A SYSTEMS APPROACH TO THE CALIBRATION
OF ECONOMY-WIDE MODELS WITH INCOMPLETE DATA

by

Michael J. Hartley

October 1985

Development Research Department
Economics and Research Staff
World Bank

The World Bank does not accept responsibility for the views expressed herein
which are those of the author(s) and should not be attributed to the World
Bank or to its affiliated organizations. The findings, interpretations, and
conclusions are the results of research supported by the Bank; they do not
necessarily represent official policy of the Bank. The designations employed,
the presentation of material, and any maps used in this document are solely
for the convenience of the reader and do not imply the expression of any
commitment, whatsoever on the part of the World Bank or its affiliates concerning
the legal status of any country, territory, city, area, or of its authorities,
concerning the delimitations of its boundaries.
A SYSTEMS APPROACH TO THE CALIBRATION OF ECONOMY-WIDE MODELS

WITH INCOMPLETE DATA

by

Michael J. Hartley*
Development Research Department
World Bank**

December 1984
(revised October 1985)

* The author wishes to express his deep sadness at the loss of a colleague and friend who otherwise would have co-authored this paper. Richard Bellman died on March 19, 1984.

** The World Bank does not accept responsibility for the author's views. These should not be attributed to the World Bank or its affiliated organizations. The findings, interpretations, and conclusions are the results of research supported by the Bank; but do not necessarily represent its official policy.
Abstract

Lawrence Klein once confided to me (circa 1966) that one of his secret dreams was to have a massive analog computer with many control dials—each calibration representing the value of a particular unknown parameter in an economy-wide model. The output was to be a set of television screens in which graphs of the actual and predicted values of each endogenous variable could be instantaneously displayed, corresponding to any configuration of dial-settings. The problem for the investigator is how to manipulate the dials. This requires either: (1) another computer attached to a robot; or (2) software consisting of an optimization algorithm wrapped around a finite set of solution and/or optimization algorithms, with visual display capabilities, as the solution is produced by a single computer; or (3) one Richard Bellman.
Table of Contents / 1

1. Introduction ......................................................... 1

2. Stochastic Dynamic Models and Incomplete Data. .................. 6
   2.1 Maximum Likelihood Estimation ................................. 8
   2.2 Minimum Distance Estimators ................................ 9

3. Calibration of Deterministic Dynamic Structural Models (Stage One) 11
   3.1 Model Specification ........................................... 11
   3.2 The Calibration Problem ....................................... 14
   3.3 The Basic Calibration Algorithm ............................... 15
   3.4 Incomplete Data and Uniqueness ............................... 20
      1. Incomplete Data ............................................ 20
      2. Uniqueness ................................................ 26

4. Calibration and Policy Analysis (Stage Two) ...................... 30
   4.1 The First-Order Approximation Model ......................... 31
   4.2 Stage Two Calibration with Complete Data .................... 33
   4.3 Stage Two Calibration with Incomplete Data .................. 38
   4.4 Direct Extensions ............................................ 45
      1. Higher-Order Approximations ................................ 46
      2. Multi-Period Policies ...................................... 50

5. Model Simplification and Respecification (Stage Three) .......... 53
   5.1 Calibration/Estimation of Missing Data ....................... 54
   5.2 A Priori Parameter Specification ............................. 56
   5.3 Model Taxonomy and Respecification ........................... 56

6. Conclusions, Extensions and Future Research ..................... 61
   6.1 Conclusions .................................................. 61
   6.2 Extensions and Future Research ................................ 68

1/ I would like to thank Miriam Bailey for her excellent word-processing.
1. Introduction

It is undoubtedly the case that construction of macroeconomic models for Less Developed Countries (LDCs) will remain somewhat of an art for some time to come—requiring judicious use of simplifying assumptions on the part of the investigator within a general taxonomy of model specifications. This paper addresses an important aspect of the problem—unlike macro models for developed countries, the available time series data for LDCs are invariably of limited duration and are incomplete. The investigator confronts data covering different time spans, sometimes with gaps; at different levels of aggregation (usually improving in detail over time); and supplemented by data obtained at irregular intervals from censuses, surveys and special studies. As data are scarce, our interest is to develop econometric methods which exploit the information present in all available observations—both complete and incomplete.

There is an extensive, now standard, econometric literature on full information systems methods for estimation of the parameters in stochastic macro models. These are usually specified as a dynamic nonlinear system of simultaneous (algebraic) equations, with normal random errors introduced additively into all behavioral equations. However, in general cases and where certain of the endogenous variables are missing, such methods will often require multiple numerical integrations—beyond the present state of computer technology—to extract the information present in the incomplete observations. The investigator must therefore "trim" the sample to include only complete observations. Unfortunately, with most LDCs, the resulting structure of estimable stochastic macro-econometric models will often be too simple to analyze most policy questions of interest.
At the other extreme, the investigator may formulate a much more realistic economy-wide model, disaggregated to encompass the various productive sectors, institutions, factors of production, etc., of importance. Such models, at present, are almost always deterministic in structure, and represent and economic theorist's macroeconomic analogue of the familiar Walrasian (1874) system of general microeconomic equilibrium (Arrow and Debreu (1954)). For a given specification of the behavioral functions and identities, given parameter values and given data on the predetermined variables, a solution for the endogenous variables may be computed. Such Computable General Equilibrium (CGE) models—see, e.g., Johanson (1960), Dervis, de Melo and Robinson (1982) or Scarf and Shoven (1984)—have been widely applied in LDCs to analyze, inter alia, counter-factual policy simulations. However, at present, the parameters in CGEs are calibrated by informal methods; and, in most instances, utilize only the data for a single recent "benchmark" year (see Mansur and Whalley (1984) or Hartley (1984a) for a discussion of "standard-practice" calibration methods). As the predictions from such CGEs are demonstrably sensitive to alternative calibrations, some concern has been expressed as to the utility of such exercises (Lau (1984)).

Given the immediate need for model-based policy analysis in LDCs, the investigator faces a genuine dilemma. The arbitrary configuration of available data precludes use of standard econometric software packages for the purposes at hand; and the standard practice calibration methods for CGEs fail

---

1/ See, however, Jorgenson (1981), for a specific example of a stochastic general equilibrium model—though estimated partially via separate treatment of the household and firm sectors, utilizing complete data samples.
to exploit all available information. To avoid this impasse, we are therefore led to explore general "systems methods" for the calibration of structural deterministic models, as these not only accommodate incomplete data more readily, but also are more easily generalized to extensions in the form of the maintained hypothesis—see Section 6.

Following the methods of quasi-linearization and identification of more general systems—see, e.g., Bellman (1956), Bellman, Kagiwada and Kalaba (1965) or Bellman and Roth (1978, 1979, 1981), we formulate a basic calibration procedure which minimizes a quadratic loss function, involving the discrepancies between the solution (simulation) path of the endogenous variables (relative to any choice of parameter vector) and the (observed) actual values, over the sample period (Hartley (1972)). Iterative minimization of the loss function requires only that a solution be computable for known parameter values, as is characteristic of CGEs. However, since the configuration of available data may be quite arbitrary, the investigator often will not be able to determine a priori whether or not a unique calibration of the model can be achieved. We employ a Singular Value Decomposition (SVD) algorithm, in part because it permits us to evaluate the numerical rank of the appropriate Jacobian matrix—see Golub (1968). This will inform the investigator, a posteriori, whether or not, upon convergence, the present model specification can be uniquely calibrated on the basis of the present use of available data. We refer to these as Stage One calibration methods.

Given the incomplete data typical of LDCs, however, a non-unique calibration will often result from Stage One. If so, this implies that we have located one member of an equivalence class of parameter values, each of which results in the same (minimum) value of the loss function; but may produce different predictions of the consequences of a given policy of
interest. This raises, first, the question of whether one may exploit any additional information from within the existing data set to resolve the parameter indeterminacy. Second, it raises the question of whether the available data are sufficient to obtain a unique calibration of the model at all—at least as presently specified. If the latter is the case, the investigator must then explore various model simplifications. These include more parsimonious parametrizations of existing model functions, and restriction of the "scope" of the model—e.g., to define specific processes generating certain of the exogenous variables, or employ extraneous "priors" to fix certain parameter values. Alternatively, a simpler structural model may be employed.

Our approach to resolve the parameter indeterminacy, without respecification of the model structure, is to introduce a further criterion. E.g., suppose one is interested in the performance of the model in predicting the consequences of a particular vector of predetermined variables—the policy assumptions and other exogenous variables of interest. Our Stage One calibration methods exploit only the information resulting from numerically optimizing the parameter values—via calculation of the sequence of model solution values relative to their observed counterparts. However, as the "real world"—and hence the structural model—is presumed to be highly non-linear, not only the solution values, but, in general, also all partial derivatives of the solution vector (with respect to the parameters and/or the predetermined (state) variables) will depend upon the underlying (primal) specification of the structural model. 1/

1/ Within finite limits.
Our approach in Stage Two, therefore, is to adopt a second quadratic loss function to be minimized, while restricting the parameter vector to lie within the equivalence class defined by Stage One. The former is based upon the difference between the corresponding coefficients in: (a) an unrestricted polynomial approximation to the reduced form of the model, and (b) a Taylor-series approximation to the restricted reduced form, obtained from computable solutions to the structural model. The expansion is made about the particular vector of predetermined variables of interest. The order of the approximations, in principle, may be varied with each endogenous variable, permitting the use of so-called "step-up/step-down" regression methods. Under present computer technology, however, except in models of small dimension, a first-order approximation model will probably have to suffice. We also consider generalization to the case where the investigator wishes to calibrate the model via approximations around a multi-period policy sequence.\(^1\)

Incomplete data considerations will further complicate these (Stage Two) procedures. Upon completion of any (expansion/contraction) phase of Stage Two, numerical SVD methods again may be employed to determine the uniqueness of the resulting calibration.\(^2\) If not unique, and if no further extensions to Stage Two are feasible or desirable, then simplifying assumptions in Stage Three must be invoked.

While the procedures in Stages One and Two are systematic (relative to a given model), if model respecifications are required in Stage Three, then Stage One/Two will have to be repeated until a unique calibration of a

---

\(^1\) Such expansions are feasible, but may be costly at present.

\(^2\) It should be stressed that the accuracy of such calculations will depend upon the word-size of the computer employed.
specific model obtains. The iterative interaction of the investigator with the model specification (e.g., in exploring simplifying/more complex assumptions in Stage Three, relative to the available configuration of data) is more difficult to systematize, and involves an inevitable subjective aspect—hence, the "art" of model building for LDCs.

In Section 2 we consider the consequences of incomplete data on the endogenous variables within the standard stochastic simultaneous equations model. In Section 3 we introduce a deterministic structural analogue and develop our basic Stage One calibration methods for both complete and incomplete data bases. In Section 4 we introduce our basic Stage Two methods and direct extensions. We describe model simplification and respecification procedures in Section 5; while reserving some concluding remarks and a discussion of possible extensions for Section 6.

2. Stochastic Dynamic Models and Incomplete Data

To permit comparison with a deterministic structural analogue, we begin by briefly considering the standard stochastic econometric model, represented as a nonlinear system of simultaneous difference equations (Haavelmo (1943)). For expository convenience, we assume all identities can be eliminated by substitution, and consider the first-order stochastic system,

\[ f(y_t, y_{t-1}, x_t, \theta) = \epsilon_t, \]  

(2.1)
where \( f \) is presumed to be an \( N \)-vector of known nonlinear behavioral functions; \( \gamma_t \) and \( \gamma_{t-1} \) are \( N \)-vectors of current and lagged endogenous variables; \( \pi_t \) is a \( K \)-vector of exogenous variables, respectively; and \( \theta \) is an \( M \)-vector of unknown structural parameters.\(^1\) The \( N \)-vector, \( \varepsilon_t \), is an unobserved random error with p.d.f., \( p(\varepsilon_t) \). The particular assumption, in which \( \varepsilon_t \) is taken to be an independent random drawing from the normal p.d.f.,

\[
p(\varepsilon_t) = N(0, \Omega),
\]

with zero mean and unknown covariance matrix, \( \Omega \), is often invoked. However, for a particular, but unknown drawing, \( \varepsilon_t \), from any choice of \( p(\varepsilon_t) \), the current actual observations, \( \gamma_t \), are assumed to be generated as a solution to \( (2.1) \), for given values of the predetermined variables, \( \gamma_{t-1} \) and \( \pi_t \). The initial value, \( \gamma_0 \), is assumed constant.

There is an extensive literature on the problem of estimating \( \theta \) and \( \Omega \) from a complete sample of data: \( \{(\gamma_t, \pi_t): t = 1, \ldots, T\} \) and the initial value \( \gamma_0 \) (see, e.g., Hendry (1976) for a brilliant synthesis, and Amemiya (1983) and Hausman (1983) for excellent surveys). Our present concern, however,

\(^1\) This is a restrictive assumption (see Hartley (1983a, 1984a)), but greatly facilitates exposition. The basic conclusions, however, remain--regardless of the need for a distinction between stochastic behavioral equations and nonlinear identities. The first-order vector autoregression is not restrictive. A system of higher order may be represented as a first-order model by simply augmenting the system with a set of dynamic identities (see Fisher (1965)). Also, Adomian (1983, 1984b) has considered more general forms of stochastic systems.

\(^2\) The exogenous vector, \( \pi_t \), may include its own lagged values (Koyck (1958), Dhrymes, (1971), Pagano and Hartley (1981)). More generally, we have vector ARIMA processes, Box and Jenkins (1971)---see Stage Three.
is how such methods can be adapted, in general, to accommodate missing data on
the endogenous variables, \( \{ Y_t \} \), given \( \{ x_t \} \) and \( \chi_0 \). \(^{1/}\) We consider, here,
only full information methods, appropriate to a CGE specification, in order to
incorporate the various types of cross-equation parameter restrictions that
are at the essence of economic (constrained optimizing) behavior. \(^{2/}\)

2.1 Maximum Likelihood Estimation

From (2.1) and (2.2), we may define the conditional p.d.f. of
\( Y_t \), given \( Y_{t-1} \) (and \( x_t \)), as:

\[
\begin{align*}
    p^*(Y_t | Y_{t-1}) &= p(\xi(Y_t, Y_{t-1}, x_t, \theta)) \cdot J_t,
\end{align*}
\]

where \( J_t = \frac{\partial f}{\partial y_t} \) is the relevant NxN Jacobian matrix. With complete data,
we maximize the likelihood, \( L(\theta, \Omega) = \Pi_{t=1}^{T} p^*(Y_t | Y_{t-1}) \); or its logarithm--
producing an exhaustive "monotonic iterative two-step" method of: \( \theta | \Omega \) and
\( \Omega | \theta \), in the special case of a normal p.d.f. (Oberhofer and Kmenta (1974)).

Suppose now, upon suitable reordering, that \( Y_t^i = [Y_{1t}^i, Y_{2t}^i] \),
where \( Y_{1t} \) is observed, but \( Y_{2t} \) is missing--the partitioning varying for
each particular \( t = 1, \ldots, T \). Assuming that \( \{ x_t \} \) and \( \chi_0 \) are complete, the

---

\(^{1/}\) Missing elements of \( x_t \) or \( \chi_0 \) may be treated as additional parameters
analogously to the deterministic model--see Section 3.4.

