*Australian Journal of Agricultural Economics*, Vol. 31, No. 2 (August 1987), pp.127-141

# SUBJECTIVE DISTRIBUTIONS AS ECONOMETRIC RESPONSE DATA

W. E. GRIFFITHS, J. R. ANDERSON and K. B. HAMAL
*University of New England, Armidale, NSW 2351*

**Analysis of actual observations of response processes is a routine procedure in applied econometrics but methods of dealing with purely subjective probability distributions as response 'data' have seemingly not hitherto been worked out. The models and methods developed here go some way to filling this void. They are illustrated by using information from a cross-sectional study of sixty Nepalese small-scale farmers. Means and variances of subjective probability distributions for rice paddy yields under different technologies are related to controlled inputs, such as fertiliser levels, and to relevant socio-economic aspects of farmers themselves, such as technological knowledge and farm size.**

A significant recent development in the field of agricultural production economics has been the recognition that risk in response phenomena, in conjunction with risk aversion on the part of decision makers, is a phenomenon worthy of empirical measurement. Estimation methods for more complete response models which explicitly introduce risk have been proposed by Just and Pope (1978), and have been extended to the case of a cross section of time series by Griffiths and Anderson (1982) and to enriched stochastic structures by Rosegrant and Roumasset (1985). Methods of dealing with conventional observational data are now well established.

However, from the point of view of modern decision theory, the information that is most relevant to decisions by individuals in the face of risk is the subjective set which encapsulates their beliefs about uncertain states of nature. The techniques of elicitation of subjective probabilities are well developed (Raiffa 1968; Savage 1971) and have found extensive application in the field of agricultural decision analysis (Anderson, Dillon and Hardaker 1977). Such subjective probability data have not been incorporated in more formal econometric models of response processes in crop and livestock production. The purpose in this paper is to explore some apposite methods for translating purely subjective first and second moments of probability distributions into corresponding observable and unobservable components that can be used in explanatory regressions. These methods are illustrated using cross-sectional survey information from Nepal.

## *Data*

Hamal (1982) interviewed sixty Nepalese smallholding farmers to obtain data on their personal and farming circumstances, as well as their subjective beliefs about paddy yields. Their major farm enterprise is the growing of rice (paddy), mainly for their own families' subsistence consumption. The primary purpose of the study was to examine the impact of risk and risk aversion on the adoption of special-purpose lines of credit. The credit was intended to assist such small farmers to adopt technologies believed to be superior, at least in the sense of inducing higher average yields of this staple grain. Fortunately, these data also lend themselves to treatment in the proposed econometric estimation.

The farmers come from two adjacent villages in south central Nepal. Being small-scale farmers, they are not wealthy, have only their small plots (average of 0·1 ha of rice per farm) and generally suffer from a very poor educational background. To facilitate the elicitation of subjective probability information, most emphasis was placed on using triangular distributions in the first round of questioning of subjective yield distributions. The adequacy of the triangular assumption was checked by asking further questions using the 'judgmental fractile' method (Raiffa 1968) of elicitation.[1] In general, the triangular distributions seemed to capture adequately the nature of the yield distributions.

### *The Models*

Following Just and Pope (1978) and Griffiths and Anderson (1982), the following response function is postulated:

$$Y = g(X) + \varepsilon h(X) \tag{1}$$

where, in this case, $Y$ is rice yield (t/ha), $X$ is a vector of inputs which describes, among other things, one of four possible technologies and $\varepsilon$ is a disturbance term with zero mean and variance $\sigma^2$. In equation (1), mean yield, $E(Y) = g(X)$, and the standard deviation of yield, $S(Y) = \sigma h(X)$, are both functions of $X$ and these functions are not necessarily the same. It is primarily this characteristic which led both Just and Pope (1978) and Griffiths and Anderson (1982) to adopt a function of this type.

Letting the subscripts $i = 1, 2, 3, 4$ and $f = 1, 2, \ldots, 60$ denote the $i$-th technology for the $f$-th farm, the specific functions $g$ and $h$ which were assumed are

$$g(X) = E(Y_{if}) = \alpha_0 + \alpha_1 D_i + \alpha_2 N_i + \alpha_3 N_i^2 + \alpha_4 B_{if} + \alpha_5 B_{if}^2 + \alpha_6 N_i B_{if} + \mu_f \tag{2a}$$

and

$$\sigma h(X) = S(Y_{if}) = \beta_0 + \beta_1 D_i + \beta_2 N_i + \beta_3 N_i^2 + \beta_4 B_{if} + \beta_5 B_{if}^2 + \beta_6 N_i B_{if} + \lambda_f \tag{2b}$$

[1] The triangular distribution, which takes its name from the shape of its probability density function, is a three-parameter distribution. It is simply elicited by asking the subject directly for the three parameters, namely, the lowest possible and highest possible value and the mode (Anderson, Dillon and Hardaker 1977, p. 26). Graphed as a cumulative distribution function, it consists of two spliced quadratic segments (Anderson, Dillon and Hardaker 1977, p. 268). The judgmental fractile method of elicitation involves determining several fractiles (beginning with the median and other quartiles) of the subject's personal probability distribution, and graphing these fractiles directly as a cumulative distribution function (Anderson, Dillon and Hardaker 1977, p. 24). The adequacy checks mentioned in the text consisted of graphing this directly elicited cumulative distribution function together with that for the correspondingly elicited triangular distribution, and visually assessing how closely they matched.

where $N_i$ = level of nitrogen applied (10 kg/ha units);
$D_i$ = dummy variable which is zero when $N_i = 0$ and unity otherwise; and
$B_{if}$ = value of labour and bullocks assessed to be used in production (Rs '000 per ha).

