional 6 observations will be.

Summary and Conclusions

In the almost universal situation where parameters of importance for decision making are not known with certainty, decisions will be subject to an additional source of risk related to the accuracy with which parameters are estimated. Bayes' criterion is the procedure consistent with expected utility maximization in the presence of estimation risk. To show the importance of estimation risk, this paper reexamines the land allocation problem in the presence of estimation risk. A simple allocation rule based on the sample estimates of the mean vector and the covariance matrix of crop returns is obtained. This land allocation rule is derived in a manner consistent with expected utility maximization, and is therefore preferable to other *ad hoc* criteria for decision making in the presence of estimation risk. The allocation rule advocated in this study yields greater expected utility than does an allocation that is an unbiased estimator of the optimal land allocation in the absence of estimation risk.

The paper also discusses how to calculate the certainty equivalent return in the presence of estimation risk (CER^e). It is argued that CER^e is not directly comparable to the certainty equivalent return in the absence of estimation risk (CER^n), and a mean to compare CER^n with CER^e *ex ante* is proposed.

The final section of the study addresses two related issues that can only be analyzed in a framework that explicitly considers estimation risk. These issues are the optimal size of an

additional (still unobserved) data sample, and the willingness to pay for an additional (still unobserved) sample of a particular size. Applying the tools developed in previous sections, conceptual solutions to both problems are presented. This section also shows how one can put dollar value on additional sample information. This issue is of relevance for experimental design and the selection of survey size.

References

- Alexander, Gordon J., and Bruce G. Resnick. "More on Estimation Risk and Simple Rules for Optimal Portfolio Selection." *J. Finan.* 40(1985):125-33.
- Anderson, T. W. *An Introduction to Multivariate Statistical Analysis*, 2nd. edition. New York: John Wiley and Sons, 1984.
- Aitchison, J., and I. R. Dunsmore. *Statistical Prediction Analysis*, Chapter 2. New York: Cambridge University Press, 1975.
- Bawa, V., S. Brown, and R. Klein. "Estimation Risk: An Introduction." *Estimation Risk and Optimal Portfolio Choice*, eds. Bawa, V., S. Brown, and R. Klein, Chapter 1. Amsterdam: North-Holland Publishing Co., 1979.
- Berger, James O. *Statistical Decision Theory and Bayesian Analysis*, 2nd edition. New York: Springer-Verlag, 1985.
- Boyle, Phelim P., and A. L. Ananthanarayanan. "The Impact of Variance Estimation in Option Valuation Models." *J. Finan. Econ.* 5(1977):375-87.
- Chalfant, J. A., R. N. Collender, and S. Subramanian. "The Mean and Variance of the Mean-Variance Decision Rule." *Amer. J. Agr. Econ.* 72(1990):966-74.
- Chen, Son-Nan, and Stephen J. Brown. "Estimation Risk and Simple Rules for Optimal Portfolio Selection" *J. Finan.* 38(1983):1087-93.
- Coles, Jeffrey L., and Uri Loewenstein. "Equilibrium Pricing and Portfolio Composition in the Presence of Uncertain Parameters." *J. Finan. Econ.* 22(1988):279-303.
- Collender, Robert N. "Estimation Risk in Farm Planning under Uncertainty." *Amer. J. Agr. Econ.* 71(1989):996-1002.
- DeGroot, Morris H. *Optimal Statistical Decisions*, Chapters 7-10. New York: McGraw-Hill, 1970.
- Freund, Rudolf J. "The Introduction of Risk into a Programming Model." *Econometrica* 24(1956):253-63.

- Frost, Peter A., and James E. Savarino. "An Empirical Bayes Approach to Efficient Portfolio Selection." *J. Finan. Quant. Analysis* 21(1986):293-305.
- Hey, John. *Uncertainty in Microeconomics*. New York: New York University Press, 1979.
- Jorion, Philippe. "International Portfolio Diversification with Estimation Risk." *J. Business* 58(1985):259-78.
- Klein, R. W., L. C. Rafsky, D. S. Sibley, and R. D. Willig. "Decisions with Estimation Uncertainty." *Econometrica* 46(1978):1363-87.
- Lence, Sergio H., and Dermot J. Hayes. "Parameter-Based Decision Making under Estimation Risk: An Application to Futures Trading." *J. Finan.* (forthcoming).
- Raiffa, Howard, and Robert Schlaifer. *Applied Statistical Decision Theory*. Boston: Harvard University Press, 1961.
- Sandmo, Agnar. "On the Theory of the Competitive Firm under Price Uncertainty." *Amer. Econ. Rev.* 61(1971):65-73.
- Shimizu, Kunio. "Point Estimation." *Lognormal Distributions--Theory and Applications*, eds. Crow, E., and K. Shimizu, Chapter 2. New York: Marcel Dekker, 1988.

Table 1. Land allocations for seven observations ($n = 7$) obtained by means of Bayes' criterion, CCS's method, and plug-in approach.

Total Acreage	Coeff. of Absolute Risk Aversion	Potatoes (acres)			Corn (acres)			Beef (acres)			Fall Cabbage (acres)		
		Bayes	CCS	Plug-in	Bayes	CCS	Plug-in	Bayes	CCS	Plug-in	Bayes	CCS	Plug-in
150	1/8500	7.5	26.4	93.4	80.5	59.2	-16.2	53.3	36.2	-24.6	8.7	28.2	97.4
	1/40000	61.8	150.4	465.5	19.4	-80.5	-435.4	4.0	-76.4	-362.5	64.8	156.5	482.4
300	1/8500	0.5	19.3	86.2	177.4	156.2	80.8	119.9	102.8	42.0	2.2	21.7	91.0
	1/40000	54.7	143.3	458.4	116.3	16.5	-338.4	70.6	-9.8	-295.9	58.4	150.0	475.9

Table 2. Certainty equivalent returns for seven observations ($n = 7$) obtained by means of Bayes' criterion, CCS's method, and the plug-in approach.

Total Acreage	Coeff. of Absolute Risk Aversion	Certainty Equivalent Return (\$)		
		Bayes	CCS	Plug-in
150	1/8500	10325	8490	-27771
	1/40000	14933	6294	-164346
300	1/8500	18291	16456	-19805
	1/40000	24374	15735	-154904

Table 3. Value of Additional Samples

Size of Observed Sample (n_y)	Certainty Equivalent Return (\$)	Dollar Value of Additional Sample (W^B) of Size		
		$n_w = 1$	$n_w = 5$	$n_w = 10$
12	15,105	114.8	269.7	273.0
18	16,718	41.3	124.0	150.2
30	18,020	11.7	43.7	64.2
100	19,402	0.8	3.9	7.5