\(^{2/}\) These occur, e.g., in complete systems of demand functions for goods and
services, factors of production, etc.--whether treated explicitly via
Deaton and Muellbauer (1980), etc.) or implicitly (Hartley (1981a),
(1983b)). Here, the difference is whether one specifies (1) a demand
system, derived analytically from the deterministic form of the primal (or
dual) problem, and then (if desired) introduces stochastic features or (2)
the primal (utility/production, etc.) function, solving the primal
(constrained optimization) problem by computer. The latter approach, we
claim, is more general--see Section 6.
likelihood function for the observed sample, \( \{Y_{1t}\} \), in general requires "integrating out" the effect of the missing \( \{Y_{2t}\} \) sequentially, as in

\[
\tilde{L}(\theta, \Omega) = \int \ldots \int \left[ \prod_{t=1}^{T} p_{T}^{*}(Y_{t} | Y_{t-1}) \right] dY_{2t}
\]

\[
= \int_{T}^{T} p_{T}^{*}(Y_{T} | Y_{T-1}) \cdot \int_{T}^{T} p_{T-1}^{*}(Y_{T-1} | Y_{T-2}) \cdot \ldots \int_{T}^{T} p_{2}^{*}(Y_{2} | Y_{1}) \cdot p_{1}^{*}(Y_{1} | Y_{0}) \cdot dY_{21} \ldots dY_{2T-1} dY_{2T}
\]

\[
\equiv \prod_{t=1}^{T} p_{t}(Y_{1t} | Y_{1t-1}).
\]

In most LDC applications, however, the order of numerical integrations (Haber (1970)) required to evaluate \( \tilde{L} \) of (2.4) in general will be well beyond present computer capabilities--see, e.g., Quandt (1983); and the information within the incomplete observations must be discarded.\(^{1/2}\)

2.2. Minimum Distance Estimators

Minimum distance estimators include nonlinear three stage least squares (Amemiya (1977)) and various "instrumental-variables" methods (e.g., Hausman (1975), Jorgenson and Laffont (1974)) and Malinvaud's (1966) method). These may be obtained via minimization of a loss function of the form,

---

1/ Numerical integration is involved because \( f \) is nonlinear in \( Y_{t} \). Consequently, \( J_{t} \) is a function of \( Y_{t} \), and \( p_{t}^{*} \) is a non-normal p.d.f., even with (2.2). See, however, Section 3.3 for a linearized model, obtained as a limiting version of (3.15), given (2.2).

2/ So-called "Bootstrap Monte Carlo" methods may also be employed in general (Hartley (1984a)), but the computational cost, at present, is prohibitive--except on small problems or with with parsimonious treatment of the model's stochastic features. Further, as \( Y_{t} \) is non-normal, use of the E-M algorithm (Dempster, Laird and Rubin (1977)) to obtain maximum likelihood estimates again requires multiple numerical integration--see below.
\[ Q(\theta, \Omega) = \frac{1}{2} \sum_{t=1}^{T} f(y_t, y_{t-1}, x_t, \theta)'. G_t . f(y_t, y_{t-1}, x_t, \theta), \quad (2.5) \]

where different members of the minimum distance class are associated with different choices of the matrices, \( G_t \) -- each a function of \( \Omega \) in the specific case (2.2). If, however, the \( \{y_{2t}\} \) are not observed, two approaches may be proposed: Either (1) minimize the expected loss, integrating out the missing \( y_{2t}'s \), or (2) replace the missing \( y_{2t} \) subvectors (within \( f_t \) and \( f_{t-1} \)) by their conditional expectations, given \( y_{1t} \) and \( y_{1t-1} \).

Approach (1) suggests minimization of the expected loss, \( 1/ \)

\[ Q(\theta, \Omega) = \frac{1}{2} \int \ldots \int \sum_{t=1}^{T} f(y_t, y_{t-1}, x_t, \theta)'. G_t . f(y_t, y_{t-1}, x_t, \theta). \rho_t(y_t | y_{t-1})dy_{2t}, \quad (2.6) \]

where the multiple numerical integrations must be performed sequentially, as in (2.4). Approach (2) requires that we also evaluate the conditional expectations of \( y_{2t} \), given both \( y_{1t} \) and \( y_{1t-1} \). Both approaches involve even more complex multiple numerical integrations than maximum likelihood estimators (MLEs), unless (with either) a specific model is adopted for mathematical convenience, i.e., in which analytical simplifications can be obtained.

The above discussion indicates that, in general, incomplete observations on the endogenous variables are extremely difficult, if not impossible, to accommodate within full-information econometric software packages for general stochastic dynamic nonlinear simultaneous equations.

---

1/ It should be noted that iterative linearization of \( f_t = 0 \) around \( y_t^* \) (see Section 3) will eliminate the \( t^{th} \) term from \( Q \).
models. In the balance of this paper, we shall therefore consider a deterministic analogue, which includes the iteratively-linearized, normal stochastic model as a special case. We then develop calibration procedures which appear tractable for incomplete data, and avoid the problems of high-order multiple numerical integrations associated with general choices of the p.d.f., \( p(\varepsilon_t) \).  

3. Calibration of Deterministic Dynamic Structural Models (Stage One)

3.1 Model Specification

We now consider a deterministic analogue to the structural model, (2.1), given by the system of \( N \) implicit behavioral equations and identities,

\[
\mathcal{F}(\mathbf{Y}_t, \mathbf{Y}_{t-1}, \mathbf{X}_t, \theta) = 0 ,
\]

(3.1)

where \( \mathbf{Y}_t \) now denotes an \( N \)-vector of solution values for the endogenous variables represented by the unknown function, \( \mathbf{X}_t \).

Thus, for a specified structural model, \( \mathcal{F} \), given the actual values of the predetermined variables, \( \mathbf{Y}_{t-1} \) and \( \mathbf{X}_t \), each choice of the parameter

---

1/ See, e.g., Prucha and Kelejian (1984) for the case of the multivariate-t distribution for \( \varepsilon_t \), in which a taxonomy of estimators analogous to Hendry (1976) is developed. Also, Gourieroux, Monfort and Trognon (1984) have recently developed a general treatment of "pseudo" MLEs.

2/ This assumes a suitable specification for \( \mathcal{F}_t \), such that \( \mathbf{J}_t \) is nonsingular.
vector, \( \theta \), will result in a particular solution vector, \( \mathbf{y}_t^* \). However, except by coincidence or design, the model solutions, \( \{\mathbf{y}_t^*\} \), will not exactly reproduce the actual values, \( \{\mathbf{y}_t\} \), and we have the deterministic/stochastic reduced form model,

\[
\mathbf{y}_t = \mathbf{y}_t^* + \mathbf{e}_t = \mathbf{g}(\mathbf{y}_{t-1}, \mathbf{x}_t, \theta) + \epsilon_t, \tag{3.3}
\]

where \( \mathbf{e}_t \) denotes a vector of discrepancies for \( t = 1, \ldots, T \). In general, as \( f \) is nonlinear in \( \mathbf{y}_t^* \), it will not be possible to obtain a closed-form expression for \( \mathbf{g} \) in (3.2). Instead, we require only that \( f \) be chosen so that \( \mathbf{y}_t^* \) is computable by numerical methods for suitable values of \( \theta, \mathbf{x}_t \) and \( \mathbf{y}_{t-1} \) (or \( \mathbf{y}_{t-1}^* \) and \( \mathbf{y}_0 \)).

---

1/ An alternative to the model, (3.1), is the dynamic specification,

\[
f(\mathbf{y}_t^*, \mathbf{y}_{t-1}^*, \mathbf{x}_t, \theta) = 0, \tag{3.1}'
\]

where \( \mathbf{y}_0 = \mathbf{y}_0^* \). In this case, the current solution value,

\[
\mathbf{y}_t^* = \mathbf{g}(\mathbf{y}_{t-1}^*, \mathbf{x}_t, \theta), \tag{3.2}'
\]

depends upon the previous solution value, \( \mathbf{y}_{t-1}^* \), as opposed to the actual value, \( \mathbf{y}_{t-1} \). Here, \( \{\mathbf{y}_t^*\} \) denotes the dynamic simulation path of the model from initial conditions. The deterministic dynamic version is considered in more detail elsewhere (Hartley (1984a)). See also Hartley (1970, 1972) or Klein (1971) for one form of a stochastic analogue, and Theil and Boot (1962) for a discussion of the final form of a linear dynamic stochastic model. In the former, the "dynamic simulation path" is the most general (i.e., "final") form for computable solutions to dynamic models. Normally, however, use of (3.1) and (3.2) is more informative.
The deterministic structural model, (3.1) and (3.2), may be simply regarded as an approximation to the "real world"--the latter being treated as either deterministic or stochastic.\(^1\) In the former case, the discrepancies, \(\{e_t\}\), are then a set of constant functions of \(\theta\), which enter the investigator's loss function. Under a stochastic interpretation for the actual values, \(x_t\), the discrepancies are random variables, and we may introduce (say) the customary specific assumption of a normal p.d.f.,

\[ e_t \sim N(0, \Sigma), \quad (3.4) \]

where \(\Sigma\) is an unknown covariance matrix.\(^2\) In this particular case, the model (3.3) differs from (2.1) in that (normal) random errors are introduced into the reduced form, (3.3), instead of the structural form, (2.1).\(^3\) This results in a system of nonlinear "regression equations", with cross-equation parameter restrictions on \(\theta\) in both \(f\) and \(g\)--except that the "regression functions", \(g\), are now only computable from a known \(f\).\(^4\) The parameters, \(\theta\) and \(\Sigma\), must then be viewed as unknown "true values," relative to the

---

\(^1\) We therefore permit a deterministic structural model to serve as an approximation to a stochastic "real world."

\(^2\) In addition, stochastic features may be introduced through \(x_t\) (Tobin (1965), Hartley (1983b)).

\(^3\) In many cases it may be important to treat the \(e_t\) as stochastic, due to random errors of measurement (Klepper and Leamer (1984)).

\(^4\) In more general models \(f\) need only be computable--say, as the solution to an agent's economic (constrained optimization) problem (Hartley (1981a, 1983b)).
stochastic reduced-form maintained hypothesis, (3.1) and (3.3).1/2/  

3.2 The Calibration Problem  

We begin with the case where complete data are available for  
\{(y_t,x_t):t=1,...,T\} and the initial value, y_0. Given any feasible value for  
θ, we may then calculate the complete model solution path, \{y_t^\ast\}. The 
calibration problem is to determine a unique value of θ for a given model,  
\(f\), such that the solution path, \{y_t^\ast\}, approximates the actual data,  
\{y_t\}, as "closely" as possible.  

To this end, we adopt (say) a quadratic loss function, \(Q(\theta)\), involving the discrepancies, \{e_t\}, defined by  

\[
Q(\theta) = \frac{1}{2} \sum_{t=1}^{T} e_t' \cdot W_t \cdot e_t = \frac{1}{2} \sum_{t=1}^{T} (y_t^\ast - y_t)' \cdot W_t \cdot (y_t^\ast - y_t)
\]

(3.5)  

where \(W_t\) is a suitably-chosen weighting matrix. If the discrepancies are  
constant, then \(W_t\) is at the discretion of the investigator. Some attractive 
choices are: (1) \(W_t = I\), a least-squares criterion for standardized data;  
use of the inverse of the diagonal matrix of "sample" variances, (2)  
\(W_t = \text{Diag}(s_{ii}^{-1})\); or the "sample" covariance matrix, (3) \(W_t = (s_{ij})^{-1}\),  
where \(s_{ij} = \frac{1}{T} \sum_{t=1}^{T} (y_{ti} - \bar{y}_i) \cdot (y_{tj} - \bar{y}_j)\). Also, we have the possible matrices  
when: (4) \(W_t\) is chosen by the investigator, jointly weighting both endogenous  

1/ We distinguish between estimation of the "true" parameters of a stochastic 
model and their calibration in a deterministic model.  

2/ We exhibit a correspondence with the specific stochastic structural form,  
(2.1) and (2.2), below.
variables, i, and the time periods, t. However, if $e_t$ is assumed to be
stochastic, with the normal p.d.f., (3.4), then the choice: (5) $W_t = \varepsilon^{-1}$,
where $\varepsilon = \frac{1}{T} \sum_{t} (y_t - y_t^*)'(y_t - y_t^*)$, will result in convergence to a "maximum
likelihood" estimate of $\theta$, relative to the stochastic reduced form
specification, (3.3). Finally, iterative linearization of $f$ around $y_t^*$ will
result in the choice: (6) $W_t = [J_t^{-1}]^{-1}$ --see below-- where an MLE for $\theta$
is now obtained relative to the model (2.1) under the particular choice for
$p(e_t)$ in (2.2). Finally, under certain types of moving-average error
processes, generated by "errors of observation" in the \{y_t\}, the parameters,
$\theta^{(n)}$, converge to an MLE for $\theta$ under the dynamic stochastic version of
(3.1)' and (3.2)' (Hartley (1970, 1972), Klein (1971)).

Regardless of the motivation, the essence of the calibration problem
is that one must know $\theta$ to calculate the forward solution, \{y_t^*\}; which,
in turn, is required to determine the value of the loss function, $Q(\theta)$,
which is then to be minimized with respect to $\theta$. This clearly suggests an
iterative approach.

3.3 The Basic Calibration Algorithm

Following the methods of quasi-linearization and identification
(Bellman (1956c), Bellman, Kagiwada and Kalaba (1965) or Bellman and Roth
(1978, 1979, 1981)), we fix $W_t$ (if dependent on $\theta$) and minimize
$Q(\theta)$ with respect to $\theta$ in $y_t^*$. This defines the first-order conditions,

$$ q(\theta) = \frac{\partial Q(\theta)}{\partial \theta} = \sum_{t=1}^{T} J'_t \cdot W_t \cdot (y_t - y_t^*) = 0 \quad (3.6) $$

where $J_t = (\partial y_t^*/\partial \theta)$ is an NxM Jacobian matrix. Written more compactly, we
seek a solution to the general nonlinear algebraic system of the form,
\[ q(\theta) = J(\theta)' \cdot W(\theta) \cdot (y - y^*(\theta)) = 0, \quad (3.7) \]

where we now employ the "stacked" definitions,

\[ J(\theta)' = J' = [J_1' \ J_2' \ldots \ J_M']_{M \times N \cdot T}, \quad (3.8a) \]

\[ W(\theta) = W = \text{Diag}[W_1, W_2, \ldots, W_{T}']_{T \times N \cdot T}, \quad (3.8b) \]

\[ y' = [y_1', y_2', \ldots, y_{T'}']_{1 \times N \cdot T}, \quad (3.8c) \]

and

\[ y^*(\theta)' = y^* = [y_1^*, y_2^*, \ldots, y_{T'}^*]_{1 \times N \cdot T}. \quad (3.8d) \]

Our simple algorithm involves: (1) the iterative replacement of the unknown function, \( y^*(\theta) \), in (3.7) by a linear Taylor series approximation, evaluated at the current iteration's parameter value; and then (2) solving the resulting linear system in \( \theta \) for the next iteration's value.

Let \( \theta^{(0)} \) denote an initial feasible parameter vector. Let \( n = 1, 2, \ldots \) be an iteration index; and for any function, \( u(\theta) \), let \( u^{(n)} = u(\theta^{(n)}) \). The linear approximation to \( y^*(\theta) \) is then given by

\[ y^*(\theta) \approx y^*(n) + J(n) \cdot (\theta - \theta^{(n)}). \quad (3.9) \]

Insertion of the r.h.s. of (3.9) for \( y^*(\theta) \), while holding fixed \( J(n) \) and \( W^{(n)} \), leads to the Modified Gauss-Newton algorithm (Hartley (1961)),

\[ \theta^{(n+1)} = \theta^{(n)} + \lambda(n) \cdot [J(n)' \cdot W^{(n)} \cdot J(n)]^{-1} \cdot [J(n)' \cdot W^{(n)} \cdot (y - y^*(n))], \quad (3.10) \]
where $\lambda^{(n)}$ is a scalar step-size parameter. Normally, $\lambda^{(n)} = 1$. However, in some cases $\lambda^{(n)} > 1$ is warranted; in others, particularly if $\theta^{(n+1)}$ with $\lambda^{(n)} = 1$ is infeasible, successive "shrinkages" in $\lambda^{(n)}$ until feasibility obtains and $Q^{(n+1)} \leq Q^{(n)}$ are required. 1/

To implement the algorithm, (3.10), it will be noted that within each iteration, $n$, we require calculation of the solution vectors,

$$\tilde{y}_t^{(n)} = g(y_{t-1}^{(n)}, x_t, \theta^{(n)})$$

(3.11)

for each sample period, $t = 1, 2, ..., T$; as well as the partial derivatives in the NTxM Jacobian matrix,

$$J^{(n)} = \left( \frac{\partial y^*_t}{\partial \theta_m} \right) = \left( \frac{\partial g_i(y_{t-1}^{(n)}, x_t, \theta^{(n)})}{\partial \theta_m} \right).$$

(3.12)

Insofar as the function, $g$, on grounds of generality, is only assumed to be computable, the elements of the Jacobian, $J^{(n)}$, must be evaluated by numerical differentiation—requiring the successive perturbation of each element, $\theta_m$, of $\theta^{(n)}$, by a small $\delta > 0$, see, e.g., Goldfeld and Quandt (1972). Computationally, this adds a further $M$ "passes" through the system—one for each element of $\theta^{(n)}$.