The alphas and betas are unknown coefficients and $\mu_f$ and $\lambda_f$ are random components which reflect specific characteristics of the $f$-th farm.

The four technologies are characterised by the level of nitrogen and the variety of rice. When $i = 1$ the nitrogen level is zero and a traditional variety of rice unresponsive to fertiliser is employed. The technologies $i = 2$, 3 and 4 correspond to nitrogen applications of 22·5, 44 and 67 kg/ha, respectively, and involve a modern, more responsive variety. The dummy variable is included to capture the effect of changing the variety; and the variable $B_{if}$ is Hamal's (1982) survey-based estimate of the value of labour and bullocks needed for a particular technology. In equations (2a, b), the mean and standard deviation are conditional on the random components $\mu_f$ and $\lambda_f$. For the population from which the sixty farms are a sample, $\mu_f$ and $\lambda_f$ are unconditionally assumed to be independent random variables with zero means and respective variances $\sigma_\mu^2$ and $\sigma_\lambda^2$. Finally, a conventional second-order 'quadratic' function in $N_i$ and $B_{if}$ is chosen to capture a possible variety of marginal effects.

In the mean function $g(X)$ it is expected that marginal expected response with respect to $N$ and $B$ will be positive, but decreasing. That is, $\alpha_2, \alpha_4 > 0$ and $\alpha_3, \alpha_5 < 0$. The coefficient $\alpha_6$ captures the effect of one input on the marginal expected response of the other; its sign could be positive or negative. The question of appropriate coefficient signs in the standard deviation function $\sigma h(X)$ is largely an empirical one. There has been evidence to suggest that marginal risk with respect to nitrogen application is positive (Roumasset, Rosegrant, Chakravarty and Anderson, in press), a property which would be reflected by $\beta_2 > 0$. Given $\beta_2 > 0$, the sign of $\beta_3$ indicates whether risk is increasing at an increasing or decreasing rate, as the level of nitrogen application increases. It would normally be expected that labour and bullocks would be risk reducing ($\beta_4 < 0$); the second-order effects represented by $\beta_5$ and $\beta_6$ are an empirical question.

Unlike in previous studies, data on yield are not available; instead, the available data are on the mean and variance of yield, derived from the subjective probability distributions that were elicited from the farmers. Given this situation, assumptions are required to derive equations from which to estimate the alphas and betas. As a first step in this direction, it seems reasonable to assume that, corresponding to equations (2a, b), the farmers possess subjective moment functions given by:

(3a)
$$E^*(Y_{if}) = \alpha^*_{0if} + \alpha^*_{1if} D_i + \alpha^*_{2if} N_i + \alpha^*_{3if} N_i^2 + \alpha^*_{4if} B_{if} + \alpha^*_{5if} B_{if}^2 + \alpha^*_{6if} N_i B_{if} + \mu^*_f$$

and

(3b) $$S^*(Y_{if}) = \beta^*_{0if} + \beta^*_{1if}D_i + \beta^*_{2if}N_i + \beta^*_{3if}N_i^2 + \beta^*_{4if}B_{if} + \beta^*_{5if}B_{if}^2 + \beta^*_{6if}N_iB_{if} + \lambda^*_f$$

The quantities $E^*(Y_{if})$ and $S^*(Y_{if})$ represent the subjective means and standard deviations for which data are available; the 'starred' coefficients can be viewed as farmers' 'perceived' coefficients. Equations (3a, b) are related to equations (2a, b) by making assumptions about the perceived coefficients.

The most general set of assumptions which is made in this regard is that the perceived intercepts $\alpha^*_{0if}$ and $\beta^*_{0if}$, and the perceived coefficients of nitrogen and nitrogen squared ($\alpha^*_{2if}$, $\alpha^*_{3if}$, $\beta^*_{2if}$, $\beta^*_{3if}$) can be explained by the socio-economic aspects of the farmers themselves. That is,

(4a) $$\alpha^*_{jif} = \delta_{0j} + \sum_{k=1}^{3} \delta_{kj} Z_{kf} + e_{jif} \qquad j = 0, 2, 3$$

and

(4b) $$\beta^*_{jif} = \omega_{0j} + \sum_{k=1}^{3} \omega_{kj} Z_{kf} + v_{jif} \qquad j = 0, 2, 3$$

where $Z_{1f}$ = area (ha) under rice for the $f$-th farm;
$Z_{2f}$ = years of experience of the $f$-th farmer;
$Z_{3f}$ = level of education measured by number of years of schooling; and
$e_{if} = (e_{0if}, e_{2if}, e_{3if})'$ and $v_{if} = (v_{0if}, v_{2if}, v_{3if})'$ represent independent random error vectors each of which is a random drawing from its respective trivariate distribution.

It is assumed that the remaining coefficients, and the farm-specific components, are correctly perceived by the farmers; that is, $\mu^*_f = \mu_f$, $\lambda^*_f = \lambda_f$,

(5a) $$\alpha^*_{jif} = \alpha_j \qquad j = 1, 4, 5, 6$$

and

(5b) $$\beta^*_{jif} = \beta_j \qquad j = 1, 4, 5, 6$$

Obviously, a more general model would also relate these coefficients to the socio-economic variables. However, any attempts in this direction are likely to lead to estimating equations which are unmanageable and where the chances of obtaining reliable estimates are slim. Since nitrogen is the most important controllable input, the choice made in (4a, b) seems a reasonable one.

