MORE REALISTIC SINGLE EQUATION MODELS THROUGH SPECIFICATION OF RANDOM COEFFICIENTS*

Max R. Langham and Michael Mara

Regression analysis with its many modifications and extensions plays a dominant role as an analytical tool in economic research. The linear regression model with random coefficients (hereafter RCR for random coefficient regression) provides a generalization of the classical linear regression model and permits a more realistic specification of the real world than does the classical model. As a consequence RCR will probably play an increasingly important role in econometric analysis of a wide class of problems--particularly as probabilistic micro-economic theory develops.

The first writings on some theoretical aspects of random coefficient models were by Hurwicz [3], and Rubin [4]. Major reference in this paper is to the basic set of consistent estimators developed by Hildreth and Houck [2]. Their estimators are a generalized extension of the earlier work by Theil and Mennes [9]. Swamy [6, 7] has been concerned with combining cross-section and time series data on a fixed set of individuals. In his work the coefficient vector was treated as random to account for interindividual heterogeneity. Froehlich [1] used Monte Carlo methods to ascertain small sample properties of various estimation procedures suggested in the literature on RCR model. Singh et al. [5] were concerned with the formulation of alternative hypotheses about the random character of the regression coefficients. They also studied methods of estimating RCR coefficients and applied the model in an analysis of structural change in the consumption function of certain countries. Zellner [10] was concerned with the aggregation problem. He showed that there is no aggregation bias for a certain class of regression models with random coefficients.

The major purpose of this paper is to discuss RCR and some of its merits and to extend it to provide a more explicit rationalization for specifying certain regressors (e.g. time) in single equation models. An example, based on a pesticide response function is presented. The approach seems particularly applicable to many environmental quality problems.

RANDOM COEFFICIENT REGRESSION

The model may be specified as follows:

(1.0) $Y_i = \beta_{io} + \beta_{i1}X_{i1} + \ldots + \beta_{iK}X_{iK}$, $i=1,\ldots,n$;
where,

(1.1) $\beta_{ik} = \beta_k + v_{ik}$, $k=0, 1, \ldots, K$;

(1.2) $\beta_k =$ a constant, the mean response of the dependent variable to a unit change in the k^{th} independent variable;

(1.3) $v_{ik} =$ unobserved random variable;

(1.4) $E(v_{ik}) = 0$.

(1.5) $E(v_{ik}v_{jg}) = \sigma_{kk}$ if $i = j$ and $g = k$
$= 0$ otherwise.

Substitute for β_{ik}'s in equation (1.0) and get:

(2.0) $Y_i = \beta_o + \beta_1 X_{i1} + \ldots + \beta_K X_{iK} + u_i$; $i = 1, \ldots, n$;

where, $u_i = v_{io} + \sum_{k=1}^{K} v_{ik}X_{ik}$;

$$E(uu') = \theta = \begin{bmatrix} \theta_{11} & & 0 \\ & \ddots & \\ 0 & & \theta_{nn} \end{bmatrix}$$ and if the X_{ik} are fixed;

(3.0) $\theta_{ii} = \sigma_{oo} + \sum_{k=1}^{K} X^2_{ik}\sigma_{kk}$ because of (1.5).

Max R. Langham is professor and Michael Mara is research assistant of the food and resource economics at the University of Florida.

*Florida Agricultural Experiment Station Journal Series No. 4891 under State Project AS 01636.

The classical linear regression model is a special case of the RCR model when $v_{ik}=0$ for $k=1,...,K$. That is, in the classical linear regression model, random variation is admitted in the intercept only. In most model applications there is no more justification to assume apriori that the intercept is subject to random variation and that the slope coefficients are all constant than there is to assume that all coefficients are subject to random fluctuation. It is intuitively appealing to assume that random fluctuations may appear in the other coefficients. If one were fitting a production function linear in raw interfarm data, for example, it seems more realistic to assume that the intercept and the marginal productivities are subject to random variation among farms than to just assume that the intercept is subject to random error.

Since the variance of u_i is a function of the X_{ik}'s, it is heteroskedastic and ordinary least squares (OLS) will yield unbiased but inefficient estimators. Hildreth and Houck [2] and Froehlich [1] have suggested several alternative methods for estimating the structure of equation (2.0) which yield consistent estimators. Each of their methods is a feasible Aitkin's estimator, i.e. a generalized least squares estimator which utilizes an estimate in lieu of actual knowledge of θ.