1/ Certain parameter vectors, $\theta$, may be infeasible, in that $f(y_t^{*}, y_{t-1}^{(n)}, x_t, \theta) = 0$ has no solution for some $t = 1, ..., T$. See Berndt, Hall, Hall, and Hausman (1974) for a more comprehensive discussion of such methods. See also Bellman (1979).
The heart of the basic algorithm thus involves repeated solution of the equation system, (3.1), for \( y_t^* \). It consists of: (1) a sequence of \((M+1)\cdot T\) "inner" solutions to obtain \( \{y_t^*(n)\} \), and its first partial derivatives, \( j^{(n)} \), for given \( \theta^{(n)} \); embedded within (2) an "outer" minimization algorithm, (3.10), for \( Q^{(n)} \) to update \( \theta^{(n)} \) to \( \theta^{(n+1)} \). The calibration algorithm then iterates between the inner and outer algorithms so that \( Q^{(n)} \) is always reduced. As the sequence \( \{Q^{(n)}\} \) is monotone decreasing and uniformly bounded from below by zero, convergence of the sequence, \( \{\theta^{(n)}\} \), to a limit point, \( \theta^{[0]} \), is assured.\(^1\) The structural model is then calibrated at

\[
\theta^{[0]} = \lim_{n \to \infty} \theta^{(n)}. \quad (3.13)
\]

If, in addition, the reduced form discrepancies, \( \{e_t\} \), in (3.3) are treated as stochastic, as in (3.4), then we may employ

\[
\text{Cov}(\theta^{[0]}) = \lim_{n \to \infty} [j^{(n)}]' \cdot \omega^{(n)} \cdot j^{(n)} - 1 = \Sigma^{[0]} \quad (3.14)
\]

as the (asymptotic) covariance matrix of \( \theta^{[0]} \), permitting "statistical" tests of hypotheses on \( \theta \).

A similar limiting result holds under the model, (2.1) and (2.2). Previous discussion of this stochastic model revealed the problems of multiple numerical integrations required to utilize incomplete data because a realistic model specification mandates that \( f \) is highly nonlinear in \( \chi_t \). This suggests the strategy of iteratively-linearizing \( f \) around a value, \( \chi_t^* \),

\(^1\) To a given level of tolerance.
such that (3.1) holds for the parameter vector, \( \theta^{(n)} \), i.e.,
\[
\bar{f}(\bar{y}_t^{(n)}, \bar{y}_{t-1}, \bar{x}_t, \theta^{(n)}) = E \varepsilon_t = 0.
\]
The linearized version of (2.1) is then:
\[
\bar{f}(\bar{y}_t^{(n)}, \bar{y}_{t-1}, \bar{x}_t, \theta^{(n)}) \approx \bar{J}_t^{(n)} (\bar{y}_t^{(n)} - \bar{y}_t) = \varepsilon_t,
\]
so that \( \bar{y}_t \) is approximately distributed as \( N(\bar{y}_t^{(n)}, \bar{J}_t^{(n)-1} \Omega \bar{J}_t^{(n)-1}) \), under normality. The MLE of \( \theta \) and \( \Omega \) may then be calculated iteratively by the basic calibration algorithm using the weighting matrix,
\[
W_t^{(n)} = [\hat{\Omega}^{(n)}]^{-1} \cdot [\hat{J}_t^{(n)}]^{-1}, \text{ with } \hat{\Omega}^{(n)} = \frac{1}{T} \sum_{t=1}^{T} \hat{f}_t^{(n)} \cdot \hat{f}_t^{(n)'}
\]
and
\[
\hat{J}_t^{(n)} = \frac{\partial \hat{f}_t^{(n)}}{\partial \bar{y}_t}.
\]
Thus, our calibration algorithm may be adapted (by adjusting \( W_t^{(n)} \)) to accommodate the special case of the stochastic structural model under "rational expectations", via iterative-linearization of the structural equations around the prevailing solution values, \( \{ \bar{y}_t^{(n)} \} \).

However, if all "disturbances" are identically zero, we have simply produced a (deterministic) version of (3.1), with an alternative weighting matrix.

Such a choice, i.e., (6), will be "efficient", in the statistical sense, only if the stochastic special case, (2.1) and (2.2), maintained to be "true", is, in fact, true. This is clearly restrictive, and suggests we entertain a variety of possible choices for \( W_t^{(n)} \).

The general calibration procedure thus places a clear premium on utilizing rapidly convergent algorithms to solve the equation system, (3.1)--see Dennis and Schnabel (1983); and suggests the need to introduce a taxonomy of admissible specifications for \( f \) to exploit its structure. For models of the CGE class, various (sparse matrix) algorithms have already been proposed--

\[ \frac{1}{t} \text{ In the absence of knowledge of } \varepsilon_t, \text{ its expectation is employed (Lucas and Sargent (1981)) to iteratively determine the point of expansion, } \bar{y}_t^{(n)}, \text{ given } \theta^{(n)}. \]
see, e.g., Scarf (1973), Drud (1977/78) and Nepomiastchy and Ravelli (1977/78). Finally, we may employ the observed vector, $\mathbf{X}_t$, as an initial value for within-sample calculations, which should place us in the neighborhood of a solution, $\mathbf{y}_t^*$.  

3.4 Incomplete Data and Uniqueness

In most LDC applications, the data set available to the investigator is not complete. We now consider how to exploit the information contained within the incomplete observations by modifying the basic calibration algorithm.

3.4.1 Incomplete Data

As with the stochastic model, we begin with the case where the matrix of exogenous variables, $\mathbf{X} = (x_{tk})$, and initial endogenous vector, $\mathbf{y}_0$, are complete, but now permit an arbitrary configuration of missing data within the matrix of endogenous variables, $\mathbf{Y} = (y_{ti})$. Let $\mathbf{D} = (d_{ti})$ denote a $\mathbf{T} \times \mathbf{N}$ indicator matrix, where $\mathbf{D}$ indicates the presence or absence of data in the matrix $\mathbf{Y}$.

---

1/ This is a most important research area with great potential payoffs. The idea is to take advantage of the configuration of zeros (i.e., "sparseness") in the Jacobian matrix, $\mathbf{J} = \left( \frac{\partial \mathbf{f}}{\partial \mathbf{y}_k^*} \right)$, and elsewhere. At present, initial solutions to CGEs of dimension, $N = 1000$, are being obtained on a CDC 176 mainframe computer in about two seconds using sparse matrix methods. See also Drud (1985) and the discussion in sections 5 and 6.

2/ The utility of this choice will vary with the realism of the investigator's choice of $\mathbf{f}$, and the reliability of the data (Klepper and Leamer (1984)).

3/ See also Bellman and Roth (1979).
d_{ti} = \begin{cases} 1, & \text{if } y_{ti} \text{ is observed} \\ 0, & \text{if } y_{ti} \text{ is missing} \end{cases} \quad (3.16)

A missing $y_{ti}$-value precludes calculation of the element, $e_{ti}$, in the discrepancy vector, $e_t$. In addition, if $y_{ti}$ is among the model's lagged endogenous variables and is missing, it will not be possible to calculate the entire solution vector, $\mathbf{y}^*_t$, in the following period; so that $e_{t+1}$ is also undefined.\footnote{Note that, if the $e_{ti}$ are stochastic and normal, as in (3.4), then (3.17) replaces each $y_{ti}$ by its conditional expectation, given $y_{t-1}$ and $x_t$. Thus we obtain the Expectation-Maximization (E-M) algorithm (Hartley (1958), Orchard and Woodbury (1972) or, even more generally, Dempster, Laird and Rubin (1977)). See also Hartley (1977a, 1977b). However, this interpretation is problematic with non-normal $e_{ti}$'s, where the expectations will generally require iterative multiple numerical integrations, which we avoid in (3.17) for any well-behaved choice of $f$.}

Finally, missing $y_{ti}$'s may affect the definition of the weighting matrices, $W_t$.

In this situation we modify the basic calibration algorithm by introducing, within the $n$th iteration, a simple missing data updating condition,\footnote{Not all elements of $\mathbf{y}_{t-1}$ need be employed as lagged endogenous variables within a particular specification of $f$ in (3.1).}

\[ y_{ti} = y^n_{ti}, \text{ if } d_{ti} = 0, \quad (3.17) \]

employed sequentially for $t=1,...,T$, given $y_0$. Thus (3.17) iteratively replaces every missing $y_{ti}$ by its corresponding model solution value, relative to the prevailing calibration, $y^n_0$. This not only sets the discrepancy $e_t = y_{ti} - y^n_{ti}$ to zero -- thereby eliminating its contribution to the loss function, $Q(n)$; but also permits calculation of a solution, $\mathbf{y}^*_t$, in the next period, by replacing all missing lagged

\[ y_{ti} = y^n_{ti}, \quad (3.17) \]
endogenous variables by their corresponding previous period's solution
values. As \( \mathbf{y}_0 \) is assumed complete, using (3.17), we may therefore calculate
the entire solution sequence, \[ \{ \mathbf{y}_t^\ast(n); t=1, \ldots, T \} \].\footnote{In addition, to calculate the elements of \( J(n) \) via finite differences, we must reapply the missing data updating condition using \( \theta(n) + \delta_m \) for each of the M separate perturbations to \( \theta(n) \). Here, we employ \[ [g_i(\mathbf{y}_{t-1}, \mathbf{x}_t, \theta(n)) - g_i(\mathbf{y}_{t-1}, \mathbf{x}_t, \theta(n) + \delta_m)]; \delta \], permitting one-sided derivatives, for the arbitrary element in \( J(n) \) of (3.12), where \( \delta_m \) is a zero vector, except for \( \delta > 0 \) in the \( m \)th position, \( m = 1, \ldots, M \).}

In short, the effect of imposing (3.17) is to permit iterative updating of all missing endogenous variables, based upon the specified model structure, (3.1), in the course of calibrating the structural parameters, \( \theta \), themselves. Thus, upon convergence of \( \{ \theta(n) \} \) to \( \theta[0] \), the algorithm will have automatically "filled in" all missing \( y_{ti} \) values with the solutions, \( y_{ti}^\ast = g_i(\mathbf{y}_{t-1}, \mathbf{x}_t, \theta[0]) \). These, in turn, have been employed in determining \( \theta[0] \), itself.

Next, we modify the parameter updating condition, (3.10), of the calibration algorithm to eliminate the missing observations within \( \{ \mathbf{y}_t \} \). Let \( N^\ast_t = \sum_{i=1}^{N^\ast_t} d_{ti} \) denote the number of \( y_{ti} \)'s observed in period \( t \); and let \( R_t \) denote an \( N^\ast_t \times N \) selection matrix, containing zeroes in each row, except for the column \( i \) associated with an observed \( y_{ti} \) value, which is

\footnote{It is also worth remarking that, given \( \mathbf{y}_0 \), if the dynamic simulation path model, (3.1)' and (3.2)', is employed, then the entire sequence, \( \{ \mathbf{y}_t^\ast(n) \} \), can be calculated directly, regardless of any missing lagged endogenous variables in \( \{ \mathbf{y}_{t-1} \} \), employed in determining \( \theta[0] \).}
set to unity.\(^1\) Thus \( R_t \) is determined by \( d_t \equiv (d_{ti}) \). Note, if \( R_t = I_N \), then \( d_t \) is the unit vector, and all elements of \( Y_t \) are observed.

Following analogous procedures to derive (3.10), we obtain the modified algorithm,

\[
\theta(n+1) = \theta(n) + \lambda(n) \cdot [J^*(n)' \cdot W^*(n)' \cdot J^*(n) - [J^*(n)' \cdot W^*(n)' \cdot (u-u^*(n))] ,
\]

(3.18)

where, to account for a possible rank deficiency, we use \( A^\top \) to denote a generalized inverse (Rao and Mitra (1977)) of \( A \). Let \( (NT)^* = \sum_{t=1}^{T} N_t^* \leq (NT) \) denote the total number of observations on all endogenous variables in the incomplete sample. Then we define

\[
J^*(n)' = \begin{bmatrix} J^*(n)' & J^*(n)' & \cdots & J^*(n)' \end{bmatrix}_M(NT)^* ,
\]

(3.19a)

\[
W^*(n) = \text{Diag}[W_1^*(n), W_2^*(n), \ldots, W_T^*(n)]_{(NT)^*x(NT)^*} ,
\]

(3.19b)

\[
u' = [u_1' \ u_2' \ \cdots \ u_T']_l(NT)^* ,
\]

(3.19c)

\[
u^*(n)' = [u_1^*(n)' \ u_2^*(n)' \ \cdots \ u_T^*(n)']_l(NT)^* ,
\]

(3.19d)

where \( J_t^* = R_t \cdot J(n) \), \( W_t^* = R_t \cdot W(n) \), \( R_t' \), \( u_t = R_t \cdot Y_t \) and \( u_t^*(n) = R_t \cdot Y_t^*(n) \) are employed to select the appropriate elements.

\(^1\) For example, if \( Y_t \) is a 4x1 vector with elements 2 and 4 observed, then

\[
R_t = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} .
\]
Finally, we note that missing $y_{t,i}'s affect certain definitions of
the weighting matrices. In choice (2), the sample variance is now redefined
as $s_{ii} = \frac{1}{T_i} \sum_{t=1}^{T} d_{ti} \cdot (y_{t,i} - \bar{y}_{i})^2$, where $\bar{y}_{i} = \frac{1}{T_i} \sum_{t=1}^{T} d_{ti} \cdot y_{t,i}$ and $T_i = \sum_{t=1}^{T} d_{ti}$. Similarly, in (3), $s_{ij} = \frac{1}{T_{ij}} \sum_{t_i=1}^{T_i} d_{ti} \cdot d_{t_j} \cdot (y_{t,i} - \bar{y}_{i}) \cdot (y_{t,j} - \bar{y}_{j})$ with $T_{ij} = \sum_{t_i=1}^{T_i} d_{ti} \cdot d_{t_j}$.

Choices (5) and (6) depend upon $\Theta$, and thus require updating each iteration. In (5), $W(t) = (\hat{\Theta}(t))^{-1}$, where $\hat{\Theta}(t) = \frac{1}{T_{ij}} \sum_{t_i=1}^{T_i} d_{ti} \cdot d_{t_j} \cdot (y_{t,i}^* - \hat{y}_{t,i}^*) \cdot (y_{t,j}^* - \hat{y}_{t,j}^*)$. Case (6) also uses the missing data updating condition, and is defined analogously.

In contrast to the iterative multiple numerical integrations required
to accommodate incomplete endogenous data in the general stochastic structural
form, (2.1), embedding the missing data updating condition within the
calibration algorithm for the deterministic structural form, (3.1), by comparison, is an extremely simple device. In principle, it permits
determination of $\Theta$, while simultaneously estimating all missing endogenous variables by their model solutions, $y_{t,i}^* = g_i(y_{t-1}, x_{t}, \Theta)$. However, even with the use of (3.17), implementation (as in the stochastic case) requires a complete set of values for both the exogenous sequence, $\{x_t\}$, and the relevant lagged endogenous variables within $y_0$. We now consider relaxing this condition.

With incomplete data on $y_0$ and/or $\{x_t\}$, our approach (in either the
stochastic model, (2.1), or the deterministic structural model, (3.1)) is to
treat each missing $y_{0,i}$ or $x_{t,k}$ element as an additional parameter to be
estimated.\textsuperscript{1} Let $d_0 = (d_{0,i})$ denote an Nx1 indicator vector, where

\textsuperscript{1} See Afifi and Elashoff (1967) for discussion of missing data procedures for a linear regression model. Other approaches will be entertained in Stage Three.
\[ \begin{aligned} 
&d_{0i} = \begin{cases} 
1, & \text{if } y_{0i} \text{ is observed or not relevant to calculating } y_1^n \\
0, & \text{if } y_{0i} \text{ is a missing lagged endogenous variable}.
\end{cases} 
\end{aligned} \tag{3.20} \]

Thus, \( y_0 \) has \( 0 \leq M^0_A = \sum_{i=1}^{N} (1-d_{0i}) \leq N \) missing lagged endogenous variables. Also, let \( C = (c_{tk}) = (c_{tk}) \) denote a \( T \times K \) indicator matrix for the exogenous variables, where

\[ c_{tk} = \begin{cases} 
1, & \text{if } x_{tk} \text{ is observed} \\
0, & \text{if } x_{tk} \text{ is missing}.
\end{cases} \tag{3.21} \]

Thus, \( x_t \) has \( 0 \leq M^A_t = \sum_{k=1}^{K} (1-c_{tk}) \leq K \) missing exogenous variables.