Given that there are various ways in which the elements in $e_{if}$ and $v_{if}$ could be related, the stochastic assumptions about these vectors need precise specification. It is assumed that $E(e_{if}e'_{ig}) = 0$ and $E(v_{if}v'_{ig}) = 0$ for $f \neq g$; that is, the error vectors are uncorrelated across farms, a reasonable assumption for survey data. It is also assumed that $E(e_{if}e'_{hf}) = 0$ and $E(v_{if}v'_{hf}) = 0$ for $i \neq h$; the error vectors for different levels of nitrogen on a given farm are uncorrelated. This assumption is a more difficult one to justify, but the firm-specific random components $\mu_f$ and $\lambda_f$, will capture characteristics peculiar to a specific farm, and hence will contribute to a composite error term which exhibits within-farm correlation. The assumption $E(e_{it}v'_{it})=0$ implies that error terms from mean perceptions are uncorrelated with error terms from perceptions about standard deviations. Relaxation of this assumption does not invalidate the estimation procedure pursued later in the paper. It does, however, imply that it might be possible to follow a more asymptotically efficient estimation procedure which jointly estimates the mean and standard deviation functions in the spirit of Zellner's (1962) seemingly unrelated regression technique. This is a possible direction for future research. Finally, it is not assumed that the elements in $e_{if}$ (and $v_{if}$) are uncorrelated; that is, the covariance matrixes of $e_{if}$ and $v_{if}$ are not necessarily diagonal.

Substituting (4a, b) and (5a, b) into (3a, b) yields the following equations for estimation:

$$\text{(6a)} \quad E^*(Y_{if}) = \delta_{00} + \alpha_1 D_i + \delta_{02} N_i + \delta_{03} N_i^2 + \alpha_4 B_{if} + \alpha_5 B_{if}^2 + \alpha_6 N_i B_{if} + \sum_{k=1}^{3} \delta_{k0} Z_{kf} + \sum_{k=1}^{3} \delta_{k2} N_i Z_{kf} + \sum_{k=1}^{3} \delta_{k3} N_i^2 Z_{kf} + \mu_f + e_{0if} + N_i e_{2if} + N_i^2 e_{3if}$$

and

$$\text{(6b)} \quad S^*(Y_{if}) = \omega_{00} + \beta_1 D_i + \omega_{02} N_i + \omega_{03} N_i^2 + \beta_4 B_{if} + \beta_5 B_{if}^2 + \beta_6 N_i B_{if} + \sum_{k=1}^{3} \omega_{k0} Z_{kf} + \sum_{k=1}^{3} \omega_{k2} N_i Z_{kf} + \sum_{k=1}^{3} \omega_{k3} N_i^2 Z_{kf} + \lambda_f + v_{0if} + N_i v_{2if} + N_i^2 v_{3if}$$

The paper is concerned with estimation of these equations, and two restricted versions of them. In the first restricted version, all coefficients other than the intercept terms are assumed to be correctly perceived by farmers. Also, it is assumed that the socio-economic variables have no bearing on farmers' perceptions. Thus, $\alpha^*_{0if} = \alpha_0 + e_{0if}$ and $\beta^*_{0if} = \beta_0 + v_{0if}$. The restrictions on (6a, b) implied by these assumptions are:

$$\text{(7a)} \quad \delta_{kj} = 0 \qquad k = 1, 2, 3 \qquad j = 0, 2, 3$$
$$e_{2if} \equiv 0 \qquad e_{3if} \equiv 0$$

and

$$\text{(7b)} \quad \omega_{kj} = 0 \qquad k = 1, 2, 3 \qquad j = 0, 2, 3$$
$$v_{2if} \equiv 0 \qquad v_{3if} \equiv 0$$

In the second restricted version, all coefficients other than the intercept terms are again assumed to be correctly perceived by farmers. However, in this case the perceived intercepts are related to the socio-economic variables. The restrictions are identical to those in (7a, b) except that the subscript $j$ now only takes the values 2 and 3. That is, ($\delta_{10}$, $\delta_{20}$, $\delta_{30}$, $\omega_{10}$, $\omega_{20}$, $\omega_{30}$) are retained in the model.

Estimation of (6a, b) and their restricted versions gives an indication of the possible relationships between the mean and variance of subjective yield distributions, various inputs, and socio-economic characteristics of the farmers. These relationships are in contrast to those considered by Just and Pope (1978) and Griffiths and Anderson (1982) who were concerned with the relationships between the *actual* mean and variance of output, and input levels. Because of the assumptions made in this paper about the way in which perceived and actual response coefficients are related, it is possible to derive some estimates of the actual coefficients; such derivations are not the major concern of the paper, however.

## *Estimation*

Feasible generalised least squares was used to estimate the parameters in (6a, b) and the restricted versions of these equations. The precise form of the generalised least squares estimator is governed by the nature of the error terms which are given by:

(8a) $$\mu_f + e^*_{if} = \mu_f + e_{0if} + N_i e_{2if} + N_i^2 e_{3if}$$

and

(8b) $$\lambda_f + v^*_{if} = \lambda_f + v_{0if} + N_i v_{2if} + N_i^2 v_{3if}$$

Each error term is characterised by a farm-specific component, $\mu_f$ or $\lambda_f$, and independent random errors $e^*_{if}$ and $v^*_{if}$. In the restricted versions where $e_{2if}$, $e_{3if}$, $v_{2if}$ and $v_{3if}$ are identically zero, $e^*_{if} = e_{0if}$ and $v^*_{if} = v_{0if}$ are homoscedastic and the models reduce to the conventional error components model (see, for example, Judge, Griffiths, Hill, Lutkepohl and Lee 1985, pp. 521–5). With the unrestricted equations, $e^*_{if}$ and $v^*_{if}$ are heteroscedastic with their variances depending on the level of nitrogen.