The variance of the error term of equation (2.0) is of particular significance to economists. The independent variables include the instruments which management controls.[1] With the classical linear regression model, one assumes that manipulation of these variables affects only the average value of Y. However, with the RCR model the action of the decision maker affects not only the average value of Y but also its variance. If the decision maker's utility is affected by both the average value of Y as well as its variance, he would want to take this into account in his decision to manipulate the independent variables. This feature of the RCR model probably makes it a closer representation of most real world decision environments than the classical linear regression model.

Estimation

The OLS estimator of the coefficient vector of equation (2.0), $\hat{\beta}=(X'X)^{-1}X'Y$, is unbiased but inefficient because of the heteroskedastic disturbance. A best linear unbiased estimator is provided by the Aitken's generalized least squares estimator:

$$(4.0)\quad \check{\beta}= (X'\theta^{-1}X)^{-1}X'\theta^{-1}Y$$

A major problem with this estimator is lack of information about θ. An alternative is to estimate θ and to use the estimate to derive a generalized feasible Aitken's estimator that is consistent and asymptotically efficient. The methods for estimating the σ_{kk} developed by Hildreth and Houck [2, p. 586-587] are based on the vector of residuals derived from the OLS regression of Y on the X's.

Hildreth and Houck [2, 592-594] show that their estimator of the σ_{kk} will lead to a consistent estimator of β. Theil [8, p. 624] shows that when OLS is used to estimate the σ_{kk} the error term is also heteroskedastic and suggests using a generalized feasible Aitken's estimator to estimate the σ_{kk}.

The σ_{kk}*'s which are estimates of the σ_{kk}'s are then used to estimate θ as per equation (3.0).

The estimated matrix θ^* is used in turn to derive a consistent estimator of the coefficient vector, i.e.

$$(5.0)\quad \beta^*= (X'\theta^{*-1}X)^{-1}X'\theta^{*-1}Y.$$

One difficulty with the OLS approach in estimating the σ_{kk} is that it can yield negative estimates of σ_{kk}. To avoid this problem Hildreth and Houck suggested two other estimators $\tilde{\sigma}$ and $\hat{\sigma}$. These two were shown to be consistent estimators and were defined as follows:

$$\tilde{\sigma}_{kk}= \max \{\sigma_{kk}^*, 0\}$$
$$\hat{\sigma}_{kk}= \text{a quadratic programming estimator.}$$

Using Monte Carlo experiments, Froehlich found that $\hat{\sigma}_{kk}$ performs as well as other estimators that have been suggested. Some of his conclusions were [1, pp. 14-16]:

1. For variance estimation $\hat{\sigma}_{kk}$ suggested by Hildreth and Houck is superior (in terms of mean square error) to the other Hildreth and Houck estimators. This the authors had conjectured. The gain over σ_{kk}^* was as much as 40 percent in samples of size 25 and 30 percent in samples of size 75.
2. Although $\tilde{\sigma}_{kk}$ is truncated and known to be biased, the bias is persistently negligible provided the true variance value is not "near" zero.
3. The more desirable two-stage procedures, although leading to substantial gains in efficiency for variance estimation, do not on the average give similar gains over ordinary least squares in estimating the mean response coefficients. This is true for sample size 25 as well as sample size 75.
4. Ordinary least squares estimation of the mean response coefficients is, on the average, a very

[1] Economic models may of course include regressors which are uncontrolled or predetermined.

satisfactory procedure even for samples as small as 25.

Singh et al. [5] presented two techniques of estimation. In the first they outlined Hildreth and Houck's technique and their second approach was a maximum likelihood estimator using a modified Gauss-Newton Technique. They concluded that the Hildreth and Houck estimator σ^* and the maximum likelihood estimator are for practical purposes identical. They also specified coefficients of both income and lagged consumption as shifting with time for some of the countries they studied.

Theil has shown [8, p. 626] that the estimation of σ^* is quite imprecise and that a considerable number of observations is needed for reasonable precision.

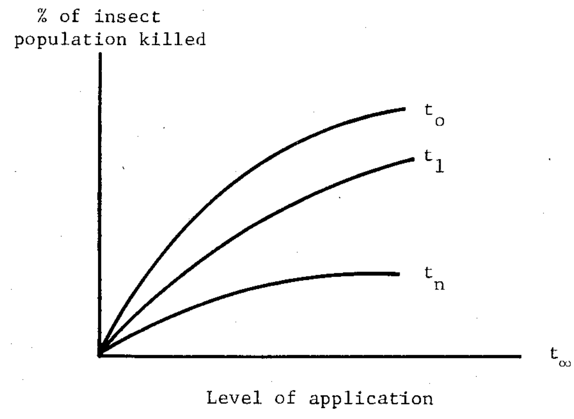

Figure 1.