Define the \( M^A_t \) element vector of "additional" parameters,

\[ \vartheta^A_t = [\vartheta_0^A, \vartheta_1^A, \ldots, \vartheta_T^A]' \cdot M^A_{x1} \tag{3.22} \]

where each element of \( \vartheta^A_t \) satisfies a restriction of the form

\[ \vartheta_{0m} = y_{0i} \quad \text{if } d_{0i} = 0, \tag{3.23} \]

for \( m = 1, \ldots, M^0_A \). Also, each element of \( \vartheta^A_{-t} , t = 1, \ldots, T \), is restricted by

\[ \vartheta_{tm} = x_{tk} \quad \text{if } c_{tk} = 0, \tag{3.24} \]

for \( m = 1, \ldots, M^A_t \). The original \( M \)-vector of structural parameters, \( \vartheta \) is then "replaced" by the \((M+M^A)\) - vector, \( [\vartheta', \vartheta^A]' \); and the basic calibration algorithm applies--now subject to the \( M^A \) restrictions of the form, (3.23) and (3.24). Note that the augmented Jacobian matrix, \( J^*(n) \) of (3.19a), will
now have the patterned form,\(^1\)

\[
J^{*}(n)' R = \begin{bmatrix}
J^{*}(n)' \\
J^{*}(n)' \\
\vdots \\
J^{*}(n)' \\
J^{*}(n)' \\
\end{bmatrix} = \begin{bmatrix}
J_{11}^{*}(n)' & J_{21}^{*}(n)' & \cdots & J_{T1}^{*}(n)' \\
0 & J_{22}^{*}(n)' & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & J_{TT}^{*}(n)' \\
\end{bmatrix}
\]  

(3.25)

where \( J^{*}(n)' \) in the second expression is given by (3.19a) in the complete data case, and \( J^{*}A(n) \) denotes the "augmented" part due to \( \phi^{A} \), in which

\[
J_{01}^{*}(n) = (\partial y_{li}^{*}(n) / \partial y_{0j}) \quad \text{and} \quad J_{tt}^{*}(n) = (\partial y_{ti}^{*}(n) / \partial x_{tk}), \quad t=1,\ldots,T,
\]

for the relevant missing elements--again calculated by numerical differentiation.\(^2\)

Finally, using (3.25), we now replace \( q^{(n)} \) of (3.7) by the augmented \((M+M^{A})\) element vector,

\[
q^{(n)} = \begin{bmatrix}
q^{(n)} \\
q^{A(n)} \\
\end{bmatrix} = \begin{bmatrix}
\partial q^{(n)} / \partial q \\
\partial q^{(n)} / \partial q^{A} \\
\end{bmatrix} = J^{*}(n)' \cdot w^{*}(n) \cdot (u-u^{*}(n)) .
\]  

(3.26)

3.4.2 Uniqueness

Our preceding discussion provides a general approach for the calibration of a deterministic structural model, (3.1), on the basis of the configuration of available data within \( Y_0, Y \) and \( X \). Modifications in the basic calibration

---

\(^1\) It should be possible to exploit the pattern of zeroes in an efficient computational design.

\(^2\)'We employ 'R' to denote 'is replaced by'.
algorithm require only the missing data indicator information, defined by \( d_0 \), \( D \) and \( C \), respectively. However, given the arbitrary configuration of missing data, the investigator may be unable to determine, \textit{a priori}, whether or not the resulting calibration, \( \theta^{[0]} \), is unique.

One case may be dispensed with immediately. Clearly, the uniqueness of \( \theta^{[0]} \) requires that the Hessian matrix,

\[
H^*(n) \equiv [J^*(n)' \cdot W^*(n) \cdot J^*(n)]
\]

in (3.18) be of full rank—with \( J^*(n) \) suitably augmented as in (3.25), if \( Y_0 \) or \( \{x_t\} \) are incomplete, \( n=1,2,\ldots \) Thus, using the so-called "order conditions", necessary for identifiability, if \( (M+M^A) > (NT)^* \), then \( \theta^{[0]} \) will not be unique. However, in cases where \( (M+M^A) \leq (NT)^* \), the investigator may not be able to determine, by analytical methods, whether or not the model is uniquely calibrable.

In general, therefore, the uniqueness of a particular model calibration is intrinsically connected with the rank of the limiting Jacobian matrix, \( J^* = \lim \limits_{n \to \infty} J^*(n) \). Accordingly, to evaluate the numerical rank of \( J^*(n) \) at each iteration, as well as to calculate the Moore-Penrose generalized inverse of \( H^*(n) \) in (3.18), we employ a Singular Value Decomposition (SVD)—see Golub (1968), Golub and Reinsch (1970) or Belsley (1974). An appropriate SVD for \( J^*(n) \) is given by

\[
J^*(n) = A(n) \cdot \Delta(n) \cdot B(n)^T \quad (NT)^* \cdot (NT)^* \cdot (NT)^* \cdot (M+M^A)^T \cdot (M+M^A)^T.
\]

Here, \( A(n) \) is an orthonormal matrix in the inner-product space defined by the weighting matrix, \( W^*(n) \), i.e., \( A(n)' \cdot W^*(n) \cdot A(n) = I \); \( B(n) \) is an
orthonormal matrix, $B(n)^T \cdot B(n) = I$; and $\Delta(n)$ is an $(NT) \times (M + M^A)$ matrix, which can be represented in the form,

$$
\Delta(n) = \begin{bmatrix}
p(n) & \ldots & 0 \\
\ldots & \ldots & \ldots \\
0 & \ldots & 0
\end{bmatrix},
$$

(3.29)

where $P(n)$ is a diagonal $m^0(n) \times m^0(n)$ matrix with strictly positive diagonal elements—the "singular values" of $J^*(n)$ (see Golub (1968)).

Substitution of (3.28) into (3.18) yields an equivalent, and computationally accurate, augmented parameter updating formula,

$$
\hat{\theta}(n+1) = \hat{\theta}(n) + \lambda(n) \cdot B(n) \cdot \Delta(n)^{-1} \cdot A(n)^T \cdot W(n) \cdot (u - u^*(n)),
$$

(3.30)

where $\Delta(n)^{-1}$ obtains from $\Delta(n)$ of (3.29), by replacing $P(n)$ by $P(n)^{-1}$.

This permits calculation of $\hat{\theta}^0 = \lim_{n \to \infty} \hat{\theta}(n)$ in cases where the $\{J^*(n)\}$ are of less than full rank. In addition, the number of functionally independent elements in $\hat{\theta}^0$ is the numerical rank, $m^0 = \lim_{n \to \infty} m^0(n)$ of $J^* = \lim_{n \to \infty} J(n)$, obtained as the number of positive diagonal elements in $P = \lim_{n \to \infty} P(n)$ of (3.29). Finally, under the stochastic interpretation for the discrepancies, an estimate of the covariance matrix, (3.14), can be calculated by

$$
\operatorname{Cov}(\hat{\theta}^0) = \lim_{n \to \infty} B(n) \cdot \Delta(n) \cdot B(n)^T,
$$

(3.31)
where $\theta^{[0]}$ is the limiting value of the (augmented) parameter vector\(^1\).

This completes our discussion of Stage One calibration methods.

We now summarize the implications of this discussion. First, as the rank of the limiting Jacobian matrix, $J^*$, satisfies the condition,

$$m^{[0]} = \text{Rank}(J^*) \leq \min \{ (M+M^A), (NT)^* \},$$

(3.32)

clearly $J^*$ will be rank deficient whenever $(NT)^* < (M+M^A)$. Thus $\theta$ will not be uniquely calibrable whenever there are fewer total observations on the $y_{ti}$'s than parameters—the customary "order condition."

Suppose, next, that there are "sufficient" numbers of observations, i.e., $(NT)^* > (M+M^A)$. In this case, in general, the investigator will not know a priori whether the particular configuration of available data permits the unique determination of $\theta$. Instead, we use the modified algorithm for incomplete data, (3.30). Upon convergence, the numerical rank, $m^{[0]}$, of $J^*$ will be "revealed" to the investigator, as the number of singular values in the diagonal matrix $P$ in (3.29) as $n \to \infty$. This "informs" the investigator that only $m^{[0]} \leq (M+M^A)$ parameters in $\theta^{[0]}$ can be uniquely determined from the present use of the configuration of available data; and we must proceed to either Stage Two or Three. Finally, if $m^{[0]} = (M+M^A)$, we have uniquely calibrated the model with $\theta^{[0]}$—including both the $M$ structural parameters in $\theta$, and the $M^A$ missing observations implicit in $\theta^A$.

\(^1\) This formula does not take into account the fact that certain missing lagged endogenous variables are calibrated (via the missing data up-dating condition, (3.17)) by prior solution values. Thus, variation is conservatively treated. In the dynamic model, (3.1)' and (3.2)', however, this problem is avoided.
In the event of a rank deficiency, then $m^{[0]} < (N+M^A)$, and there is an equivalence class, $\Xi$, of parameter values, $\theta$, which satisfy the equation, $Q(\theta) = Q^{[0]}$ -- each of which will result in the same (minimum) value, $Q(\theta^{[0]}) = Q^{[0]}$, of the loss function. Our algorithm uses the SVD to calculate the Moore-Penrose generalized inverse of $H^*(n)$. This implicitly imposes a normalization rule to obtain a "unique" value, $\theta^{[0]}$, viz., that $\theta^{[0]}$ has the shortest length, $(\theta' \cdot \theta)^{1/2}$, of all members, $\theta$, of the equivalence class, $\Xi$ (see Rao and Mitra (1977)). Such a choice is convenient, but arbitrary. However, different members of $\Xi$ may result in different predictions of the endogenous variables for the same vector of predetermined variables. Thus, the immediate question of interest is whether one may exploit any additional information within the existing data set, in order to choose between alternative members of the equivalence class. If not, then the investigator must either simplify the model, or await the passage of time until sufficient data are available to uniquely calibrate the present structural model.

4. Calibration and Policy Analysis (Stage Two)

Our Stage One calibration procedures are based on minimizing a loss function involving the discrepancies between the observed elements of $\chi_t$ and the solution values, $g(z_t, \theta)$, where it is now convenient to let $z_t$ denote an $L$-vector ($K \leq L \leq N+K$) of the relevant lagged endogenous variables, plus all $K$ exogenous predetermined variables. However, as $g$ is presumed to be highly nonlinear in $z_t$, in general, its derivatives with respect to $z_t$ will
also be unknown, but computable functions of $\hat{\theta}$. In Stage Two we attempt to exploit this fact by minimizing the distance between: (1) a Taylor series approximation to the restricted reduced form, and (2) a corresponding polynomial approximation to the unrestricted reduced form, while restricting $\widehat{\theta}$ to lie within the Stage One equivalence class, $\mathcal{Q}(\widehat{\theta}) = \mathcal{Q}^{[0]} = \mathcal{Q}(\hat{\theta}^{[0]})$. The former requires a priori specification of a point of expansion, $z^*$, for the predetermined variables. An appealing choice is $z^* = [y_T' \quad x_{T+1}']'$; where $y_T$ is the last sample value of the relevant lagged endogenous variables, and $x_{T+1}$ is the "policy" vector for period $T+1$, of interest to the investigator. For an alternative purpose, one might wish to employ the within-sample mean vector, $\bar{z}$, for the predetermined variables, $z$, calculated from all available data.

4.1 The First-Order Approximation Model

We begin by considering a first-order Taylor series approximation to the restricted reduced form, relative to $z^*$, and adopt the model.

---

1/ In practice, Stage Two approximations will be constrained by the computational cost and accuracy of calculating more than second-order numerical partial derivatives (Goldfeld and Quandt (1972)). These occur in certain extensions to the first-order approximation model (see below).

2/ The vector, $x_{T+1}$ will contain both policy instruments and assumptions for the remaining uncontrolled exogenous variables. If $y_T$ has any missing values, $y_{T_1}$, we use $y_{T_1}^{(m)}$, via the missing data updating condition. This is of importance since the data in $y_T$ do not all come at once--rather there are "data availability lags" of different lengths.

3/ Thus, the appropriate choice of $z^*$ in Stage Two depends upon the purpose to which the model is to be put.

4/ The superscript "[ ]" denotes the order of the approximation.
\[ y_t = g(z^*, \theta) + \frac{\partial g(z^*, \theta)}{\partial z_t} \cdot (z_t - z^*) + e_t^{[1]} \Rightarrow y_t^* = y_t + e_t^{[1]} \quad (4.1) \]

Note that the approximation to \( g(z_t, \theta) \) is exact when \( z_t = z^* \). Let \( \pi_0 \) denote the \( N \)-vector of predictions from the structural model,

\[ \pi_0 \equiv (\pi_0_i) = g(z^*, \theta) = (y_t^*) \quad (4.2) \]

associated with a given \( z^* \), for any \( \theta \). Also, let \( \pi_1 \) denote the \( N \times L \) Jacobian matrix of impact multipliers,

\[ \pi_1 \equiv (\pi_{1ix}) = \frac{\partial g(z^*, \theta)}{\partial z_t} = (\frac{\partial y_t^*}{\partial z_t}) \quad (4.3) \]

again evaluated at the "policy vector", \( z^* \), of interest. Using (4.2) and (4.3), we may "reparametrize" the first-order approximation model as

\[ y_t = \pi_0 + \pi_1 \cdot (z_t - z^*) + e_t^{[1]} \Rightarrow y_t^* = \pi_1 \cdot e_t^* + e_t^{[1]} \quad (4.4) \]

where \( \pi_1 \) is the \( N \times (L+1) \) matrix of linearized reduced form parameters, \( \pi_1 \equiv [\pi_0 : \pi_1] \), and \( e_t^{[1]} = [1 : (z_t - z^*)]' \) is an \( L+1 \) vector of levels of regressors. If we ignore imposing the implied restrictions on the elements of \( \pi_1 \), i.e., the fact that each element is a specific, but unknown computable function of \( \theta \), then (4.4) may be viewed as the unrestricted
linearized reduced form, corresponding to the restricted version (4.1).\(^1\)

The basic idea in Stage Two is to choose \( \hat{\theta} \) to minimize the "distance" between: (1) the restricted reduced form parameters, \( \pi_0(\hat{\theta}) \) of (4.2) and \( \pi_1(\hat{\theta}) \) of (4.3); and (2) estimates of their unrestricted counterparts, \( \hat{\pi}_0 \) and \( \hat{\pi}_1 \), respectively; while restricting \( \theta \) to lie within the Stage One equivalence class, \( \Xi = \{ \theta : Q(\theta) = Q^{[0]} \} \).

4.2 Stage Two Calibration with Complete Data

We first consider the complete data case, with \( M^A = 0 \). Let \( \pi_i^{[1]} \) denote the \( i^{th} \) row of \( \Pi^{[1]} \) in the unrestricted reduced form, (4.4), with corresponding restricted counterpart,

\[
\pi_i^{[1]}(z^*,\theta) = [g_i(z^*,\theta); \frac{\partial g_i(z^*,\theta)}{\partial z'}].
\]

Then, we may employ the least squares estimates,

\[
\hat{\pi}_i^{[1]} = [p^{[1]'} \cdot p^{[1]^{-1}} \cdot p^{[1]'} \cdot y_i] \equiv G^{[1]} \cdot y_i,
\]

of \( \pi_i^{[1]} \), where \( p^{[1]} = [p^{[1]}, \ldots, p^{[1]}] \) is an \((L+1) \times T\) regressor matrix, \( T > (L+1) \), and \( y_i \) is a \( T \times 1 \) vector of observations on the \( i^{th} \) endogenous

---

\(^1\) Another way to view the matter is that the \( M^A \) parameters in \( \theta \) within (4.2) and (4.3) are either already more parsimonious parametrizations of the \( N \times (L+1) \) elements in \( \Pi^{[1]} \), or, if not, eventually will become so, as the extensions of Section 4.4 are introduced.
variable—evaluated separately for \( i = 1, \ldots, N \). Finally, we employ

\[
\text{Cov}(\hat{\pi}_i^{[1]}) = \hat{\pi}_i^{[1]} = s_i^2 [p_i^{[1]'} p_i^{[1]}]^{-1}
\]

(4.7)

to estimate the covariance matrix of \( \pi_i^{[1]} \), where

\[
s_i^2 = \frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{p}_i^{[1]'} \hat{\pi}_i^{[1]})^2.
\]

We now adopt a quadratic loss function for the Stage Two first-order approximation problem,

\[
Q^{[1]}(z^*, \theta) = \frac{1}{2} \sum_{i=1}^{N} (\hat{\pi}_i^{[1]} - \pi_i^{[1]}(z^*, \theta))^2 \cdot \xi_i^{[1]} \cdot (\hat{\pi}_i^{[1]} - \pi_i^{[1]}(z^*, \theta))^2,
\]

(4.8)

Thus, in (4.8), we employ the least squares estimates, \( \hat{\pi}_i^{[1]} \), of the linearized unrestricted reduced form parameters as a "fixed target"; and seek a member, \( \hat{\theta}_i^{[1]} \), of the equivalence class, \( \xi \), which minimizes the "distance" to the corresponding restricted reduced form parameters, \( \pi_i^{[1]}(z^*, \theta) \), using the inverses of the \( \xi_i^{[1]} \), \( i = 1, \ldots, N \), as weights.