The methods used to estimate the variance components $\sigma^2_\mu$, $\sigma^2_\lambda$, $\sigma^2_e = E(e^{*2}_{if})$ and $\sigma^2_v = E(v^{*2}_{if})$ for the restricted homoscedastic equations, and the data transformation which yields the feasible generalised least squares estimator, are outlined in Judge et al. (1985, pp. 521–5). Models with dummy variables for farms, and the farm-constant socio-economic variables omitted, are used to estimate $\sigma^2_e$ and $\sigma^2_v$. The variance components $\sigma^2_\mu$ and $\sigma^2_\lambda$ are estimated from the models after averaging within farms and omitting variables $N_i$ and $D_i$ whose within-farm averages do not vary from farm to farm.

Estimation of the unrestricted equations given in (6a, b) is less conventional because of the presence of both a farm-specific error component and a heteroscedastic error component. A suitable generalised least squares procedure will be described in terms of the mean function (6a). Let $C$ be the covariance matrix for $e_{if} = (e_{0if}, e_{2if}, e_{3if})'$, then, under the assumptions outlined earlier, $e^*_{if}$ is an independent random variable with mean zero and variance:

(9) $$\sigma_i^2 = c_{11} + 2c_{12}N_i(2c_{13} + c_{22})N_i^2 + 2c_{23}N_i^3 + c_{33}N_i^4$$

where the $c_{ij}$ are elements of $C$. If the observations are ordered first according to $i$ with $f$ fixed, and then according to $f$, then the covariance matrix for the vector of disturbance terms $(\mu_f + e^*_{if})$ is given by $I \otimes V$ where $V = \sigma_\mu^2 J + \Sigma$, $J$ is a $(4 \times 4)$ matrix of ones, $\Sigma$ is a diagonal matrix with non-zero elements $(\sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2)$, $I$ is the identity matrix of dimension $(60 \times 60)$, and $\otimes$ denotes Kronecker product. A feasible generalised least squares estimator of the coefficients in (6a) is:

(10) $$\hat{\gamma} = [X'(I \otimes \hat{V}^{-1})X]^{-1}X'(I \otimes \hat{V}^{-1})y$$

where $\hat{V}$ is an estimator for $V$, $\gamma$ is a $(16 \times 1)$ vector containing the unknown coefficients in (6a), $X$ is a $(240 \times 16)$ matrix of observations on the explanatory variables in (6a), and $y$ is a $(240 \times 1)$ vector of observations on $E^*(Y_{if})$.

The remaining problem is to suggest estimators for $\sigma_\mu^2$ and $\sigma_i^2$, $i = 1, 2, 3, 4$. The first step for the $\sigma_i^2$ is to estimate (6a) by least squares with the $\mu_f$ treated as coefficients of dummy variables, and with the $Z_{kf}$ (but not the interaction terms $N_i Z_{kf}$ and $N_i^2 Z_{kf}$) omitted because of their collinearity with the dummy variables. Let the least squares residuals from this regression be given by $\hat{e}^*_{if}$. Following the specification in (9), a test for heteroscedasticity is given by regressing $(\hat{e}^{*2}_{if}/\tilde{\sigma}^2)$ on a quartic equation in $N_i$ where $\tilde{\sigma}^2 = \sum_{i=1}^{4} \sum_{f=1}^{60} \hat{e}^{*2}_{if}/240$. The test is completed by comparing half the regression sum of squares from this regression with a critical value of the $\chi^2_{4\,\text{d.f.}}$ distribution (Breusch and Pagan 1979). Unfortunately, this procedure breaks down because the data contain only four different settings of $N_i$. As an approximating alternative, an analogous procedure with a cubic equation in $N_i$ (and the $\chi^2_{3\,\text{d.f.}}$ distribution) was followed. For both equations (6a, b), the null hypothesis of homoscedasticity was rejected. Assuming that it is satisfactory to omit $N_i^4$, estimates of the $\sigma_i^2$ are given by the predicted values from the regression of $\hat{e}^{*2}_{if}$ on a constant term, $N_i$, $N_i^2$ and $N_i^3$ (Goldfeld and Quandt 1972; Amemiya 1977).

The first step toward estimation of $\sigma_\mu^2$ is to find least squares estimates of (6a), averaged over $i$. The residuals from this procedure can be used to find an estimate $\hat{\sigma}_*^2$ of the variance of an averaged disturbance which is given by:

(11) $$\sigma_*^2 = \sigma_\mu^2 + \sum_{i=1}^{4} \sigma_i^2/4^2$$

Then, an estimator for $\sigma_\mu^2$ is given by $\hat{\sigma}_\mu^2 = \hat{\sigma}_*^2 - \sum_{i=1}^{4} \hat{\sigma}_i^2/16$.

The estimates $\hat{\sigma}_i^2$ and $\hat{\sigma}_\mu^2$ are used to form $\hat{V}$ so that the feasible generalised least squares estimator in (10) can be computed. Corresponding procedures are used to estimate (6b).