COEFFICIENTS SPECIFIED AS STOCHASTIC FUNCTIONS

RCR can be thought of as a special case of a more general specification in which the coefficients of the model are in turn stochastic functions of other variables. Certain phenomena associated with quality of the environment problems, for example, suggest this rationale for the use of RCR. If insects are known to develop a resistence to such substances as D.D.T. so that if one expressed the percentage of an insect population which was killed with a given level of application one would expect the function to shift in time as perhaps suggested by Figure 1.

The curve shifts in time because the insect is in fact a different (more resistent) life form. Much the same phenomenon exists when there is a shift in technology–e.g. hybrid corn is a different product than open pollinated corn. Much the same phenomenon may also exist when there is an improvement in the quality of a human resource due to experience or education. In each of these cases one may not know enough to specify the precise explanatory variables that creates the change in the response of the dependent variable, but one can perhaps measure the extent of the structural change as a function of proxy variables.

For example, a model which describes the phenomenon[2] in Figure 1 may be expressed as follows:

$$(6.0)\quad Y_t = \beta_{to} x_t^{\beta_{t1}} \qquad t=1,\ldots,n$$

where

$$(6.1)\quad \beta_{to} = e^{(\gamma_0+\gamma_1 t+u_t)}$$

$$(6.2)\quad \beta_{t1} = \delta_0+\delta_1 t + v_t$$

$$\text{or } \ell n\, Y_t = (\gamma_0+\gamma_1 t+ u_t) + (\delta_0+\delta_1 t + v_t)\, \ell n\, x_t$$

---

[2] We used an alternative specification of this phenomenon and some experimental data (supplied by P. H. Clark and M. M. Cole of the Entomology Research Division, Agricultural Research Service, USDA, Gainesville, Florida) on succeeding generations of a colony of body lice to estimate equation (9.0). Each generation had been subjected to specified levels of lindane (a chlorinated hydrocarbon). A shock model was postulated as follows:

$$(9.0)\quad y_{ig} = 100 - \frac{\beta_{ig}}{X_{ig}} \ ;$$

$$(9.1)\quad \beta_{ig} = \gamma_0 + \gamma_1 g + u_{ig};\ \text{where}$$

y_{ig} = percent of the g^{th} generation killed when exposed to the i^{th} concentration of lindane;
x_{ig} = percent concentration of the i^{th} dosage of lindane applied to the g^{th} generation;
β_{ig} = random coefficient which is a stochastic function of the generation number (equation 9.1); and
u_{ig} = is a spherical disturbance. This assumption is not completely valid because the range of y_{ig} is truncated. A similar criticism exists when one uses ordinary least squares to fit demand functions, production functions, etc.

This model can be estimated with ordinary least squares by regressing $(100-y_{ig})x_{ig}$ on g. If the function shifts as depicted in Figure 1 then $\gamma_1 > 0$. The results when the function was fitted to 75 observations were: $\hat{\beta}_g = .00454 + .03026\ g$, with $\bar{R}^2 = .34$. $\quad$ (.00481)

The estimated standard error of the coefficient of g (in parentheses) indicates that the null hypothesis should be rejected. The variance of the error is $\sigma^2 x_{ig}^{-2}$. This inverse relationship between dosage and variance of kill would encourage greater pesticide use of the decision maker received utility from a lower variance in the percent pests killed.

$$(7.0)\ \ell n\, Y_t = \gamma_0 + \gamma_1 t + \delta_0\, \ell n\, x_t + \delta_1 t \ell n\, x_t + w_t$$
$$\text{where } w_t = (u_t + v_t\, \ell n\, x_t)$$

Here β_{to} and β_{tl} are random functions of time. Equation (7.0) is a fixed coefficient model with a heteroskedastic disturbance and one could argue for the specification of this equation directly. However, the specification provided by equations (6.0) to (6.2) is more explicit. As a consequence, the reader knows more about the rationale behind the model. The heteroskedastic disturbance permits a change in the variance of the dependent variable with a change in the regressor. This result will in many applications, as was indicated above, permit a more realistic model specification.