Let \( g^{[1]} \) denote the \( M \times 1 \) vector of first partials of \( Q^{[1]} \), i.e.,

\[
g^{[1]}(z^*, \theta) = \frac{\partial Q^{[1]}(z^*, \theta)}{\partial \theta} = \sum_{i=1}^{N} \xi_i^{[1]} \cdot (\hat{\pi}_i^{[1]} - \pi_i^{[1]}(z^*, \theta)),
\]

(4.9)

---

1/ In actual computations, one should employ a Gram-Schmidt Decomposition (GSD) of \( p^{[1]} \) see, e.g., Lawson and Hanson (1974) or Pagano and Hartley (1980). Note also that \( C^{[1]} \) is invariant with \( i \).

2/ Note, unlike Stage One, that the quadratic loss function in Stage Two depends upon the particular choice of \( z^* \).
where, for i=1,..., N, we require the Jacobian matrices,

\[ J_i = \frac{\partial g_i}{\partial \theta} = \left[ \frac{\partial^2 g_i}{\partial \theta \partial z'} : \frac{\partial^2 g_i}{\partial \theta \partial z} : \frac{\partial^2 g_i}{\partial \theta^2} \right] \]  

(4.10)

We then define the Lagrangian function,

\[ \Lambda(\theta, \phi) = Q[1](z^*, \theta) - \phi(Q(\theta) - Q[0]) \]  

(4.11)

with scalar "multiplier", \( \phi \), (Bellman (1956b)) which yields the system of M+1 first-order conditions,

\[ \frac{\partial \Lambda}{\partial \theta} = g[1](z^*, \theta) - \phi \cdot g(\theta) = 0 \]  

(4.12a)

\[ \frac{\partial \Lambda}{\partial \phi} = Q(\theta) - Q[0] = 0 \]  

(4.12b)

Here, we seek a solution to (4.12a) and (4.12b) in the situation where \( Q[0] \) and \( z^* \) are now known constants; whereas \( Q(\theta), g(\theta) \) and \( g[1](z^*, \theta) \) are all unknown, but computable functions of \( \theta \). In addition, as output from Stage One, we have an initial value, \( \theta[0] \), which already satisfies (4.12b) with \( g(\theta[0]) = 0 \).

Following the quasi-linearization methods of Stage One, we rewrite (4.9) more compactly as

---

1/ It should be noted that in a first-order approximation model we require the second-order cross-partial, \( (\partial^2 g_i(z^*, \theta)/\partial \theta \partial z') \), which must be calculated by numerical differentiation (see Goldfeld and Quandt (1972)). Each numerical second partial requires three additional function evaluations—raising a question of the present cost, but not feasibility, of basic Stage Two procedures in problems of large dimensions.
\[ q^{[1]}(\xi^*, \delta) = J^{[1]}(\xi^*, \delta)' \cdot \hat{z}^{[1]} - 1 \cdot (\hat{\pi}^{[1]} - \bar{\pi}^{[1]}(\xi^*, \delta)) , \]  

(4.13)

where we again employ the compact "stacked" representations,

\[ J^{[1]}' = J^{[1]}(\xi^*, \delta)' = [J_{1}' \ J_{2}' \ ... \ J_{N}']_{Mx(L+1)N} , \]

(4.14a)

\[ \hat{z}^{[1]} = \text{Diag}[\hat{z}^{[1]}_1 \ \hat{z}^{[1]}_2 \ ... \ \hat{z}^{[1]}_N] \]

(4.14b)

\[ \hat{\pi}^{[1]}' = [\hat{\pi}^{[1]}_1 \ \hat{\pi}^{[1]}_2 \ ... \ \hat{\pi}^{[1]}_N]_{Lx(L+1)N} , \]

(4.14c)

\[ \bar{\pi}^{[1]}' = \bar{\pi}^{[1]}(\xi^*, \delta)' = [\bar{\pi}^{[1]}_1(\xi^*, \delta)' \ ... \ \bar{\pi}^{[1]}_N(\xi^*, \delta)']_{Lx(L+1)N} . \]

(4.14d)

Note that, in addition to the linear approximation, (3.9), of \( y^* \), which is contained within \( q \), we also require the approximations,

\[ \bar{\pi}^{[1]}(\xi^*, \delta) = \bar{\pi}^{[1]}(\xi^*, \delta) + J^{[1]}(\xi(n)).(\delta - \delta(n)) \]  

(4.15a)

and

\[ Q(\delta) = Q(n) + q(n)'.(\delta - \delta(n)) , \]

(4.15b)

relative to \( \delta(n) \), as well as the joint expansion.

\[ \phi(\chi^{*} - \chi) = \phi(n).(\chi^{*} - \chi(n)) - \phi(n).J(n).(\delta - \delta(n)) + (\phi(\delta(n)).(\chi^{*} - \chi(n)), \]

(4.15c)

relative to both \( \delta(n) \) and \( \phi(n) \). Hence, if we replace \( y^*, \bar{\pi}^{[1]}, Q \) and \( \phi(\chi^{*} - \chi) \) by their linear approximations, while evaluating all remaining functions of \( \delta \) at \( \delta(n) \), the solution to the equation system, (4.13a) and (4.13b), may be
obtained via the modified algorithm:

\[
\begin{bmatrix}
\tilde{g}_{(n+1)} \\
\phi_{(n+1)}
\end{bmatrix} = \begin{bmatrix}
\phi_{(n)} \\
\phi_{(n)}
\end{bmatrix} + \lambda_{(n)} \cdot \begin{bmatrix}
\tilde{u}_{[1](n)} & g_{(n)} \\
g_{(n)} & 0
\end{bmatrix}^{-1} \begin{bmatrix}
\tilde{g}_{[1](n)} - \phi_{(n)} \\
-\lambda_{(n)}
\end{bmatrix},
\]

(4.16)

where \( \tilde{u}_{[1](n)} \) is the \( M \times M \) symmetric matrix,

\[
\tilde{u}_{[1](n)} = j_{[1]}'(n) - j_{[1]}(n) + \phi_{(n)} \cdot j_{(n)}'. \ w_{(n)} \cdot j_{(n)}
\]

(4.17)

\( \phi_{(n)} \) is the scalar Lagrange multiplier, and \( \lambda_{(n)} \) is a scalar step-size parameter.

Implementation of (4.16) requires calculation of \( Q_{(n)} \) of (3.5), \( g_{(n)} \) of (3.7), \( \tilde{g}_{[1](n)} \) of (4.13), and \( \tilde{u}_{[1](n)} \) of (4.17), for given values of \( \phi_{(n)} \) and \( \lambda_{(n)} \). These, in turn, require that, relative to \( \phi_{(n)} \), we again calculate the forward solution sequence, \( \tilde{X}_{ti} = \{y_{*}\}_{ti} \) of (3.8d); its first-partial, \( j_{(n)} = (3y_{*}/3\phi_{m}) \) of (3.8a); and the weighting matrix, \( W_{(n)} \) of (3.8b)—just as in the Stage One calibration. However, in a first-order Stage Two approximation, in addition, we now utilize the information obtained by calculating: (a) \( \tilde{u}_{[1](n)} \) of (4.5) and (4.14d) and (b) \( j_{[1](n)} \) of (4.10) and (4.14a). These rest upon calculation of: (i) the solution values, \( y_{*}(n) = g_{i}(z, \phi_{(n)}) \); (ii) their first partials with respect to \( \phi \) and \( z \), i.e., \( 3g_{i}(z, \phi_{(n)})/3\phi \) and \( 3g_{i}(z, \phi_{(n)})/3z \), respectively; and, finally, (iii) the second-order cross partials, \( 3^2 g_{i}(z, \phi)/3\phi 3z \)—all for a
particular, invariant choice of $z^*$. Thus, fixing $z^*$, we require $3xM + M + L + 1$ additional model solutions within each iteration $n$ in order to calculate the elements of $w^{[1]}(n)$ and $J^{[1]}(n)$ in Stage Two—over and above the $(M+1)^T$ already noted for Stage One. In problems of large dimension, this may suggest skipping Stage Two, and proceeding directly to Stage Three. This will depend upon the current state of computer technology and the available research budget.  

4.3 State Two Calibration with Incomplete Data

We now consider the incomplete data case for Stage Two, in which we again permit an arbitrary configuration of available data within $y_0$, $Y$, and $X$—as indicated by the pattern of binary elements within $d_0$, $D$, and $C$, respectively. Our approach is identical to that already employed in Stage One. All missing endogenous variables, $y_{ti}$, are iteratively replaced by their current solution values, $g_i(z_t, e(n))$, via use of the missing data updating condition, (3.17). Missing predetermined (state) variables, $y_{0i}$ and $x_{tk}$, will be treated as additional parameters—thus augmenting the $M$-vector of structural parameters, $\theta$, by the $M^A$-vector, $\theta^A$, of (3.22). Further, $\theta^A$ must satisfy the $M^A$ restrictions of the form, (3.23) and (3.24). Thus, in the augmented framework, the dimension of $\theta(n)$ within the $n^{th}$ iteration is $M + M^A$, requiring appropriate adjustments in the definitions of $g^{(n)}$ (given earlier by (3.26)); as well as $g^{[1]}(n)$ and $U^{[1]}(n)$ within

1/ In calculation of numerical second derivatives, some benefit may be obtained from that fact that $z^*$ is invariant relative to $\{\theta(n)\}$—particularly in the neighborhood of $\theta[0]$.

2/ We look forward to accumulating some computational experience on such matters, once the appropriate computer software has been written.
(4.16) -- both now required to implement Stage Two.

Inspection of (4.13) and (4.17) reveals that we must again define the first-order augmented Jacobian matrix -- this time given by,

\[
\begin{bmatrix}
J[1](n)' \\
J[1]A(n)'
\end{bmatrix}
= \begin{bmatrix}
J[1](n)' & J[1](n)' & \cdots & J[1](n)'
\end{bmatrix},
\]

as with (3.25); where the \( J[1](n)' \) are defined by (4.10), evaluated at \( \hat{\theta}(n) \) and \( \bar{z}^* \); and the \( M_A \times L \) matrices, \( \hat{J}[1](n)' \), \( i = 1, \ldots, N \), are now defined by \( J[1](n)' = \frac{\partial^2 g_i(z^*, \hat{\theta})}{\partial y_0 \partial z_k} \) and \( J[1](n)' = \frac{\partial^2 g_i(z^*, \hat{\theta})}{\partial x_{tk} \partial z_k} \), \( t = 1, \ldots, T \). Thus, incomplete data in the \( M_A \) elements within \( Y_0 \) and \( X \) results in a further \( M_A \times (L+1) \cdot N \) second-partial calculations.

Next, we note the consequences of incomplete data upon: (1) the estimates of the unrestricted reduced form parameters, \( \hat{\pi}[1] \), of (4.6) and (4.14c), and (2) the inverse weighting (estimated covariance) matrices, \( \hat{\pi}[1] \), of (4.7) within \( z[1] \) of (4.14b). With complete data, \( \hat{\pi}[1] \) is invariant across iterations in \( \hat{\theta}(n) \); and thus was termed a "fixed target" for the restricted reduced form coefficients, \( \hat{\pi}[1](\bar{z}^*, \hat{\theta}) \) of (4.5) and (4.14d). With incomplete data, however, this is no longer the case. As before, given the augmented parameter vector, \( \hat{\theta}(n) \), and employing the missing data updating condition, (3.17), we replace all missing values iteratively, before proceeding to the parameter updating formula. Thus, we must now
calculate

\[ \hat{\pi}_{i}(n) = [p^{(1)}(n)'p^{(1)}(n)]^{-1}[p^{(1)}(n)'\gamma_{i}(n)] = G^{(1)}(n)\gamma_{i}(n) \quad (4.19) \]

and

\[ \hat{\pi}_{i}^{2}(n) = s_{i}^{(2)}[p^{(1)}(n)'p^{(1)}(n)]^{-1}, \quad (4.20) \]

with \( s_{i}^{(2)} = \frac{1}{T} \sum_{t} (y_{t}^{(n)} - p_{t}^{(1)}(n)' \hat{\pi}_{i}(n))^{2} \) within each iteration \( n \), using \( \hat{\pi}_{i}(n) \) and \( \hat{\pi}_{i}^{2}(n) \) in (4.13) and (4.17) to calculate \( q^{(1)}(n) \) and \( u^{(1)}(n) \), respectively. \(^2\)

Finally, to explicitly accommodate a possible rank deficiency in the \((M+M^{A}+1)\)-order bordered Hessian matrix,

\[ H^{(1)}(n) \equiv \begin{bmatrix} u^{(1)}(n) & q(n) \\ q(n)' & 0 \end{bmatrix}, \quad (4.21) \]

in (4.16), we must again resort to a generalized inverse. One approach is to first apply partitioned inversion formulas to \( H^{(1)}(n) \) (Rao and Mitra (1977)),

---

1/ Henceforth, whenever required by rank deficiency, '-' will be interpreted as a generalized inverse, '-', with the appropriate SVD employed in computation.

2/ Thus \( \hat{\pi}_{i}(n) \) now becomes a "moving target" for \( \hat{\pi}_{i}(n) \) relative to \( \gamma_{i}(n) \). In addition, the inverse of the weighting matrix, \( \hat{\pi}_{i}^{2}(n) \), fails to reflect the fact that certain of the elements in \( \gamma_{i}(n) \) and \( p^{(1)}(n) \) are themselves estimated. Again, we use a GSD and SVD in actual computations.
and then calculate the generalized inverse, \([u^{[1]}(n)]^{-1}\), by its SVD -- both for computational accuracy, and to determine the numerical rank, \(m^{[1]}(n)\), at issue.\(^1\) The outcome is that the Stage Two parameter updating condition, (4.16), may be recast via the (incomplete data) algorithm,

\[
\begin{bmatrix}
\tilde{\theta}^{[1]}(n+1) \\
\tilde{\phi}^{[1]}(n+1)
\end{bmatrix} = \begin{bmatrix}
\tilde{\theta}^{[1]}(n) \\
\tilde{\phi}^{[1]}(n)
\end{bmatrix} + \lambda(n) \begin{bmatrix}
c^{[1]}(n) \\
c^{[1]}(n)
\end{bmatrix} \begin{bmatrix}
c^{[1]}(n) & c^{[1]}(n) \\
\end{bmatrix} \begin{bmatrix}
q^{[1]}(n) - \phi(n). q(n) \\
-q(n)
\end{bmatrix}
\]  

(4.22)

where (see, again, Rao and Mitra (1977)),

\[
c^{[1]}(n) = u^{[1]}(n)^{-1}. [I + c^{[1]}(n) \cdot q(n) \cdot q(n)', \cdot u^{[1]}(n)^{-1}] 
\]  

(4.23a)

\[
c^{[1]}(n) = -u^{[1]}(n)^{-1} \cdot q(n) \cdot c^{[1]}(n) 
\]  

(4.23b)

\[
c^{[1]}(n) = -(q(n)'), \cdot u^{[1]}(n)^{-1} \cdot q(n)^{-1} 
\]  

(4.23c)

The important point to note is that, in addition to \(Q^{(n)}\) and \(q^{(n)}\) -- already required for Stage One, all Stage Two formulas require, in addition, only: (1) \(q^{[1]}(n)\) of (4.13) -- suitably augmented as described above; and (2) the generalized inverse of \(u^{[1]}(n)\). The latter may be calculated via the SVD,

\[
u^{[1]}(n) = \Delta^{[1]}(n). \Delta^{[1]}(n). \Delta^{[1]}(n),
\]

(4.24)

---

\(^1\) I would like to thank Professor Marcello Pagano for discussions on this point.
in the Moore-Penrose "unique" form, where $A^{[1]}(n)$ and $B^{[1]}(n)$ are both orthonormal $(M+M^A)$-order square matrices, and $\Delta^{[1]}(n)$ is a diagonal matrix of non-negative elements, with $m^{[1]}(n)$ singular values, $0 \leq m^{[1]}(n) \leq (M+M^A)$.