## *Results*

The coefficient estimates and $t$-ratios for the alternative models are given in Tables 1 and 2 for $E^*(Y)$ and $S^*(Y)$, respectively. In each case, the results are presented for the complete models as outlined (6a, b), prefaced with the restricted models (7a, b), and the slightly less restricted models which are described as (7′a, b). Also included are some corresponding revised models which were based on the significance and plausibility of the original coefficients. Tables 3 and 4 contain the adjusted coefficients of determination ($\bar{R}^2$) for each equation and, because this latter measure is based on the transformed variables and is therefore not necessarily a reliable basis for comparison, the $\bar{R}^2$ statistics from the corresponding dummy variable equations. Also included are values for the Akaike information criterion, some $F$ statistics, the variance estimates, and the value of the $\chi^2$ statistics for testing for heteroscedasticity in equations (6a, b). The critical value of the $\chi^2_{3\,\text{d.f.}}$ distribution at a 5 per cent significance level is 7·81.

Apart from some constants, the Akaike information criterion (AIC) version employed is the one outlined by Amemiya (1980), namely,

$$\text{(12)} \qquad \text{AIC} = -\frac{2}{T}\log L(\hat{\theta}) + \frac{2K}{T}$$

where $\hat{\theta}$ = the maximum likelihood estimate of all the unknown parameters; $L(\hat{\theta})$ = the maximum of the likelihood function; $K$ = the number of parameters (the dimension of $\hat{\theta}$); and $T$ = the number of observations. To obtain the entries in Tables 3 and 4 the errors were assumed to be normally distributed, the log-likelihood function was evaluated at the generalised least squares estimates (rather than the maximum likelihood estimates), and the constant $2/T$, and the constant term in the log-likelihood function, were ignored. Because model selection criteria such as the Akaike information criteria are rather *ad hoc* (Judge et al. 1985, pp. 862–81), these values should be regarded as a rough guide, rather than a rigorous basis for model selection. In general, if the Akaike information criterion were the sole criterion, the model for which it had the minimum value would be chosen.

The models (7′a, b) and (7a, b) differ from those in (6a, b) in two respects. In (6a, b), more variables are included and the errors are heteroscedastic. Thus, if the Akaike information criterion is considerably lower for equations (6a, b), as indeed is the case, this could be attributable to the additional variables, the heteroscedasticity assumption, or both. To shed more light on this question, both versions of (7′a) and (7a), and the revised version of (6a) were re-estimated using the heteroscedasticity assumption and the variance estimates obtained with the full version of (6a). This procedure yields, for all the equations, residual sums of squares which are comparably weighted. Under these circumstances, the conventional $F$ test, which uses the differences in residual sums of squares to test for the significance of omitted variables, will be asymptotically justified. In Table 3, the $F$ values obtained from testing for the omitted variables in each equation, relative to the full version of model (6a) are presented. The values given in Table 4 were obtained in a similar way, and were obtained relative to equation (6b).

## TABLE 1

*Equations for Mean Yield*

| | Const.[a] | Variety dummy | $N$ | $N^2$ | $B$ | $B\times N$ | $B^2$ | Area | Exper. | Educ. | Area $\times N$ | Exper. $\times N$ | Educ. $\times N$ | Area $\times N^2$ | Exper. $\times N^2$ | Educ. $\times N^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| First-round models | | | | | | | | | | | | | | | | |
| (7′a) | 0·575 | −0·226 | 0·440 | −0·0239 | −0·830 | −0·0881 | 0·580 | | | | | | | | | |
| | (4·19)[b] | (−3·00) | (8·48) | (−4·70) | (−1·13) | (−2·93) | (2·18) | | | | | | | | | |
| (7a) | 0·623 | −0·224 | 0·437 | −0·0237 | −0·810 | −0·0871 | 0·569 | −0·413 | −0·0450 | 0·0100 | | | | | | |
| | (4·36) | (−2·98) | (8·46) | (−4·68) | (−1·10) | (−2·90) | (2·14) | (−1·12) | (−1·69) | (1·37) | | | | | | |
| (6a) | 2·047 | −0·239 | 0·443 | −0·0302 | −0·452 | −0·0568 | 0·418 | −0·0782 | −0·0376 | 0·0089 | −0·167 | −0·0095 | −0·00006 | 0·0134 | 0·0014 | 0·00008 |
| | (4·58) | (−4·07) | (8·98) | (−5·74) | (−0·74) | (−2·08) | (1·90) | (−0·17) | (−1·16) | (0·99) | (−1·28) | (−1·01) | (−0·02) | (0·78) | (1·10) | (0·24) |
| Revised models | | | | | | | | | | | | | | | | |
| (7′a) | 0·306 | −0·235 | 0·367 | −0·0254 | 0·740 | −0·0319 | | | | | | | | | | |
| | (5·61) | (−3·10) | (9·08) | (−5·03) | (5·12) | (−2·04) | | | | | | | | | | |
| (7a) | 0·355 | −0·232 | 0·365 | −0·0251 | 0·728 | −0·0320 | | −0·346 | −0·0433 | | | | | | | |
| | (6·15) | (−3·06) | (9·05) | (−4·99) | (5·08) | (−2·05) | | (−0·94) | (−1·64) | | | | | | | |
| (6a) | 1·395 | −0·245 | 0·373 | −0·0315 | 0·569 | | | | | | −0·0718 | −0·0164 | | | 0·0021 | |
| | (7·27) | (−4·33) | (11·02) | (−6·68) | (4·23) | | | | | | (−1·99) | (−1·97) | | | (1·89) | |

[a] Because of the different methods of transformation, the interpretation of the constant is not the same for all models.
[b] Numbers in parentheses are $t$-values for the null hypothesis of zero coefficients.