Since one can approximate any continuous function with a polynomial of suitable order, one could with ample data estimate a coefficient which is some complicated continuous function of time.[3]

One could also estimate coefficients that have discontinuities. For example, the function depicted in Figure 2 could be accurately specified by an RCR model using zero-one variables as follows:

$$(8.0)\ Y_t = \beta_{ot} + \beta_{1t} x_t$$
$$(8.1)\ \beta_{ot} = \beta_o + u_t$$
$$(8.2)\ \beta_{1t} = \gamma_0 + \gamma_1 Z_t + v_t$$
i.e.
$$Y_t = \beta_o + \gamma_o x_t + \gamma_1 x_t Z_t + (u_t + v_t x_t)$$
$$\text{where } Z_t = 0 \text{ if } t < t_o$$
$$= 1 \text{ if } t \geq t_o$$

It should be stressed that the specification of coefficients as stochastic functions of time is for illustrative purposes only. Coefficients in the model

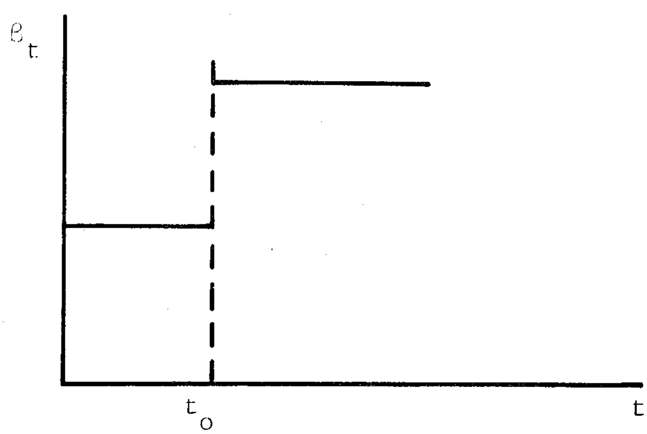

Figure 2.

could certainly be specified as functions of other variables. For example, if the coefficients in a model were some function of χ the model in (8.0) could be depicted as

$$Y_t = [f_o(\chi_t, u_t)] + [f_1(\chi_t, v_t)] X_t$$

ECONOMIC MODELS AND RCR

In this section an economic model is considered in which (1) the variables manipulated by the decision maker affect not only the mean outcome of the decision but also the variance of the outcome, and (2) the decision maker's utility is also affected by the variance of the outcome. In such models, RCR would seem more appropriate for empirically fitting underlying relationships that have traditionally been fitted with ordinary least squares.

Assume, for example, that the utility which a producer expects to derive from production is a monotonically increasing function of his expected profits and a monotonically declining function of the variance of profits, i.e.,

$$U = f(E\pi, \sigma_\pi^2).$$

Also, for simplification, assume that expected revenue comes from one commodity which is produced with two variable inputs X_1 and X_2 according to the function,

$$EY = g(X_1, X_2)$$

and variance function

$$\sigma_y^2 = h(X_1, X_2).$$

Lastly, assume that X_1 and X_2 are purchased at fixed prices p_{x_1} and p_{x_2} that Y is sold at a fixed price p_y, and that the producer faces a capital constraint given by K.

If the functions f, g, and h were known one could find those values for X_1 and X_2 which maximized the producer's expected satisfaction directly by maximizing U subject to the capital constraint. In the absence of knowledge of the function f an alternative would be to maximize expected profits subject to an acceptable (to the decision maker) variance constraint and the capital constraint, i.e.,

[3] A generalization of the model where the coefficients are stochastic functions of time is available from the authors.

$$\max NR = p_y\, g(X_1, X_2) - p_{x_1} X_1 - p_{x_2} X_2 - FC$$

s.t.

$$p_{x_1} X_1 + p_{x_2} X_2 \leq K$$
$$h(X_1, X_2) \leq \sigma^2$$
$$X_1, X_2 \geq 0$$

where, FC = fixed costs.

The first inequality arises because one would not want to require that capital be exhausted if it were more profitable to leave some idle. The second arises because a variance smaller than σ^2 would be perfectly acceptable to the decision maker.

A graph of the problem might exist as in Figure 3. Here only capital is binding on the solution. Of course, the variance or both variance and capital could be binding.

