(832)

(a) UNIVERSITY OF NAIROBI
(b/) Institute for Development Studies
Working papers

SEARCH PROCEDURES FOR OPTIMAL POLICY IN STOCHASTIC, TIME-DEPENDENT, NONLINEAR SIMULTANEOUS EQUATIONS MODELS

By

Robert James Whitacre

WORKING PAPER NO.215

Department of Economics
University of Nairobi
P.O. Box 30197
Nairobi, Kenya

June, 1975

IDS/WP 215

SEARCH PROCEDURES FOR OPTIMAL POLICY IN STOCHASTIC, TIME-DEPENDENT, NONLINEAR SIMULTANEOUS EQUATIONS MODELS

by

Robert James Whitacre

Synopsis

Several alternative numerical approximation techniques suitable for computer simulation are derived which are capable of solving stochastic, time-dependent, nonlinear simultaneous equations models. Their applicability in the Kenyan context in the solution of 'Jembe', the most recent World Bank Model of the Kenyan Economy, named 'Jembe', as well as deriving the solution of the model contained in I.D.S. Working Paper number 216 by this same author.

1. Introduction

This paper derves numerical approximation techniques applicable to optimal policy specifications upon the time path of controllable variables of stochastic, time-dependent, nonlinear simultaneous equations models where the feed links between successive stages constitute a multivariate vector of finite dimension and time horizon. Specifically optimization requires approximating the infimum of a loss function defined over the space of possible endogenous variable values.

Selecting a macroeconomic policy package extending over several time periods under specified objective criteria resulting from a stochastic non-linear econometric model falls within the scope of this analysis.

It is known[1] that nonstochastic simulation, that is repeated solution of such a model with the exclusion of the error terms incorporated in the structural equations, will lead to policies which: (1) fail to optimize the designated criteria, (2) are statistically biased; the degree of bias being indeterminable, and (3) are incapable of being compared with posterior historical values since confidence intervals on the time path of the endogenous variables will not be established.

Stochastic simulation, repeated solution of the model with error terms attached to the structural equations drawn from the corresponding structural equation error term distributions, has been shown /12/ to asymptotically yield an error distribution on the path of the endogenous variables which converges uniformly to the true multivariate distribution function. Such convergence is, of course, conditional upon the policy specified. Furthermore it is argued /12/ that the Bayes' Principle, choosing that policy which minimizes expected risk (or in economic terminology, that policy which maximizes expected utility), is capable of finding an optimum policy when the set of potential policies is finite.

---

1 See R.J. Whitacre /12/ for a mathematical demonstration of these conclusions.

This present investigation analyzes policy determination under the Bayes' Principle when the set of potential policies is unrestricted, i.e, all specifications on the exogenous variables which do not violate the original constraints of the problem are admissible. It is found that most of the algorithmic techniques, both gradient and nongradient methods, which can be utilized to approximate an optimum in the nonstochastic case can be included as components of stochastic maximization algorithms. In addition, some conditions under which such extensions are valid and several alternative stochastic maximization algorithms are then suggested.

Section 2, "Principles of Stochastic Maximization," compares stochastic maximization with nonstochastic maximization and specifies necessary conditions for an algorithmic technique to converge to an optimum.

Section 3, "Algorithmic Techniques of Stochastic Maximization," develops various alternative stochastic maximization algorithms and Section 4, "Applications", briefly describes an application of the theoretical stochastic maximization techniques which is currently being conducted. These findings will ultimately constitute a Monte-Carlo type study comparing several of the alternatives available.

2. Principles of Stochastic Maximization

Only in the nonstochastic univariate case where search for a unimodal function is restricted to a finite region has it been possible to find an optimal search technique.[2] No such measure of efficiency has been proposed for the various n-dimensional search techniques, and even for unimodal multivariate functions a "best" technique cannot be found as information empirically gathered depends upon the technique employed to gather this information. Different techniques will have different degrees of relative efficiency depending upon the particular objective function. This arises as different techniques attempt to exploit various types of information which may or may not be present in a

---

2. The Fibonacci method is known to be preferable by several criteria when the number of experiments is of major concern.

particular objective function. Different objective functions for maximization will normally vary in both the nature and quality of information in different regions in which the maximization process is conducted. In the general problem, evaluation of alternative methods must be examined in light of the particular problem being investigated; this should be done analytically where possible. However, Monte-Carlo type methods should prove as useful in this field as they do presently in the comparison of alternative statistics in econometric investigations.[3]

For stochastic models, there are general characteristics which must be contained within any search procedure attempting to find that policy which minimizes the risk function. Every such procedure must be sequential in nature, specifically for each policy specification a sample set must be drawn from the structural disturbances, the resulting system of simultaneous equations solved, the objective function evaluated and the resulting information incorporated into the new posterior evaluation of the risk function. Then this evaluation of the risk function is compared with the current risk function evaluations for alternative policies and finally a decision made as to the next policy to be evaluated.