## TABLE 2

*Equations for the Standard Deviation of Yield*

| | Const.[a] | Variety dummy | $N$ | $N^2$ | $B$ | $B \times N$ | $B^2$ | Area | Exper. | Educ. | Area $\times N$ | Exper. $\times N$ | Educ. $\times N$ | Area $\times N^2$ | Exper. $\times N^2$ | Educ. $\times N^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| First-round models | | | | | | | | | | | | | | | | |
| (7′b) | 0·152 | 0·331 | 0·0374 | 0·0014 | −0·256 | −0·0209 | 0·116 | | | | | | | | | |
| | (2·29)[b] | (10·04) | (1·66) | (0·65) | (−0·81) | (−1·60) | (1·01) | | | | | | | | | |
| (7b) | 0·195 | 0·334 | 0·0222 | 0·0017 | −0·0143 | −0·0117 | 0·019 | −0·284 | −0·0414 | −0·0010 | | | | | | |
| | (2·36) | (10·26) | (1·02) | (0·83) | (−0·05) | (−0·93) | (0·17) | (−2·52) | (−5·08) | (−0·46) | | | | | | |
| (6b) | 0·483 | 0·329 | 0·0587 | −0·0028 | −0·0790 | −0·0067 | 0·035 | −0·072 | −0·0220 | −0·0014 | −0·139 | −0·0125 | 0·0005 | 0·0131 | 0·0013 | −0·00007 |
| | (2·88) | (14·44) | (2·89) | (−1·40) | (−0·33) | (−0·58) | (0·40) | (−0·47) | (−1·99) | (−0·46) | (−2·43) | (−3·02) | (0·41) | (1·82) | (2·47) | (−0·45) |
| Revised models | | | | | | | | | | | | | | | | |
| (7′b) | 0·117 | 0·327 | 0·01124 | 0·0008 | | | | | | | | | | | | |
| | (31·01) | (10·80) | (0·75) | (0·48) | | | | | | | | | | | | |
| (7b) | 0·195 | 0·327 | 0·01124 | 0·0008 | | | | −0·284 | −0·0418 | | | | | | | |
| | (23·56) | (10·80) | (0·75) | (0·48) | | | | (−2·50) | (−5·13) | | | | | | | |
| (6b) | 0·442 | 0·327 | 0·0509 | −0·0032 | | | | −0·093 | −0·0224 | | −0·128 | −0·0124 | | 0·0122 | 0·0013 | |
| | (15·84) | (17·18) | (3·96) | (−2·06) | | | | (−0·61) | (−2·05) | | (−2·28) | (−3·07) | | (1·72) | (2·49) | |

[a] Because of the different methods of transformation, the interpretation of the constant is not the same for all models.
[b] Numbers in parentheses are $t$-values for the null hypothesis of zero coefficients.

TABLE 3

*Statistics for Mean Yield Equations*

| | $\bar{R}^2$ | AIC | $F$ | Variance estimates and related statistics | |
|---|---|---|---|---|---|
| First-round models | | | | | |
| (7′a) | 0·911 (0·939)[a] | −311 | 1·411 | $\hat{\sigma}_\mu^2=0\cdot0390$<br>$\hat{\sigma}_e^2=0\cdot0115$ | |
| (7a) | 0·910 (0·938) | −312 | 1·038 | $\hat{\sigma}_\mu^2=0\cdot0361$<br>$\hat{\sigma}_e^2=0\cdot0115$ | |
| (6a) | 0·977 (0·940)′ | −329 | | $\hat{\sigma}_\mu^2=0\cdot0370$<br>$\hat{\sigma}_1^2=0\cdot0151$<br>$\hat{\sigma}_2^2=0\cdot0051$ | $\hat{\sigma}_3^2=0\cdot0039$<br>$\hat{\sigma}_4^2=0\cdot0072$<br>$\chi^2=37\cdot4$ |
| Revised models | | | | | |
| (7′a) | 0·909 (0·938) | −310 | 1·644 | $\hat{\sigma}_\mu^2=0\cdot0384$<br>$\hat{\sigma}_e^2=0\cdot0116$ | |
| (7a) | 0·908 (0·937) | −310 | 1·457 | $\hat{\sigma}_\mu^2=0\cdot0360$<br>$\hat{\sigma}_e^2=0\cdot0116$ | |
| (6a) | 0·978 (0·939) | −335 | 1·040 | $\hat{\sigma}_\mu^2=0\cdot0369$<br>$\hat{\sigma}_1^2=0\cdot0168$<br>$\hat{\sigma}_2^2=0\cdot0050$ | $\hat{\sigma}_3^2=0\cdot0038$<br>$\hat{\sigma}_4^2=0\cdot0076$<br>$\chi^2=44\cdot8$ |

[a] $\bar{R}^2$s in parentheses are those from the dummy variable models used to obtain the variance estimates.

## *Mean yield*

The revised models for mean yield presented in Table 1 were based on several considerations. First, in the original models, the coefficient of $B^2$ (squared value of labour and bullocks) was positive, indicating increasing marginal returns to this factor. Since this is unlikely in an economy where labour is relatively abundant, $B^2$ was dropped from the revised models. This led to more plausible marginal effects for $B$ as reflected in the coefficients of $B$ and $B\times N$. In the final model (6a) where interaction terms were included and heteroscedasticity assumed, the coefficient of $B\times N$ was not significant and so this variable was also dropped.