The problem will be recognized as a non-linear programming problem. The area of feasible solutions is shaded. If both constraints were binding, the Lagrangean function would be:

$$F = p_y\, g(X_1, X_2) - p_{x_1} X_1 - p_{x_2} X_2 - FC + \lambda_1 (p_{x_1} X_1 + p_{x_2} X_2 - K) + \lambda_2 (h(X_1, X_2) - \sigma^2),$$

where λ_1 and λ_2 are multipliers. The necessary conditions for a maximum are given by

$$\frac{\partial F}{\partial \lambda_i} = \frac{\partial F}{\partial X_i} = 0, \quad i = 1, 2.$$

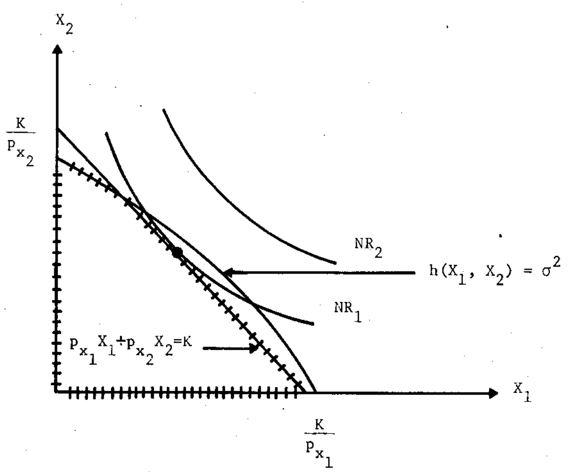

Figure 3.

Kuhn-Tucker theory would indicate a zero value for λ_2 (a measure of the marginal expected profits associated with the decision maker's willingness to accept an increment in the variance of profits) for the problem in Figure 3. In this case the decision maker's willingness to accept an additional unit of variance in profits would have no effect on expected profits attainable at current levels of resource use.

The specification of the model indicates that values of the inputs affect not only the mean value of Y and hence expected profits but also the variance of Y. RCR provides a statistical model consistent with this economic model. Fitting of the RCR model can provide a means of estimating functions g and h under certain specifications.

CONCLUDING REMARKS

The idea of specifying certain coefficients to be stochastic functions of time was suggested earlier in those cases where one unit of dependent variable at time t was not homogeneous with a unit at time t+k. The examples of more resistent life forms and hybrid corn were used. Changes in the quality of certain inputs provide another situation where coefficients which are stochastic functions of time may make sense in model specification. If labor, for example, improves in quality over time and one cannot specify the true causes of the quality change, allow for associated changes in productivity by specifying the coefficients to be stochastic functions of a time proxy. Or, if measurement indicates that the coefficients of certain inputs vary with time under conditions of no change in the quality of inputs, the cause may be attributed to more efficient ways of coordinating the inputs (management).

However, the main point is that the RCR model permits one to explicitly assume that the level of the regressors affects the variance of the outcome as well as its expected value and hence allows greater realism in economic model specification than does the standard fixed coefficient model. If one is only interested in point estimates of the means of the random coefficients, Froehlich's work [1] indicates that ordinary least squares remains quite satisfactory as an estimating technique.

REFERENCES

[1] Froehlich, B. R., "Estimation of a Random Coefficient Regression Model," unpublished Ph.D. Thesis, University of Minnesota, 1971.

[2] Hildreth, Clifford and J. P. Houck, "Some Estimators for a Linear Model with Random Coefficients," *Journal of the American Statistical Association,* 63, 322, June, 1968, 584-595.

[3] Hurwicz, Leonid, "Systems with Nonadditive Disturbances," *Statistical Inference in Dynamic Economic Models,* edited by T. C. Koopmans. Cowles Commission Monograph No. 10, 1950, 410-418.

[4] Rubin, Herman, "Note on Random Coefficients," *Statistical Inference in Dynamic Economic Models,* edited by T. C. Koopmans. Cowles Commission Monograph No. 10, 1950, 419-421.

[5] Singh, Balvir, A. L. Nagar and Baldev Raj, "Measurement of Structural Change (An Application of Random Coefficient Regression Model)," unpublished paper, March, 1972.

[6] Swamy, P. A. V. D., "Efficient Inference in a Random Coefficient Regression Model," *Econometrica,* 38, 2 March, 1970, 311-323.

[7] ——, "Statistical Inference in Random Coefficient Regression Models," unpublished Ph.D. Thesis, University of Wisconsin, 1968.

[8] Theil, Henry, *Principles of Econometrics.* New York: John Wiley and Sons, Inc., 1971.

[9] Theil, H. and L. B. M. Mennes, "Multiplicative Randomness in Time Series Regression Analysis," Mimeographed Report No. 5901, Econometric Institute of the Netherlands School of Economics, 1959.

[10] Zellner, Arnold, "On the Aggregation Problem: A New Approach to a Troublesome Problem," Report No. 6628, Center for Mathematical Studies in Business and Economics, University of Chicago, 1966, earlier draft July 25, 1965.