Thus in any logical method for stochastic maximization there are two components of any algorithmic procedure which are of primary importance: (1) when should sample evaluations of the current value of the empirical risk function cease for a given policy and (2) what criteria should be used to determine the succeeding policy to be evaluated.

Denoting by T_n the current empirical frequency distribution of the endogenous variables for sample size n(that is n drawings have been taken from the structural disturbances and the model solved for the endogenous variables for each drawing) subject to a given specification on the endogenous variables x it has been shown in [10] that

(1) $\lim_{n \to \infty} T_n = T,$

where T is the true posterior frequency distribution of the endogenous

---

3. From a system viewpoint the present state of econometric analysis analysis for macroeconomic models must be considered seriously deficient. Attention has been focused upon the estimation of econometric simultaneous equations models to the near exclusion of the ultimate objective for formulating such models; (optimum policy determination.

variables. Thus if we denote the risk function of a policy x subject to T_n by $r(x,T_n)$ it is the case that

(2) $\lim_{n\to\infty} r(x,T_n) = r(x,T)$.

For any finite n however the equality given in equation (2) need not hold and, in general, it is in possible to deduce whether the current posterior evaluation $r(x,T_n)$ will be greater or less than $r(x,T)$. Therefore any search procedure which permanently eliminates from further search cannot be shown to approach the optimum policy in the asymptotic limit.

A decision to stop evaluation of the current policy might be based upon such diverse criteria as (1) a random mechanism, (2) prior decisions before the initiation of the iteration process, (3) the path of moments of T_n or $r(x,T_n)$ for the current policy being evaluated and those that have already been evaluated, (4) the nature of r, e.g. the specification and number of the constraining stochastic simultaneous equations and the structure of losses associated with the various states that may arise, (5) the costs of evaluation, (6) the time required for evaluation, and (7) the purpose to which the model is to be put.

When deciding what criteria should be used to determine the next policy to be evaluated the proceeding list is applicable. But the proper direction and length of movement from the present policy to a new one to be evaluated must also be considered.

If direction can be chosen from but a finite number of choices, the direction of movement need not necessarily be in the direction of greatest improvement in the objective function. If, on the other hand, the choice is unrestricted, it would seem desirable that the movement be guided by the gradient of the risk function at this point. Determination of the gradient requires that the nature of the risk function be known within a neighborhood of the present policy.

Distance of movement along the desired direction must consider the length of increment for such movement. If step sizes are too small convergence to an optimal policy will be very slow even if the right direction is chosen. If step sizes are too big the value of the risk function may decrease even though the true optimum lies in the proper direction.

Only in special circumstances will an extremal value of the risk function even in the asymptotic limit be known to be a global optimum; the basic requirement is clearly the same as that encountered in geometric programming, that the risk function be convex with respect to the exogenous policy variables x. For even a moderately complicated problem of the type under discussion it would be extremely difficult if not impossible to prove convexity of the risk function in the asymptotic limit. Such proof, however would be purely superfluous in the case of algorithmic stochastic maximization. The values of T_n for any given policy x are, of course, conditional upon the drawings made from the structural disturbances. Different sets of drawings will normally yield different values of T_n and therefore of $r(x,T_n)$. Even if it were known therefore that r(x,T) were convex in x this would not guarantee that $r(x,T_n)$ was convex in x even if T_n for each x corresponded to the same sets of drawing from the structural disturbances. In such circumstances where a function is not known to contain a single optimum a widely accepted procedure is to begin the search at a number of initial base points widely dispersed. Indeed this was one of the motivations for the recommended finite set of diverse policies to be evaluated in /12/. Then the maximum of the suboptimums generated from these initial search points should be chosen.

On the basis of this discussion any procedure which starts with alternative specifications of the error terms and then proceeds to find the optimal policy contingent upon these error terms seems destined to failure. It appears necessary that a policy be specified subsequently followed by repeated drawings from the disturbance terms to guarantee convergence to an optimum policy. Thus any penalty function approach where the exogenous variables are included in the maximization procedure seems invalid (the penalty function approach, to be covered in the next section, is valid however when this is not the case).