Initially the socio-economic variables (area, experience and education) appeared to have little influence on the farmers' perceptions of mean yield, with experience in (7a) being the only significant variable at a 10 per cent significance level. However, the value of the Akaike criterion is considerably lower for (6a), suggesting that, in total, the socio-economic variables and their interactions with $N$ and $N^2$ could have some influence on subjective mean yield. Alternatively, the

TABLE 4

*Statistics for the Standard Deviation of Yield Equations*

| | $\bar{R}^2$ | AIC | $F^a$ | Variance estimates[b] and related statistics | |
|---|---|---|---|---|---|
| First-round models | | | | | |
| (7′b) | 0·920 (0·937)[c] | −515 | 8·272* | $\hat{\sigma}_\lambda^2=0\cdot632$<br>$\hat{\sigma}_v^2=0\cdot247$ | |
| (7b) | 0·919 (0·936) | −528 | 4·716* | $\hat{\sigma}_\lambda^2=0\cdot322$<br>$\hat{\sigma}_v^2=0\cdot247$ | |
| (6b) | 0·970 (0·949) | −563 | | $\hat{\sigma}_\lambda^2=0\cdot349$<br>$\hat{\sigma}_1^2=0\cdot302$<br>$\hat{\sigma}_2^2=0\cdot071$ | $\hat{\sigma}_3^2=0\cdot085$<br>$\hat{\sigma}_4^2=0\cdot096$<br>$\chi^2=56\cdot4$ |
| Revised models | | | | | |
| (7′b) | 0·920 (0·932) | −516 | 6·254* | $\hat{\sigma}_\lambda^2=0\cdot657$<br>$\hat{\sigma}_v^2=0\cdot266$ | |
| (7b) | 0·920 (0·932) | −530 | 2·900* | $\hat{\sigma}_\lambda^2=0\cdot373$<br>$\hat{\sigma}_v^2=0\cdot266$ | |
| (6b) | 0·970 (0·945) | −572 | 0·123 | $\hat{\sigma}_\lambda^2=0\cdot401$<br>$\hat{\sigma}_1^2=0\cdot352$<br>$\hat{\sigma}_2^2=0\cdot069$ | $\hat{\sigma}_3^2=0\cdot083$<br>$\hat{\sigma}_4^2=0\cdot011$<br>$\chi^2=67\cdot0$ |

[a] $F$ values which are starred are significant at the 5 per cent level.
[b] The variance estimates have been multiplied by $10^2$.
[c] $\bar{R}^2$s in parentheses are those from the dummy variable models used to obtain the variance estimates.

reduction in the value of the criterion may have occurred because the heteroscedasticity assumption in (6a) is more plausible. To investigate these possibilities further, the $F$ values described above were calculated, and experimentation with the omission of some variables was carried out. First, education was dropped because it seemed to have little effect in both the equations for mean and standard deviation. Then, the non-interaction terms, as well as (Area×$N^2$), were dropped from (6a). This procedure did not improve equation (7a), but in (6a) it did lead to significant coefficients (at the 10 per cent level) for (Area×$N$), (Experience×$N$) and (Experience×$N^2$). On the other hand, none of the $F$ values was significant. This last result, by itself, may mean that the socio-economic variables, and their corresponding interaction terms, are not important. All the results taken together lead to the conclusion that there is weak evidence supporting the inclusion of some of the socio-economic variables, but only through the interaction terms, and there is strong evidence in favour of the heteroscedasticity assumption.

Thus, from the models in Table 1, the revised version of (6a) is regarded as the best choice.

The fact that the socio-economic variables appear only to be important through interaction terms with $N$ and $N^2$ implies that these variables do not influence the farmers' perceptions about the *level* of mean yield, but they do influence perceptions about the *response* of mean yield to nitrogen. Specifically, from the revised (6a),

$$\partial E^*(Y)/\partial N = 0{\cdot}373 - 0{\cdot}0718 \text{ Area} - 0{\cdot}0164 \text{ Experience} - 0{\cdot}063N + 0{\cdot}0042N \times \text{Experience} \tag{13}$$

and evaluated at the sample means of area and experience, 0·106 ha and 2·1 years, respectively, this derivative becomes:

$$\partial E^*(Y)/\partial N = 0{\cdot}331 - 0{\cdot}054N \tag{14}$$

Thus, as expected, the 'average' farmer expects a diminishing marginal mean yield with respect to nitrogen. Maximum mean yield occurs at $N = 0{\cdot}331/0{\cdot}054 = 6{\cdot}13$ which is equivalent to 61·3 kg/ha.

Before turning to the results for the standard deviation of yield, some brief comments on the coefficient of the variety dummy variable will be made. Since the traditional variety is always used when $N$ is zero and the modern variety is always used when $N$ is positive, the response function measures the response of the modern variety to $N$, and it does not make sense to attempt to use the model to predict the response of the traditional variety to $N$. However, even though it does involve extrapolation beyond the range of the data, it may be reasonable to use the model to predict mean yield of the modern variety when $N$ is zero. Under these circumstances, the coefficient of the variety dummy gives the reduction in yield expected if the modern variety is used instead of the traditional variety when $N$ is zero. A reduction would, in fact, be expected because the modern variety does not perform well without nitrogen application.

*Standard deviation of yield*

In the equations for the standard deviation of yield in Table 2, the variables $B$, $B \times N$ and $B^2$ consistently failed to show any significant influence on $S^*(Y)$, and consequently, they were dropped from all the revised models. Rather surprisingly, $N$ and $N^2$ were not significant in the models without interactions with the socio-economic variables but, nevertheless, because of their obvious importance in (6b), they were retained in all the models.