Finally, let $m^{[1]} = \lim_{n \to \infty} m^{[1]}(n)$, denote the numerical rank of the limiting matrix, $U^{[1]} = \lim_{n \to \infty} U^{[1]}(n)$.

There are four possible qualitative outcomes to the problem of minimizing $Q^{[1]}(\tilde{z}, \hat{\theta}^*)$, with respect to $\hat{\theta}$, subject to $Q(\hat{\theta}) = Q[0]$, for for a given $\tilde{z}^*$:

**Case [1a]:** The best outcome is to obtain a unique solution with zero minimand. This occurs when $m^{[1]} = (M+M^A)$ and $Q^{[1]}(\hat{\theta}^{[1]}, \tilde{z}^*) = 0$. It implies that we have located a member of the equivalence class, $\Xi$, such that the predictions, (4.2), and impact multipliers, (4.3), obtained from both the restricted and unrestricted estimates of a first-order approximation to the reduced form, are identical, given the configuration of available data. Hence, $\hat{\theta}^{[1]}$ may be employed within the structural model to generate predictions, $\hat{\Pi}_0^*$, and impact multipliers, $\hat{\Pi}_1^*$, associated with the policy, $\tilde{z}^*$. 1/

**Case [1b]:** The next best outcome is a zero minimand, $Q^{[1]}$, but a non-unique structural calibration, $m^{[1]} < (M+M^A)$. This implies that there is a subset of the equivalence class of structural calibrations, all of which are consistent with unrestricted estimates of the predictions and impact multipliers associated with a given policy, $\tilde{z}^*$. More information, however,

---

1/ The "point" at which this occurs may be termed "exact calibration" of $\theta$, relative to a first-order approximation model--see below.
must be employed in the calibration process before unique values of $\theta$ can be obtained. Such parameter values are, therefore, suspect.

**Case [1c]:** The third possibility is a positive minimand at $\theta^{[1]}$, $\theta^{[0]}$, i.e., $Q^{[1]}(z^*, \theta^{[1]}) > 0$ with $Q(\theta^{[1]}) = Q^{[0]}$. This implies that we have obtained the "closest" structural calibration, $\theta$, within $\Xi$, to the unrestricted estimates of a first-order reduced form approximation. Such estimates may be unique ($m^{[1]} = M + M_A$), or non-unique ($m^{[1]} < M + M_A$). However, $\theta^{[1]}$ is, at least, an improvement over $\theta^{[0]}$ in that the former is: (1) "closer" (in terms of the loss function, (4.8)), i.e., $Q^{[1]}(z^*, \theta^{[0]}) > Q^{[1]}(z^*, \theta^{[1]})$; and (2) contains more functionally independent parameters, i.e., $m^{[1]} > m^{[0]}$—see below. 1/

**Case [1d]:** The worst case obtains when $\theta^{[1]} = \theta^{[0]}$, i.e., there is no change in $\theta$. In this event, the additional Stage Two perturbations, relative to $z^*$ and the available data, contain no additional information about $\theta$ in the model, $f$.

In summary, each of the cases above is a verifiable outcome. If cases [1a] or [1b] obtain, then the structural model may be properly employed

---

1/ However, the "closest-possible" calibration may be "good-enough" (Bellman (1966a)).
for policy analysis—however, in [1b], \( \theta \) is not unique. Otherwise, if [1c] or [1d] obtains, we should, presumably, employ the unrestricted reduced form estimates for prediction. In all cases except [1a], calibration and policy analysis would benefit from either: (1) additional information or (2) "tightening" the structural specification of the model further.

One important point, however, remains. Note from (4.17) that we can rewrite \( U_1(n) \) as

\[
U_1(n) = K_1(n)', K_1(n) + K(n)', K(n) = [K_1(n)', K(n)'] \cdot \begin{bmatrix} K_1(n) \\ K(n) \end{bmatrix}.
\]

To see this, suppose we apply a Cholesky Decomposition (CD)—see, e.g., Belsley (1974)—to each of:

\[
\hat{W}(n) = \hat{v}(n)', \hat{v}(n)
\]

and

\[
\delta_1(n) = \hat{v}[1](n)', \hat{v}[1](n)
\]

in which \( v(n) \) and \( v[1](n) \) are lower triangular matrices of order \((NT)^* \) and \((L+1).N\), respectively. Then, in view of (4.17), we have

\[
K(n) = \phi(n)^2 . v(n). j(n)
\]

and

\[
K[1](n) = v[1](n). j[1](n)
\]

---

1/ The arguments favoring the use of structural deterministic models to generate policy analyses and predictions are the same as those advanced in the case of the stochastic econometric model—see, e.g., Christ (1966) or Klein (1974). The present discussion indicates exactly when a particular calibration of a structural model may be employed for predictions and, conversely, when the unrestricted reduced form must be employed. For a contrary view, see Friedman (1953). We regard the matter, inter alia, as one of parsimony of parametrization (Occam's razor). Structural calibration both requires and provides more information.
in (4.25). We now recall that the rank of $K^{(n)}$ within Stage One is already known to be $m^0$; whereas (4.25) implies that the rank, $m^{[1]}$, associated with $U^{[1]}$, must satisfy the inequality,

$$m^0 \leq m^{[1]} \leq (M+M^A)$$

(4.28)

In short, the effect of the Stage Two first-order approximation algorithm is to produce an alternative calibration, $\theta^{[1]}$, with (1): the same Stage One loss-function value, $Q(\theta^{[1]}) = Q^0$, but (2) with a possible increase in the number of functionally independent parameters from $m^0$ to $m^{[1]}$, given $z^*$. Whether or not a rank increase obtains is equivalent to whether or not the further perturbations of the model utilized in Stage Two were informative about the structural form, $f$, relative to the given data base at hand. This, we must, in general, determine by computer, using numerical methods, such as those given above.

4.4 Direct Extensions

In principle, the Stage Two calibration approach may be extended in at least two ways: (1) expansion of the order of the Taylor series approximation to the restricted and unrestricted reduced forms; and (2) extension of the policy horizon of interest to the investigator to more than one-period. However, in the first case, the current state of computer technology may constrain the feasibility of such extensions. It is
instructive, however, to consider each of these in turn.\textsuperscript{1/}

4.4.1 Higher-Order Approximations

Consider replacing (4.1) by a second-order Taylor series approximation model, defined for the \(i\)th endogenous variable as

\[
y_{ti} = g_i(z^{*}, \theta) + \frac{\partial g_i(z^{*}, \theta)}{\partial z_1} \cdot (z_t - z^{*}) + \frac{1}{2} \cdot \frac{\partial^2 g_i(z^{*}, \theta)}{\partial z_1^2} \cdot (z_t - z^{*})^2 + \ldots + e_{ti}^{(2)}
\]

(4.29)

\[
y_{ti}^{(2)} + e_{ti}^{(2)} = y_{ti}^{[2]} + e_{ti}^{[2]} = \pi_{0i} + \pi_{1i} \cdot (z_t - z^{*}) + \frac{1}{2} \cdot \pi_{2i} \cdot (z_t - z^{*})^2 + \ldots + e_{ti}^{(2)}
\]

with corresponding unrestricted analogue,

\[
y_{ti} = \pi_0^{(2)} + \pi_{1i} \cdot (z_t - z^{*}) + \frac{1}{2} \cdot \pi_{2i} \cdot (z_t - z^{*})^2 + \ldots + e_{ti}^{(2)}
\]

(4.30)

Here, \(\pi_0 = (\pi_{0i})\) and \(\pi_1 = (\pi_{1il})\) are given by (4.2) and (4.3), respectively; whereas in a second-order approximation,

\[
\pi_2 = [\pi_{21} \; \pi_{22} \ldots \; \pi_{2N}]
\]

(4.31)

defines the \(N \cdot L^2\) "new" unrestricted reduced form parameters.\textsuperscript{2/}

By analogy with the first-order case, we again consider the arbitrary \(i\)th unrestricted second-order reduced form equation,

\[
1/\text{For each feasible extension, a taxonomy of verifiable outcomes, analogous to cases [1a]-[1d] in the first-order approximation model, will obtain.}

\[
2/\text{The number, } N \cdot \lambda_0^2, \text{ of structural parameters within the restricted reduced form in the (possibly augmented) vector, } \theta, \text{ remains the same.}
\]
\[ y_{ti} = \mathbf{p}_{t}^\prime \Pi_{i} + e_{ti}, \quad (4.32) \]

where now the regressor-vector, \( \mathbf{p}_{t} \), is augmented to

\[ \mathbf{p}_{t}^{[2]} = [1; (z_{t} - \bar{z})^\prime; \text{vec } \mathbf{M}_{z_{t} - \bar{z}}^\prime] = [\mathbf{p}_{t}^{[1]}; \text{vec } \mathbf{M}_{z_{t} - \bar{z}}^\prime], \quad (4.33) \]

in which \( \text{vec } \mathbf{M}_{z_{t} - \bar{z}}^\prime \) is a stacked "innovation-vector" of squares and cross-products of the elements of \( (z_{t} - \bar{z}) \), i.e.,

\[ \text{vec } \mathbf{M}_{z_{t} - \bar{z}}^\prime = \left[ \frac{1}{2} (z_{t1} - z_{1}^\ast)^2; (z_{t1} - z_{1}^\ast) (z_{t2} - z_{2}^\ast) \ldots \frac{1}{2} (z_{tL} - z_{L}^\ast)^2; \right. \]

\[ \ldots; (z_{t1} - z_{1}^\ast) (z_{tL} - z_{L}^\ast) (z_{t2} - z_{2}^\ast) (z_{tL} - z_{L}^\ast) \ldots \frac{1}{2} (z_{tL} - z_{L}^\ast)^2 \right], \quad (4.34) \]

and the corresponding \( 1 + L + [L(L+1)/2] \)-element parameter vector, \( \Pi_{i}^{[2]} \), is defined by

\[ \Pi_{i}^{[2]} = \left[ \Pi_{i0}; \Pi_{i1i}; \text{vec } \Pi_{i2i} \right] = \left[ \Pi_{i}^{[1]}; \text{vec } \Pi_{i2i} \right]. \quad (4.35) \]

Here, \( \text{vec } \Pi_{i2i} \) is conformable with \( \text{vec } \mathbf{M}_{z_{t} - \bar{z}}^\prime \) of (4.34), i.e.,

\[ \text{vec } \Pi_{i2i} = \left[ \Pi_{i11}; \Pi_{i12} \Pi_{i22} \ldots \Pi_{i1L} \Pi_{i2L} \ldots \Pi_{iLL} \right], \quad (4.36) \]

which involves placing the lower triangle of the \( L \times L \) symmetric matrix, \( \Pi_{2i} \), in an \( L(L+1)/2 \) element vector. Finally, the corresponding restricted reduced form coefficient vector is given by
Given these preliminaries, several points should be made. First, exactly the same procedures may be applied within Stage Two to the second-order approximation model, (4.29), and its unrestricted analogue, (4.32), as were employed in the first-order case to (4.1) and (4.4), respectively. We simply replace the superscript '1' by '2', and utilize the appropriate augmented definitions. However, while there is no conceptual difficulty with a second-order approximation model, inspection of the Jacobian matrix, 
\[ J^2(z^*, \theta) \]
where
\[ J^1(z^*, \theta) \]
is defined by (4.10), and
\[ J^2 = \left[ J^2 \right] \]
in which \[ J^2 = \left[ J^2 \right] \]. In short, to implement a second-order extension, we require numerical third-order derivatives. Thus, at present, such an extension may pose computational feasibility and/or cost problems.

Second, all successive approximation models are nested. E.g., in comparing [2] with [1], (4.29) and (4.32) contain (4.1) and (4.4), respectively, as special cases. It follows that the unrestricted reduced form
regression coefficients, $\pi_{1i}^2$, may be estimated using familiar "step-up/step-down" procedures (see, e.g., Seber (1977)), by augmenting the regressor matrix, $\mathbf{P}_1$, in (4.6), by the recursion formula,

$$\mathbf{P}_2 = [\mathbf{P}_1; \text{vec } \mathbf{M}_z \mathbf{z}^{\star}]$$

involving the $L(L+1)/2$ additional columns, $\text{vec } \mathbf{M}_z \mathbf{z}^{\star}$.  

Third, problems of incomplete data may be handled analogously to the procedures employed in the first-order approximation model. In this case, however, the Jacobian matrix, $\mathbf{J}^{1}(n)'$, of (4.18) must now be augmented as follows:

$$\mathbf{J}^{2}(n)' = [\mathbf{J}^{1}(n)'; \mathbf{J}^{2}(n)']$$  \hspace{1cm} (4.40)

where $\mathbf{J}^{1}(n)'$ is given by (4.18) and

$$\mathbf{J}^{2}(n)' \equiv \begin{bmatrix} \mathbf{J}^{2}(n)' \\ \vdots \\ \mathbf{J}^{2}(n) \mathbf{A}(n)' \end{bmatrix} = \begin{bmatrix} \mathbf{J}^{2}(n)' & \cdots & \mathbf{J}^{2}(n)' \\ \mathbf{J}^{2}(n)' & \cdots & \mathbf{J}^{2}(n)' \\ \vdots & \cdots & \vdots \\ \mathbf{J}^{2}(n)' & \cdots & \mathbf{J}^{2}(n)' \end{bmatrix}$$  \hspace{1cm} (4.41)

in which $\mathbf{J}^{1}(n)'$ is given by (4.39), evaluated at $\hat{\theta}(n)$, and

---

1/ In this connection, it should be noted that to preserve "degrees of freedom" for calibration, we may wish to make only a partial second-order expansion--using only a subset of elements within $\mathbf{M}_z \mathbf{z}^{\star}$, etc. In addition, proceeding systematically through step-up procedures will enable us to discover a minimal partial second-order approximation model which achieves exact calibrability--if at all. Thus, if the elements within $\mathbf{M}_z \mathbf{z}^{\star}$ are "ordered", then at some "point" we cross a "threshold" with respect to structural calibrability.

2/ See also Pagano and Hartley (1980) for the use of the GSD and SVD in the context of "step-up/step-down" procedures with polynomial distributed lags.
\[ J^{[2]}(A(n))' = \begin{bmatrix} J^{[2]}(A(n)) & J^{[2]}(A(n)) \ldots J^{[2]}(A(n)) \end{bmatrix}_t. \]

where

\[ J^{[2]}(A(n))_0 = (\frac{\partial^3 g_i(x^*, y(n))}{\partial y_0 \partial z_k \partial z_l}) \]

and

\[ J^{[2]}(A(n))_t = (\frac{\partial^3 g_i(x^*, y(n))}{\partial x^t \partial z_k \partial z_l}) \]

-- again involving third partials.

However, if computable, then by arguments analogous to the first-order case, the convergent iterate, \( g^{[2]} \), will have rank, \( m^{[2]} \), satisfying

\[ m^{[0]} \leq m^{[1]} \leq m^{[2]} \leq (M + M^A). \]

Higher-order approximations of order \( r \geq 2 \) are obviously only constrained by computational feasibility. If feasible, they would involve the same constrained, augmented regression methods (previously discussed in the matter of \( [1] + [2] \)), where in "stepping-up" to a higher-order approximation, calculation of \( g^{[r]} \) draws upon calculations employed to obtain \( g^{[r-1]} \). In principle, \( g^{[r]} \), obtained from an \( r \)th-order approximation model, will be of rank \( m^{[r]} \), where \( m^{[r-1]} \leq m^{[r]} \leq (M + M^A) \). Clearly, as \( r \to \infty \), the approximation at \( x^* \) becomes exact for any \( z_t \), i.e., \( Y_t^{[r]} = Y_t \).