3. Algorithmic Techniques of Stochastic Maximization

3.1 Quantifying the Problem

It is now necessary to quantify further the original specification of the problem. Denote by y the matrix of endogenous variables, where the first subscript refers to the time index while the second refers to the number of endogenous variables to be determined from the constraining simultaneous equations of the model. If there are m time periods and n endogenous variables then

3-1 $y = \underset{\sim}{[}\bar{y}_{ij}\underset{\sim}{]}_{m \times n}$

Similarly let us denote by x the matrix of exogenous variables which define a policy,

3-2) $x = \underset{\sim}{[}\bar{x}_{ik}\underset{\sim}{]}_{m \times p}$

and b be the matrix of estimated coefficients in the constraining equations, assumed here to be constant throughout the model (the analysis can easily be modified to where this need no longer hold),

3-3) $b = \underset{\sim}{[}\bar{b}_{rs}\underset{\sim}{]}_{r \times s}$.

Suppose that there are z constraining equations for which there are also associated error terms u_v^t, $1 \leq v \leq z$ at time period t.
Then the system of constraining equations at time period t is given by

3-4) $g_1(y^t, x^t, b, u_1^t) = 0$

$\vdots$

3-4A) $z(y^t, x^t, b, u_z^t) = 0$

Here the t superscript indicates the t^{th} row of the matrices y and x respectively. The error terms u_v are assumed drawn from the error distributions of the associated equations of the constraining simultaneous equations model.

To relate the system of equations 3 - 4 to the usual applied econometric analysis it is merely necessary to set $u_i^t = 0 \forall i = 1, \ldots, Z$. The resulting model is then

3-4B) $g_1(y^t, x^t, b, 0) = 0.$

$\vdots$

$g_z(y^t, x^t, b, 0) = 0.$

Given this simultaneous system the standard technique has been to specify alternative policies (i.e., alternative x^t) and solve for the corresponding y^t; A specific objective function defined over y^t guiding the alternative x_t not being utilized,

---

* In the case of the Wharton model this was, until quite recently, accomplished via Newton's method.

Before continuing the specification of the problem, let us note that for problems with inequality constraints all such problems can be converted into the form of 3-4 by the addition of approporiately specified slack variables. Also the presence of equality restrictions forces a choice only of those points y^t which satisfy the equations 3-4. Such a point may be difficult to obtain and it will usually be necessary to conduct a separate search for a feasible initial search point before optimization can begin. One practicable method is to find that y^t which minimizes the sum of squares

$$3\text{-}4C) \qquad S=\sum_{k=1}^{z} g_k 2,$$

thus converting our search for a feasible point into one of a search in an unrestricted form; then many numerical or analytical methods are potentially useable. If a solution exists, it will, of course be at S=0. In the case where we are analyzing repeated samples from our error distributions attached to g_1 to g_z solution with respect to one specification upon the u_y does not necessarily imply that a solution will exist for a different specification upon the u_v, as the system of equations to be solved is nonlinear. Obviously if at any time during the solution procedure it becomes impossible to solve the set of constraining simultaneous equations for some drawing of the u_v from the set of structural disturbances then there is a serious specification error present within the constraining equations.

Specification of each interaction in the simulation procedure for a given policy and error specification in the form of minimization contained in equation 3-4 implicitly assumes that failure to satisfy each equation is given equal weight in the overall minimization process. This would be satisfactory if solution always proceeds to a specification of y* such that S=0. However as will be argued in the next section determination of the endogenous variables will typically stop short y* and there will usually be an error resulting in termination of the process at some value $\bar{y}$ close to y*. Since $\bar{y}$ will depend upon the specification of the underlying objective function it is suggested here that instead of equation 3-4 that the following objective function be utilized.

$$3\text{-}4D) \qquad S = \sum_{k=1}^{z} A_k g_k^2$$

where the coefficients A_K are weighted according to the associate explanatory power of the associated equations g_K. Intuitively those equations which have been estimated with but little accuracy (i.e., those with a low R^2) we would not be overly concerned with violating whereas we would with those of high explanatory power. Moreover those identities which are explicitly specified within the model we would not want to violate at all. As a practical suggestion the coefficients A_K can be given the values of the associated correlation coefficients (i.e., the R^2 values) while for those equations which are identities can be assigned the value $A_k = 1$.