The influence of the socio-economic variables was much more clear cut than in the equations for mean yield. Education had little effect, but both area and experience exerted strong influences as reflected by the $t$-values, the $F$ values, and the values for the Akaike information criterion. Also, there was strong evidence of heteroscedasticity. Thus, for a preferred model, the revised equation (6b), which excludes education and the terms involving $B$, was an obvious choice. In this equation, it could be argued that area should be dropped because of its non-significance; however, it was retained because of its clear importance in the interaction terms.

For the response of the standard deviation to nitrogen:

$$(15)\quad \partial S^*(Y)/\partial N = 0{\cdot}0509 - 0{\cdot}128\ \text{Area} - 0{\cdot}0124\ \text{Experience} - 0{\cdot}0064N + 0{\cdot}0244N\times\text{Area} + 0{\cdot}0026N\times\text{Experience}$$

which, evaluated at mean levels of area and experience, is:

$$(16)\quad \partial S^*(Y)/\partial N = 0{\cdot}0113 + 0{\cdot}0016N$$

This result means that, for the 'average' farmer, the perceived standard deviation of yield increases with the level of nitrogen application. This accords with the bulk of the literature based on experimental and field data (Roumasset, Rosegrant, Chakravarty and Anderson, in press).

## *Conclusion*

Collection of subjective response data has to date been very limited but, as further such data are elicited, demand must grow for econometrically efficient processing of the data into forms suitable for economic analysis and interpretation. The approaches described represent a start to the development and use of appropriate methods. Some of the complexities confronted arose from the cross section of pseudo-experimental data sets elicited. However, it is likely that every set of subjective data will feature a structure requiring consideration of some special error components.

The empirical results are of mixed quality, but at least the preferred subset of the most general specifications seems plausible. They reinforce other, often intuitive, notions that farmers in impoverished circumstances in developing countries perceive risk in the use of 'modern' inputs such as mineral fertilisers applied to supposedly high yielding varieties. Also, they suggest that farmers' perceptions can depend on characteristics of the farmers themselves and of their farms. The consequences of such perceptions of mean and variance, or of risk generally (Antle 1983), on farmers' choice of technology, must also depend on their aversion or otherwise to risk — a topic which has not been broached in this work, but which is taken up variously by, *inter alia*, Anderson (1973), Roumasset (1974), Anderson and Hamal (1983), Rosegrant and Roumasset (1985) and Griffiths (1986). There is still a pressing need for further methodological and empirical work before the matter of risk in response processes can be judged to be adequately understood.

## *References*

Amemiya, T. (1977), 'A note on a heteroscedastic model', *Journal of Econometrics* 6(3), 365–70, 'Corrigenda' (1978) 8(2), 265.

—— (1980), 'Selection of regressors', *International Economic Review* 21(2), 331–54.

Anderson, J. R. (1973), 'Sparse data, climatic variability, and yield uncertainty in response analysis', *American Journal of Agricultural Economics* 55(1), 77–82.

——, Dillon, J. L. and Hardaker, J. B. (1977), *Agricultural Decision Analysis*, Iowa State University Press, Ames.

—— and Hamal, K. B. (1983), 'Risk and rice technology in Nepal', *Indian Journal of Agricultural Economics* 38(2), 217–22.

Antle, J. M. (1983), 'Testing the stochastic structure of production: a flexible moment-based approach', *Journal of Business and Economic Statistics* 1(3), 192–201.

Breusch, T. S. and Pagan, A. R. (1979), 'A simple test for heteroscedasticity and random coefficient variation', *Econometrica* 47(5), 1287–94.

Goldfield, S. M. and Quandt, R. E. (1972), *Nonlinear Methods in Econometrics*, North-Holland, Amsterdam.

Griffiths, W. E. (1986), 'A Bayesian framework for optimal input allocation with an uncertain stochastic production function', *Australian Journal of Agricultural Economics* 30(2/3), 128–52.

—— and Anderson, J. R. (1982), 'Using time-series and cross-section data to estimate a production function with positive and negative marginal risks', *Journal of the American Statistical Association* 77(379), 529–36.

Hamal, K. B. (1982), 'Risk aversion and perception, and adoption of new technology and associated credit in Nepalese paddy production', M.Ec. dissertation, University of New England, Armidale.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lutkepohl, H. and Lee, T. C. (1985), *The Theory and Practice of Econometrics*, 2nd edn, Wiley, New York.

Just, R. E. and Pope, R. D. (1978), 'Stochastic specification of production functions and economic implications', *Journal of Econometrics* 7(1), 67–86.

Raiffa, H. (1968), *Decision Analysis*, Addison-Wesley, Reading, Massachusetts.

Rosegrant, M. W. and Roumasset, J. A. (1985), 'The effect of fertiliser on risk: a heteroscedastic production function with measurable stochastic inputs', *Australian Journal of Agricultural Economics* 29(2), 107–21.

Roumasset, J. (1974), 'Estimating the risk of alternate techniques: nitrogenous fertilization of rice in the Philippines', *Review of Marketing and Agricultural Economics* 42(4), 257–94.

——, Rosegrant, M. W., Chakravarty, O. and Anderson, J. R. (in press), 'Fertilizer and crop yield variability: a review' in J. R. Anderson and P. B. R. Hazell (eds), *Variability in Cereal Yields and Implications for Agricultural Research and Policy*, Johns Hopkins University Press, Baltimore.

Savage, L. J. (1971), 'Elicitation of personal probabilities and expectations', *Journal of the American Statistical Association* 66(336), 783–801.

Zellner, A. (1962), 'An efficient method of estimating seemingly unrelated regressions and tests of aggregation bias', *Journal of the American Statistical Association* 57(297), 348–68.