However, this limiting expansion does not necessarily imply \( \lim_{r \to \infty} m^{[r]} \) achieves \( M + M^A \), as higher-order expansions may be uninformative with respect to \( g \) -- depending upon the model, \( f \), and the available data base.

4.4.2 Multi-Period Policies:

An alternative direction for extension of the first-order approximation model is to replace the single-period policy, \( x^* = Z_{T+1} = [Y_T' x_{T+1}]' \), by the multi-period policy sequence, \( \{z_{T+S} = [Y_{T+S-1}' x_{T+S}]': s=1, \ldots, S\} \). However, for

\[ 1/ \] This may be viewed as an asymptotic argument with respect to \( T \).
S > 1, the requisite future values of the lagged endogenous variables, 
\( y_{T+s-1} \; s=2, \ldots , S \), are not observed, and thus, via the missing data 
updating condition, are replaced by their sequential predictions, 
\( y_{T+s-1} = g(y_{T+s-2}, x_{T+s-1}, \theta) \). Thus, we employ \( y_T \) (the initial value) and 
\( [x_{T+s} : s=1, \ldots , S] \) (the policy sequence), as the "new" point of expansion. We 
illustrate with a two-period policy model (S=2), retaining a first-order 
Taylor series approximation.

The two-equation, first-order model, associated with the two-period 
policy sequence, \( \{x_{T+1}, x_{T+2}\} \), given the initial state of the system, \( y_T \), 
can be represented as the linear approximation to the restricted reduced form:

\[
\begin{align*}
\Delta y_T &= g(y_T, x_{T+1}, \theta) + \frac{\partial g(y_T, x_{T+1}, \theta)}{\partial y_T} \cdot (y_T - y_T) \\
&\quad + \frac{\partial g(y_T, x_{T+1}, \theta)}{\partial x_{T+1}} \cdot (x_T - x_{T+1}) + \epsilon_{1t} = \epsilon_{1t} + \epsilon_{1t} \\
\Delta y_{T+1} &= g(y_{T+1}, x_{T+2}, \theta) \\
&\quad + \frac{\partial g(y_{T+1}, x_{T+2}, \theta)}{\partial y_{T+1}} \cdot \frac{\partial g(y_{T+1}, x_{T+2}, \theta)}{\partial y_T} \cdot (y_T - y_T) \\
&\quad + \frac{\partial g(y_{T+1}, x_{T+2}, \theta)}{\partial x_{T+1}} \cdot \frac{\partial g(y_{T+1}, x_{T+2}, \theta)}{\partial x_{T+1}} \cdot (x_T - x_{T+1}) \\
&\quad + \frac{\partial g(y_{T+1}, x_{T+2}, \theta)}{\partial x_{T+2}} \cdot (x_{T+1} - x_{T+2}) + \epsilon_{2t+1} = \epsilon_{2t+1} + \epsilon_{2t+1},
\end{align*}
\]

where (4.44a) is identical to (4.1), and (4.44b) exhibits the application of 
the chain rule for (numerical) derivatives in calculation of multi-period 
policies, resulting from sequential use of the missing data updating.
We reparametrize the approximation model, (4.44a) and (4.44b), as follows:

\[
Y_t = \Pi_{10} + \Pi_{110} \cdot (Y_{t-1} - Y_T) + \Pi_{111} \cdot (x_t - x_{T+1}) + \epsilon_{1t}^{[1]} \tag{4.45a}
\]

and

\[
Y_{t+1} = \Pi_{10} + \Pi_{210} \cdot (Y_t - Y_T) + \Pi_{211} \cdot (x_{t+1} - x_{T+1}) + \Pi_{212} \cdot (x_{t+1} - x_{T+2}) + \epsilon_{2t+1}^{[2]}
\tag{4.45b}
\]

Inspection of (4.45a) and (4.45b) reveals a recursive structure with complicated forms of nested cross-equation parameter restrictions. By analogy with the single-period policy model, we may either: (1) estimate the successive models sequentially, or (2) estimate both equations jointly—recognizing the parametric structure.

Approach (1) involves first estimating \( \Pi_{1}^{[1]} \equiv [\Pi_{10} \Pi_{110} \Pi_{111}] \) in (4.45a) as \( \hat{\Pi}_{1}^{[1]} = \hat{\Pi}_{11}^{[1]} \), using T-1 (incomplete) observations on (4.6). Then, given \( \hat{\Pi}_{110} \) and \( \hat{\Pi}_{111} \), we then may estimate \( \Pi_{2}^{[1]} \equiv [\Pi_{20} \Pi_{210} \Pi_{211} \Pi_{212}] \) via linear regressions for (4.45b), using T-1 (incomplete) observations. This procedure is the simplest way of imposing the nested cross-equation restrictions, while only using, linear regressions, sequentially, to retain the appropriate parametric form. Approach (2) involves iterative nonlinear optimization, and will not be discussed further.

---

1/ In discussion of multi-period policies our notation distinguishes between the vectors \( Y_T \) and \( x_{T+1} \) in \( z \). Also, for notational convenience only, we assume \( L-K=N \).

2/ An S-equation analogue follows the same sequential pyramid pattern.
We then form a quadratic loss function— for $S=2$,

$$Q_2^{[1]}(\eta_{T', x_{T+1}, x_{T+2}}, \theta) = \frac{2}{s=1} \sum_{i=1}^{N} \left[ \pi_{si}^{[1]} - \pi_{si}^{[1]}(Y_T, x_{T+1}, \ldots, x_{T+s}, \theta) \right]^2$$

and, as before, minimize $Q_2^{[1]}$, subject to $Q(\theta) = Q^{[0]}$. Upon convergence to $\theta_2^{[1]}$, the rank, $m_2^{[1]}$, of the appropriate Jacobian matrix, $J_2^{[1]}$, reveals whether or not the calibration is unique. Incomplete data are treated as before, and extensions beyond a 2-period policy horizon are straightforward. With the latter, it should be stressed, however, that despite repeated application of the chain-rule, only numerical first partial derivatives are required— albeit, sequentially, even when $S > 2$.

Finally, we note joint-expansions of both the $r$th-order of the approximation model and the $S$th policy horizon, as with $\theta_S^{[r]}$, are feasible, to the extent that $(r+1)^{st}$-order numerical derivatives can be calculated, where $r, S > 1$.

5. **Model Simplification and Respecification (Stage Three)**

We arrive at Stage Three, either: (1) directly from Stage One, with a calibration, $\theta^{[0]}$— containing $m^{[0]} < (M+M^A)$ functionally independent elements; or (2) from Stage Two, with the additional assumption of a given policy sequence, $\{x_{T+s}: s=1, \ldots, S\}$, and an $r$th-order approximation to the reduced form, yielding the calibration, $\theta_S^{[r]}$, and numerical rank, $m_S^{[r]} < M^A$. In either case, the computer "informs" the investigator that a unique structural calibration cannot be achieved on the basis of the present
use of available data, so that the model must be further restricted. In Stage Three we consider three possible procedures: (i) calibrate/estimate certain of the missing \( x_{tk} \) and/or \( y_{0i} \) values; (ii) fix the values of certain of the structural parameters in \( \theta \); or (iii) respecify the model, \( f \), in (3.1).

We briefly consider each of these, fully recognizing our discussion is not exhaustive (Bellman (1966b)).

5.1 Calibration/Estimation of Missing Data

All missing endogenous variables within \( Y \) are handled in the course of the model calibration process, using the missing data updating condition. However, there are \( M_0^A \) missing values within \( Y_0 \) and \( M_t^A \) within \( x_t \), \( t=1,\ldots,T \). Together, these result in the augmentation of \( \theta \) by the \( M^A \) additional elements within \( \theta^A \). If the rank, \( m \) (i.e., either \( m^0 \) or \( m^r \) ) exceeds \( M \), then, by "filling-in" the missing values within \( Y_0 \) and \( X \), it may be possible to uniquely calibrate \( \theta \), given these simplifying assumptions. As the model, (3.1), is designed to "explain" \( Y \), given \( Y_0 \) and \( X \), an extraneous deterministic/stochastic "model" to calibrate/estimate \( Y_0 \) and/or \( X \) must be adopted by the investigator.

We consider first \( X = (x_k) = (x'_k) = (x_{tk}) \). The simplest procedures involve methods of interpolation or extrapolation of each individual time series, \( \{x_{tk}: t=1,\ldots,T\} \), using the observed values \( (c_{tk}=1) \) to estimate those missing \( (c_{tk}=0) \). E.g., LaGrange interpolation polynomials could be employed. Depending upon the configuration of available

1/ Model simplification may also be desired if (say) \( m = (M+M^A) = (NT)^* \), leaving no "degrees of freedom" in the calibration of \( \theta \) and a singular \( \Sigma \).
data, parametric models may also be utilized. Univariate time series models (say) of the ARIMA form (Box and Jenkins (1970)), polynomial time trends; etc.; may be entertained.

An alternative is to adopt a multivariate (deterministic/stochastic) process for $x_t$. Suppose the $x_t$ are independent drawings from the normal distribution, $N(\mu, \phi)$. We then employ all available data within $\mathbf{X}$ to estimate $\hat{\mu}$ and $\phi$, as in

$$\hat{\mu} = \left(\mu_k\right), \quad \text{where} \quad \mu_k = \frac{1}{T} \sum_{T}^{t} c_{tk} x_{tk},$$

and

$$\hat{\phi} = \left(\phi_{k',k}\right), \quad \text{with} \quad \phi_{k,k'} = \frac{1}{T} \sum_{T}^{t} c_{tk} c_{tk'} (x_{tk} - \mu_k) (x_{tk'} - \mu_{k'}).$$

Then for a particular period, $t=1, \ldots, T$, upon reordering, suppose that $x_t' = [x(1)' , x(2)' ]$, where $x(1)$ is observed, but $x(2)$ is missing. Then we may estimate $x(2)$ by its conditional expectation, given $x(1)$, as in

$$\hat{x}(2) = \hat{\mu} + \hat{\phi}(2,1) \cdot (x(1) - \hat{\mu}).$$

This approach, in principle, may be generalized to vector auto-regressive processes, etc. These may either be calibrated (by the methods of sections 3 and 4) or estimated from incomplete data bases, where possible.

In most cases the investigator will simply retain the $A_0$ missing elements of $\mathbf{y}_0$ as additional parameters. However, if desired, one may also attempt to "fill in" a missing $y_{0i}$-value via backwards extrapolation from the observed $\{y_{ti}\}$, etc.1/

In both situations, the computer software should permit the investigator to make judicious a priori estimates of missing $y_{0i}$ or $x_{tk}$ values directly; as well as select from a pre-programmed taxonomy of systematic missing data "filling-in" options, with visual display capabilities. Each element, $y_{0i}$ or $x_{tk}$, which has been filled-in is inserted into $\mathbf{X}$ or $\mathbf{x}_0$, and the corresponding indicator variable, $I$.

---

1/ This approach is "not natural", in the light of (3.1).
\( c_{tk} \) or \( d_{01} \), is reswitched from 0 to 1. We thus return to Stage One with a reduced \( M^A \) value, and a new configuration of "available data."

5.2 A Priori Parameter Specification

An alternative device is to shrink the dimensionality, \( M \), of the parameter space in Stage Three, by fixing certain of the elements of \( \theta \), \text{a priori}. The investigator invariably has (point/stochastic) "priors" as to the values of certain parameters--often derived from studies of "similar" countries, etc.\(^1\) Other parameters will require formal calibration on the basis of the available data, given those values fixed in advance.\(^2\) We thus return to Stage One with a reduced value of \( M \), but, in this case, with the same configuration of data.

5.3 Model Taxonomy and Respecification

In cases where the rank resulting from experimentation with various missing data and/or \text{a priori} parameter specification options--each configuration requiring a separate pass through Stage One/Two--fails to indicate a unique parameter calibration with "plausible" results, the investigator must then entertain a "simpler" model than \( f(x^*_t, y_{t-1}, x_t, \theta) = 0 \) of (3.1) to "explain" \( y_t \), or any sub-vector of \( y_t \)-- adjusting for differences in the levels of aggregation.

The problem is to define what is meant by "simpler." The taxonomy that one adopts clearly involves a further \text{subjective} aspect. Not only must the investigator now adopt a mathematical representation of the way the "real

\(^1\) Zellner (1971). Stochastic priors may be treated by Monte Carlo methods.
\(^2\) See Hartley (1984a) for further discussion.
economy" works, but also we require a nested sequence of parametric specifications within a given taxonomy, ranging from simple to complex.

Given any particular "economic theory", we shall consider this problem as consisting of two aspects: (1) the "degree-of-approximation" of the various behavioral functions within a model specification of a given level, and (2) the level of the model specification of a given type. The former may be viewed instructively in terms of so-called "flexible functional forms"—see, e.g., Christenson, Jorgenson, Lau (1973), Diewert (1973), Wales (1977), Gallant (1981) or Barnett (1982, 1983, 1984). The latter is appropriately defined via the "level-of-aggregation" embodied in the model specification.

Clearly, with a fixed available data base, there are trade-offs between a more complex over-all model and the degree-of-approximation employed in modeling of each its behavioral functions.

One ingenious taxonomic representation of an economy of the CGE class is the so-called Transactions-Value (T-V) Approach to model specification (Drud, Grais and Pyatt (1983)). The approach rests upon the Social Accounting Matrix (SAM), developed, by R.A. Stone and his collaborators, as a framework to represent the monetary transactions embodied in a system of national accounts—see, e.g., Pyatt and Roe (1977), Pyatt and Thorbecke (1976) or King (1981). As there are few alternative taxonomic systems to date, we briefly describe the essentials of the T-V Approach to illustrate this phase of Stage Three.1/

A SAM is simply an $A^{th}$-order square matrix, $v_t = (v_{tij})$, where $v_{tij}$ denotes the value of a transaction between "account i" and "account j" in period $t=1,...,T$. The $v_{tij}$ may be unrestricted in sign: so that $v_{tij} > 0$

denotes an \( i:j \) gross payment, and, conversely, \( j:i \) when \( v_{tij} < 0 \). In most economies, one finds \( V \) is sparse, i.e., many elements, \( v_{tij} \) are identically zero. However, the associated accounting system must "balance", in the sense that the row sums, \( v_{t.i} = \sum_{j=1}^{A} v_{tij} \), and column sums, \( v_{t.j} = \sum_{i=1}^{A} v_{tij} \), are equated. Thus, for each account, \( i=1,...,A \), we require \( v_{t.i} = v_{t.i} \).

By suitable definition of the \( A \) accounts in a SAM, one may capture the basic macro-economic circular flow, as represented in the system of national accounts; and, further, "explained" by the basic Walrasian general equilibrium model of the behavior of households, firms and government (see Dervis, de Melo and Robinson (1982)). This may be done at various levels of aggregation.