Rewriting the system of equations in 3-4 we have

3 - 6) $g(y,x,b,u) = 0$ where

3- 7) $g(y,x,b,u) = g^1(y^1,x^1,x^1,b,u^1)$

$$\vdots$$

$$g^m(y^m,x^m,b,u^m),$$

and in 3-6 0 is a tm column vector of zeroes.

Now given a policy x and a drawing u from the structural equations, both over the m time periods involved, the solution vector y is obtained. Different drawings of u, if drawn from the structural error distributions, generate the posterior error distribution $T_n(y)$. The solution for each drawing of error terms u(n) generates the matrix $y(n) = y(n/x,u)$. Given a loss function defined over y denoted by h(y) then for n observations upon the error terms u drawn from the structural disturbances for a given policy x expected loss is given by

$$3 - 8) \quad r(x,T_n) = \frac{\sum_{t=1}^{n} h(y(i)|x).}{n}$$

From Section 2 we have equation 2-2 stating that $\lim_{n\to\infty} r(x,T_n)=r(x,T)$ which implies that $\lim_{n\to\infty} \text{var}(r(x,T_n))=0$. However for small sample size the variance of the risk function for any policy might be quite high as the frequency distribution T even in the asymptotic limit cannot be

expected to have zero variance.[4]

3-2 Classification of Search Techniques

The initial distinction must lie between those which require exact solution of equation 3-6 for each drawing from the structural disturbances and those which do not. Particularly given a policy specification and a specified error matrix an attempt can be made to find the optimum policy where some estimate $\bar{y}$ of y is utilized, implying that a resulting estimate $\bar{h}$ of h is used, i.e., $\bar{h}/\bar{y}(u)/$ is used as an estimate of $h/\bar{y}(u)/$. Since for any drawing u the use of h is used in determining $r(x,T_n)$ which is itself utilized as a statistic measuring $r(x,T)$, or used in finding the median as previously mentioned, it seems that terminating the search for an approximation of y at a level $\bar{y}$ would considerably ease the computational burden. Specifically if we write the risk function as $r(x,y,T)$ we will utilize approximations $r(x,\bar{y},T_n)$. Also it is not obvious that in the general case an exact solution of 3-6 can readily be found. Indeed most algorithmic techniques that can be utilized to solve 3-6 would happen along the exact solution with probability equal to zero.

Since 3-6, can be regarded as a deterministic system once the error terms have been incorporated from a specific drawing from the structural equations, any of the nonstochastic solution approaches can be utilized to obtain the approximation $\bar{y}$ of y. This may be done as in the conversion to an unrestricted form as in 3-4B or directly by successive substitution as in the Gauss-Seidel method.[5]

Since the conversion into 3-4 of equations 3-6 may be regarded as an unconstrained maximization problem any of the algorithmic techniques

---

4 Since the distribution function for small samples must be taken as unknown a better measure of $r(x,T)$ might be the median of $h(y(i)/x)$, $i=1,\ldots,n$. The median presumably is an effective method of preventing an overpowering influence of outlier observations upon the algorithmic solution process. Since many potential stochastic maximization processes might presumably start off the search with but a few observations upon each potential policy the median might prove a better statistic in such early stages. Whether or not the median should be used and if used when it should be replaced would depend upon the particular algorithmic solution technique under examination as well as the particular problem. Such determination would probably have to take place through Monte-Carlo type numerical methods rather than analytic techniques.

5. The Gauss-Seidel method can be found in Fisk /_ 3 _/

suitable for maximization which have been already developed for maximization of unconstrained nonstochastic systems are applicable at this stage. The overall error term allowable for the resulting solution vector will of course have to be determined by the stochastic maximization process involved in finding an approximation to the optimal policy x^* determined from $r(x^*,T) = \max r(x,T)$. A short listing of the more important, e.g. those which have shown themselves to be relatively efficient must include in the direct search methods the Sequential Simplex Method (as it is simple to program, has the smallest number of points to start, needs only one experiment for each new movement, and is the easiest to understand), the Rosenbrock technique, its modification by Davies, Swann, and Campey, the Campey and Nichols method, the Nelder and Mead technique, and the Complex method of Box; and in the gradient methods those of Kelley, Hildebrand, Cauchy, the method of conjugate directions, Carol and Fletcher-Reeves. A complete listing of these various techniques here would be inappropriate as they have appeared in the literature. The original sources outlining these techniques as well as others can be found in the bibliography. The available Monte-Carlo studies, of which there are as yet very few, seem to indicate that the gradient methods appear to converge more rapidly than the nongradient techniques; although more evidence is clearly needed. For this stage of the stochastic maximization problem, the gradient may be fairly easy to calculate as the objective function for maximization is known, e.g. it is given by equation 3-4D, and the necessary partial derivatives can be determined by straightforward differentiation and evaluation.

In approximating $r(x,y,T)$ by $r(x,\bar{y},T_n)$ there will be a tradeoff between the computational burden determining $\bar{y}$ as an approximation of y at each stage of the iteration procedure as compared to utilizing T_n as an approximation to T. In addition it is very difficult to deduce quantitatively what this trade-off will be; indeed it is most likely impossible upon an analytical level. For given specifications of a stochastic constraining model as in 3-6 the distribution of $\bar{y}$ from the true values y for given drawings of the error terms will depend upon both the algorithmic solution procedure chosen, the nature of the equations in the model, and the particular sets of error terms drawn from the structural equations. While the rate of convergence of T_n to T will also depend upon the same types of data, with of course T being unknown a priori. Convergence of any algorithmic process to the true maximum of the stochastic maximization problem; that is, any algorthmic process which examines a sequence of points x_i such that $\lim x_i = x^*$ where x^* is

that policy such that $r(x^*,T)$ = max $r(x,T)$, must necessarily have contained in it a technique that guarantees that lim $\bar{y}$ = y and lim T_n = T.

To determine an initial starting point for stochastic maximization through whatever algorithmic process is chosen it appears that it would normally be of interest to compare x^* with that policy x^{**} which is defined by $r(x^{**},0) = \max(x,0)$; that is, a comparison between the stochastic simultaneous equations model with its deterministic counterpart. Solution for x^{**} takes place either through one of the projected gradient approaches, the penalty function approach, Box's Complex Method, or one of the other computationally efficient algorithms.

Also when simulating the model choice of the matrix u of error terms should be determined from the data base used to estimate the parameters b. Specifically the matrix u should to as great an extent as possible be the calculated residuals for the structural equations obtained from the statistical estimation process utilized to determine the specification of the model. This technique implies a fuller use of the prior information available than if the disturbance terms were generated independently of the data base. Since meaningful results of this type of simulation are necessarily large sample properties it will normally be necessary to generate disturbance matrices u from the presumed structural equation disturbance distributions in addition to those obtained from the data base as calculated residuals in the estimation process. This proceeds unless it is possible to generate data for the model within the context of the model. In this latter case estimation of the parameters should be done with respect to the entire set of data used in the simulation.

Solution of equations 3-6 are necessary for evaluation of the objective function at each stage of the simulation procedure given a policy (that is for each specification of u). This is probably best obtained through one of the gradient techniques while choice of a direction and length of movement in terms of the policy vectors x is probably best done through one of the non-gradient approaches. This arises for several reasons. Consider the resetting situation if a gradient approach were utilized for movements in terms of policy. Within the context of stochastic simulation there is initially the prior problem of what, if any, meaning the concept of the gradient can have for a stochastic function. If h(y) is continuous and g(y,x,b,u) is continuous where the distribution functions are continuous in the disturbance terms u and the variables y,x, and u are all defined with connected domains then the risk function

r(x,T) will be continuous and the gradient for any policy specification x will be well defined in the usual manner. However for any T_n, that is for any empirical frequency distribution, the risk function $r(x,T_n)$ will not have gradients that are well defined as the distribution T_n will necessarily be discrete. Since it is the risk function r(x,T) we are ultimately concerned with the existence of gradients for $r(x,T_n)$ is not really of direct interest. What is of direct relevance in any gradient method approach is obtaining approximations of the gradient for a specific policy x, e.g. the gradient of r(x,T), based upon the information available from the various functions $r(x,T_n)$ where both x and T_n are variables. Rephrasing this, any potential gradient method approach which is to be utilized in stochastic simulation is necessarily interested in obtaining good statistics which attempt to measure the gradient at the various policy vectors. To this author's knowledge the estimation of a gradient for either a known or unknown function has not been analysed within the relevant literature and thus pertinent criteria have not been developed for evaluating the worth of alternative statistical techniques.

Since the dimensionality of x is equal to (m)(n) this implies that since the risk function is not known a priori near any policy x that at least (m)(n) evaluations of the risk function near the current policy x need to be evaluated to even consider using a gradient method. That is to approximate the gradient it will prove necessary to estimate the risk function in all of the possible directions of movement, the extent to which a move in a particular direction being taken is determined by the increase in the risk function in this direction. If the number of experiments taken in each direction is small then there will very likely be a significant problem with outliers arising in the sample; outliers carrying a disproportionate weight in evaluating the risk function when the sample in any direction is small. Since the number of directions, equal to Im)(n) in number will normally be quite large the chance of outliers occurring in one or more directions will be significant; thus implying that in small samples our estimation of the gradient can be significantly misleading, especially if the variance of T_n is large for small n. Since the efficiency of any gradient procedure explicitly depends upon correct specification of the gradient at a point this can have serious consequences. A series of movements of a stochastic gradient algorithmic in the proper direction can be totally eliminated by one 'bad' move made by misestimation of the gradient. Also since (m)(n) will probably be a large number a direct search procedure, by moving immediately to another point to be evaluated, can utilize mn evaluations at this new point resulting in $r(x,T_{\cdots})$ which can be

close to r(x,T). This direct movement, involving approximately the same number of calculations, yields an estimate of r(x,T) whereas the gradient approaches might still be considering a direction of movement. Thus it is recommended that the second stage of the stochastic maximization procedure be one of the non-gradient approaches previously mentioned.

As an overall algorithmic stochastic maximization strategy once a model has been specified, estimated, an initial policy specified, and the necessary error terms drawn from the underlying error distributions is first to determine which of the available techniques best serves as the solution technique for equations 3-6. This can be done by specifying a value of S to be achieved in equation 3-4D and seeing which of the alternatives chosen performs best on an average basis. Most likely the best technique will depend both upon the specification of the model and the form of equation 3-4D as well as the value of S desired; however due to the paucity of Monte-Carlo studies in this area it is conceivable that one technique will turn out to be superior in the majority of potential applications. In this regard current research being undertaken by the author examines this point peripherally (see section 4). Once the technique of approximation for the first stage of the stochastic maximization procedure has been determined it is recommended that either the Sequential Simplex technique be applied or the Complex Method where the initial error allowable on y, namely $\bar{y}$ be set rather large and the number of determinations at each policy be rather small (perhaps 10 evaluations of alternative error terms).

In relationship to the sequential simplex technique the error allowable to $\bar{y}$ should be decreased and the sample size should be increased once a terminal simplex is reached. In the original specification of the complex method stopping only occurred when five consecutive equal function evaluations occurred. This clearly is not feasible with an objective function that is a random variable, i.e., $r(x,\bar{y},T_n)$. It is suggested here that both $\bar{y}$ be more closely specified and n increase once five given consecutive function evaluations fall within two standard deviations of the minimal standard error of these five policies. Thus both of these algorithmic processes would become more 'refined' the closer they approached a solution. Also the above suggestions are easily programmed

4. Applications

Currently the author is attempting to find both an efficient technique for solution of a large macroeconomic model in terms of policy

for a well-defined set of policy objectives. Specifically the author, in conjunction with support from the University of Nairobi and the Government of Kenya, will apply the theoretical analysis outlined here in determination of the size, scope, and distribution of fiscal policy to be derived from an already existing macroeconomic model of the Kenyan economy. It is hoped that besides fulfilling these objectives that several of the alternatives suggested in this analysis which seem viable as potential stochastic maximization algorithms for macroeconomic models can be compared through a Monte-Carlo study sometime in the near future.

REFERENCES

1. Beveridge, G.S.G. and Schechter, R.S., Optimization: Theory and Practice. Mcgraw-Hill, New York, 1970.

2. Ferguson, T.S., Mathematical Statistics, Academic Press, New York, 1967.

3. Fisk, P.R., Stochastically Dependent Equations. Harper New York, 1967.

4. Hood, W.C. and Koopmans, T.C. Studies in Econometric Method, Wiley, New York, 1953.

5. International Bank for Reconstraction and Development, The Second Decade; A Basic Economic Report on Kenya, Vol. I. V.

6. Loeve, M.M., Probability Theory, Van Nostrand, 3rd edition, Princeton.

7. Computer Simulation Experiments with Models of Economic Systems, Naylor, T. (ed.) Wiley, New York, 1971.

8. The Design of Computer Simulation Experiments, Naylor, T. (ed.), Duke University Press, Durham, N.C., 1969.

9. Slater C., and Walsham, G. "A General Systems Simulation of the Kenyan Economy, "July, 1974 I.D.S. Working Paper 174.

10. Wald, A., Sequential Analysis, Wiley, 1947, New York.

11. Wald, A., Statistical Decision Functions. Wiley, 1950, New York.

12. Whitacre, R. "Policy Determination Through Simulation of Non-linear Econometric Models." April, 1975, I.D.S. Working Paper 213.