The T-V Approach involves providing the investigator with a stored library of possible specifications, which may be associated with any not-identically-zero \((i,j)\) cell within \( V \). Some are purely monetary transactions

\[ 1/ \text{Alternatively, by selective transposition, all non-zero elements may be made positive. The choice depends upon whether one wishes to take advantage of the "positivity" or "triangularity" of } V \text{ in numerical calculations. This, however, involves "netting-out" the gross transactions and results in a "reduced form" representation of markets.} \]

\[ 2/ \text{In the case of LDCs, the reported transactions data, } \{v_{tij}\}, \text{ may not balance in a particular year. Byron (1978) suggests minimizing a quadratic loss function to locate the "closest" balanced solution to the observed data, prior to calibration of } \theta. \text{ In our approach, this is not necessary, in that the model equations contain the "balancing identities", which may or may not be satisfied by the officially reported data. Balanced accounts may then be obtained as a solution to the complete calibrated model, fixing (via constraints) all endogenous variables to be retained at reported values (Hartley (1984a)). Thus, we take account of the complete model (including the behavioral equations) and all available data to balance accounts.} \]

\[ 3/ \text{For example, Drud, Grais and Pyatt (1983) employ the following five basic accounts: (1) factors of production; (2) domestic institutions; (3) production activities; (4) commodities; and (5) the rest of the world. The degree of disaggregation entertained within each basic account then provides a taxonomy of models.} \]
(such as a lump-sum government transfer to households). Other transactions are the product of a price, \( p_{ti} \), and a quantity, \( q_{tij} \) (e.g., the expenditure of (the suitable aggregate of) households on a particular class of goods and services). We denote these by

\[
\nu_{tij} = \begin{cases} 

v_{tij}, & \text{if } (i,j) \in I_1 \\
p_{ti}q_{tij}, & \text{if } (i,j) \in I_2 \\
0, & \text{if } (i,j) \in I_3
\end{cases}
\]  

(5.1)

where the index sets, \( I_1, I_2 \) and \( I_3 \) are specified by the investigator in advance.\(^1\) In the case of \((i,j) \in I_2\), a nonlinear identity (see above), the row, \( i \), represents the value of demand, and the column, \( j \), the value of supply of a particular good, factor of production, etc. The T-V Approach permits: (1) various functional specifications for the \( v_{tij} \) (but, at present, not directly for the \( q_{tij} \))\(^2\), (2) permits the investigator to designate any particular \( v_{tij}, p_{ti} \) or \( q_{tij} \) as endogenous (an element of \( Y_t \)) or exogenous (an element of \( x_t \)); and, if the former, to (3) draw upon a library of explicit functions, \( f \), from within the general form:\(^3\)

\[
\begin{align*}
& f_k(Y_t, Y_{t-1}, \theta) = \nu_{tij} - f_k^*(V_t, V_{t-1}, Q_t, Q_{t-1}, P_t, P_{t-1}, Y_t, Y_{t-1}, \theta) = 0, \\
& f_\rho^*(X_t, X_{t-1}, \theta) = q_{tij} - f_\rho^*(V_t, V_{t-1}, Q_t, Q_{t-1}, P_t, P_{t-1}, Y_t, Y_{t-1}, \theta) = 0,
\end{align*}
\]

(5.2a)

\(^1\) The matter is analogous to whether one specifies an expenditure or demand/supply system. Equation (5.1) permits either. See, however, Section 6.2

\(^2\) This is for convenience, not generality.

\(^3\) Under a strictly economic definition.
This embeds the Walrasian general market equilibrium/disequilibrium model—the latter if \( p_{ti} \) is not specified by an equation, (5.2c), or is exogenous—within a system of national accounts. If \( p_{ti} \) is endogenous, and if not specified by an equation, (5.2c), then the corresponding \( \{q_{ti}\} \) and \( \{q_{t.i}\} \) are jointly determined by the "market system" in each period \( t \). Otherwise (with T-V), if exogenous or (5.2c) holds, either the demand-side or the supply-side is presumed known (to the investigator) to constrain—and determines the quantity, \( q_{tij} \), transacted—see Goldfeld and Quandt (1975). This leads to a so-called fix-price model—possibly with choices of multiple "regimes" in each period (Quandt (1984)). So-called "shadow prices" may then be calculated by relaxation of (5.2c), etc.

In such a formulation, one should presumably wish to choose from among "flexible functional forms" in determining (5.2a) or (5.2b)\(^{1/2} \)—both

---

1/ The T-V system library is, at present, somewhat restrictive. It implicitly assumes profit/utility maximizing behavior for firms/households under perfectly competitive conditions, and focuses upon the "elasticity of substitution" parameter, \( \sigma_{ij} \)—distinguishing between \( \sigma_{ij} = 0 \) (Leontief), \( \sigma_{ij} = 1 \) (Cobb-Douglas) and a constant \( \sigma_{ij} \) (C.E.S.) specification—see Drud, Grais and Pyatt (1983). This, however, is typical with present CGE specifications (Mansur and Whalley (1984)).

2/ This raises the question of where to introduce the functional specifications (Hartley (1981a, 1983b), Bellman and Hartley (1983/1984)).
subject to complicated (possibly implicit) cross-equation parameter restrictions. The level of the model specification may be adjusted by the degree of disaggregation permitted with respect to the model accounts.

Finally, it should be evident from (5.2a) and (5.2b) that such a CGE model is a special case of the dynamic model, (3.1).

In short, whatever model typology is adopted for

$$f(Y_t, Y_{t-1}, X_t, \theta) \in \{f^{(h)}(Y_t, Y_{t-1}, X_t, \theta): h=1, \ldots, H\}$$

(5.3)

in (3.1)--ranging from "simple" (h=1) to "complex" (h=H) models, the rank, m(h), attained, relative to (M(h) + M(h)), and given the configuration of (NT)* observations, determine whether a "step-up" or "step-down" through the model taxonomy is warranted, before returning to Stage One.

6. Conclusions, Extensions and Future Research

6.1 Conclusions

This paper has proposed a multi-stage systems approach to the problem of specifying and calibrating macro-economic models in the context of arbitrary configurations of incomplete data. As a vehicle to convey the basic ideas, we have employed the standard econometric model, consisting of a dynamic, nonlinear system of simultaneous behavioral equations and identities. We have considered both the customary stochastic representation and a deterministic analogue. In the former, incomplete data either require high-order numerical integrations, beyond present computer capabilities; or, in certain specific exceptions, iterative linearizations around the model solution values--effectively reducing the model to a deterministic form.
Accordingly, we have devoted most of our discussion to deterministic structural models, though a variety of special cases may be given specific stochastic interpretations—e.g., if the normal distribution is maintained for either the disturbances or the discrepancies.

Our systems approach may be broken down into three stages. In Stage One, starting from any feasible parameter vector, we generate a sequence of model solutions over the sample period, as well as the consequence of successive perturbations in the elements of that vector. This provides us with sufficient information: (1) to evaluate the (quadratic) loss function—involving the discrepancies between the model solution values and the observed actual values; and also (2) to determine a suitable direction in the parameter space (via the parameter updating condition) which will reduce the value of the loss function. The process iterates until convergence obtains.

Incomplete data on any of the endogenous variables are treated by our missing data updating condition—involving the iterative replacement of all missing values by our prevailing model solutions. Missing exogenous variables and initial endogenous variables are treated as parameters—thus augmenting both the dimension of the parameter vector and the relevant Jacobian matrix—here, the first partials of the solution vector with respect to the (possibly augmented) vector of structural parameters. Given the arbitrary configuration of missing data, the resulting Jacobian matrix may be rank deficient. We employ a Singular Value Decomposition, both to determine the numerical rank of the Jacobian and to calculate the Moore-Penrose generalized inverse, required in parameter updating.

1/ Perturbations are a way of extracting information about the model, \( f \), relative to a given data base.
A rank deficiency (upon convergence) in Stage One implies that, relative to the present use of available data, there is an equivalence class of parameter values which produces the same (minimum) value of the loss function. The investigator must therefore either: (1) proceed to Stage Two, in which he/she attempts to choose between alternative members of this class, by extracting further information (via different types of perturbations) from the available data, relative to the given model; or (2) go directly to Stage Three, in which the model itself is simplified.

In Stage Two we employ two alternative approximations to the reduced form of the deterministic structural model: (a) a Taylor series expansion of the restricted (structure-based) reduced form; and (b) an unrestricted polynomial approximation of the same order, here utilizing only knowledge of the relevant set of predetermined variables. The former requires an a priori choice by the investigator for the values of the predetermined variables that are to serve as the point of expansion. A "natural" choice is the policy vector of interest to the investigator. For any given order of approximation in the two respective reduced-form models, the Stage Two procedure is to select that parameter vector, from within the Stage One equivalence class, which brings the restricted and unrestricted reduced form parameter estimates as "close" as possible.

We have considered a basic Stage Two model involving a first-order approximation model and a single-period policy. This is shown to require the same calculations as in Stage One--the model simulation path over the sample period, and the effect of successive perturbations of each element in the prevailing parameter vector. In addition, we also require the information provided by evaluation of the model solution associated with the chosen policy vector; the effect of successive perturbations in each element of that policy.
vector; and the effect of joint perturbations in both the prevailing parameter vector and the policy vector. Incomplete data can be handled exactly as in Stage One. Extensions to higher-order approximation models and/or to multi-period policies are provided. An $r$th-order approximation model requires $(r+1)$-order numerical derivatives. Thus, the feasibility and cost of higher-order extensions varies with the state and price of computer technology.

Of particular interest, in the context of arbitrary configurations of incomplete data, is the fact that, at the completion of any phase of Stage One/Two, the computer will inform the investigator whether or not a unique calibration has been achieved, and whether the calibrated structural model may be utilized to predict the consequences of a given policy. In the former, there is an isomorphism between what we have termed structural calibrability in the deterministic case, and the notions of identifiability and estimability in the stochastic analogue—see the classic treatment by Fisher (1966) and the discussion in Leamer (1978). In the econometric literature, the concept of identifiability is viewed as a logical problem and prior to parameter estimation:

"(Consider) the situation arising in the very frequent case in which a structural equation to be estimated is part of a system...of equations all of which hold simultaneously. As we shall see, in such a case, the nature of the theoretical model to be estimated itself implies that the parameters of a given equation cannot logically be inferred on the basis of empirical data alone. Structural estimation is impossible without the a priori information concerning the equation to be estimated. Such information must be provided from a source outside the real-world data at hand—either from economic theory (©) or from the results of other studies of different types of data (©).... Without such additional information, structural estimation is a logical impossibility...no matter how extensive and complete those observations may be." (Fisher (1966, pp.1 and 2, parentheses added).

---

1/ Identification in the engineering/systems literature is usually equivalent, termwise, to estimation in the econometric literature.
We view the matter (locally) as an a posteriori rank deficiency in the appropriate Jacobian matrix, revealed to the investigator, ex post, through a Singular Value Decomposition (Golub (1968)).

The literature on identifiability is replete with discussions of special cases involving the so-called "rank and order conditions" of appropriate matrices. No general treatment (to my knowledge) of the problem is available for either local or global identifiability, and in many instances, the rank of the appropriate Jacobian matrix (e.g., Fisher (1966), ch.5 or Sargan (1983)) cannot be evaluated analytically. The result is that, in such cases, the investigator will not know a priori whether the model is identifiable, and whether it is uniquely estimable from incomplete data samples. This is an unsatisfactory state of affairs, and suggests the use of numerical methods—see also, Bellman, Gluss and Roth (1964), Bellman (1970a) and Hartley (1981b).

In contrast, with a deterministic model we are able to provide a complete taxonomy of qualitative outcomes with respect to calibrability of and prediction from structural models. We will not however be able to distinguish between a lack of calibrability due to incomplete data, logical "under-calibrability" (in Fisher's sense) or multi-collinearity—all of which may result in a rank deficiency in the relevant Jacobian matrix. In short, we have an identification problem as to the causes of a lack of calibrability.1/

It should be noted that our Stage One/Two methods require only that we be able to compute a solution to the specified model, given suitable values for the predetermined variables in any period and the unknown structural

1/ This is important for model respecification in Stage Three, and requires further investigation of possible numerical methods.
parameters. Our specific algorithms require various types of perturbations in either the prevailing parameter vector and/or the policy vector(s) employed as the point(s) of expansion in Stage Two (Bellman (1967)). Such marginal and/or joint perturbations are employed as numerical partial derivatives, wherever required by analytical arguments. The numerical accuracy and cost of our methods will therefore be constrained by the word-size, speed, immediate-access storage capacity and price of the prevailing computer technologies. Whether or not such perturbations are informative depends upon the model at hand and the data base available. It is therefore an empirical question, ex post.

Stage Three involves the interaction of the investigator with the model system and the available data base. In deterministic models, this involves three types of possible procedures. In the first, the investigator selects from a variety of pre-programmed options to "fill in" any of the missing values of the system's state (exogenous or initial endogenous) variables, utilizing the values that have been observed. This has the effect of reducing (or even eliminating) the dimension, $M^A$, of the augmented parameter vector $(\theta^A)$, associated with missing data in Stages One/Two.

In the second procedure, the investigator may elect to fix certain of the $M$ structural parameters in $\theta$, calibrating the remainder on the basis of available data and the above simplifying assumptions. Systematic

---

2/ For models of small dimension, a personal computer will already permit Stage One/Three systems approaches--particularly since the marginal cost of additional computation is near zero. In large problems, use of a "super computer" may be warranted in order to fully exploit all information via Stages One, Two and Three--particularly when data are scarce. See, however, Bellman (1967) or Bellman and Angel (1972).
experimentation with alternatives--iterating between Stages One/Two and Three --may suffice to produce a "satisfactory", uniquely calibrated model, \( f \), of (3.1), based upon the available configuration of data, and subject to the prescribed simplifying assumptions.

In the event that no satisfactory model emerges from such experiments, the investigator must then explore simplifications in the "scope" of the model specification, \( f \). We have distinguished between two simplification strategies: (1) a more parsimonious parametrization of given level; and (2) simplification in the overall level of complexity of the model. In the former, extending the basic ideas of the T-V Approach for CGEs (Drud, Grais and Pyatt (1983)), we propose a stored library of nested "flexible functional forms", differing in the degree-of-approximation (and, hence, the number of parameters)--ranging from "simple" to "complex" forms.

The level of overall model complexity we shall associate with the level of disaggregation adopted for (say): the industrial classification of sectors of production; types of factors of production (labor, capital, intermediates, etc.); types of households (urban, rural, etc.); types of goods and services; types of governments (central, regional, local); types of banks (central, other); etc. Suitable "step-up/step-down" procedures within a taxonomy of such models of a given type will depend on the data base at hand and the objectives of model construction. However, there is nothing, in principle, to preclude development of the appropriate computer software to treat CGEs and other types of macroeconomic models within our overall
6.2 Extensions and Future Research

We conclude by taking note of some potential extensions to our systems approach, and indicate some problems for future research.

First, we should stress that the model specification does not require that the behavioral functions in \( f \) of (3.1) be given as known, closed-form parametric expressions. Rather, we simply require a computable model solution, \( Y_t \), given \( Y_{t-1}, x_t \) and \( \theta \). For example, under a neoclassical economic theory, individual economic agents (or their aggregate) are often assumed to solve a constrained optimization problem in determining their current "economic" decisions. Under such neoclassical behavioral models, the observed sub-sector of (say) consumer demands, \( y^{(1)} \), may be modeled as the solution, \( y^{(1)*} \), to a utility maximization problem,

---

1/ In this connection, micro-level survey or census data, available at benchmark intervals, may be pooled with macro-level time series by defining a suitable taxonomy of households, firms, etc. and adopting a corresponding level of disaggregation in model specification—see Fox (1983), Hartley (1981b, 1984a) or Bellman and Hartley (1983/1984). This leads to so-called micro/macro general equilibrium models (Hartley (1980)). This is important for LDCs—particularly in Africa, as it represents a quick way of expanding the data base, without necessarily increasing the number of parameters in the overall system.

2/ With the introduction of various visual display capabilities (Tufte (1983)) at selected points in our Stage One/Two/Three systems approach, and with a systematic library of "control options" for model simplification or respecification, appropriate computer software would permit the investigator to play a challenging video game—in guiding the computer to a suitable, fully-calibrated model. Sound transforms may be similarly employed (Hartley (1984a)).

3/ A CCE determines only relative prices. To determine absolute price levels, we propose using the quantity theory of money (Friedman and Schwartz (1963)), as a "closure rule", while modeling the velocity of money circulation. See also Klein (1983), Hartley (1984b), Adomian (1984a) and Bellman and Cooke (1963).
subject to an income constraint, $y_t^{(3)*} = \mathcal{X}_t^{(1)*} - \mathcal{X}_t^{(2)*}$, and embedded within a larger (CGE) model; where $\mathcal{X}_t^{(1)*}$ and $\mathcal{X}_t^{(2)*}$ are model solution sub-vectors for quantities demanded and prices of consumption goods and services, respectively; and $y_t^{(3)*}$ denotes household disposable income--each a sub-vector of $\mathcal{X}_t^{*}$. The resulting solution for the joint model, and hence the demand system, $y_t^{(1)*} = f^{(1)*}(\mathcal{X}_t^{*}, \mathcal{X}_{t-1}^{*}, \mathcal{X}_t, \theta^{(1)})$, can be calculated numerically from a specification of the utility function, $U$, without requiring an analytic solution to specify $f(-)$ directly. This greatly expands the generality of possible neoclassical model specifications that can be accommodated econometrically. Profit maximizing firms may be modeled analogously.\footnote{The substantive difference is that profit (output) is observable, whereas utility is latent (Hartley (1981a, 1983b)).} Given a parametric specification of a production function, optimal values of all variable inputs, in principle, may be determined numerically, and jointly with any set of input/output prices, under profit maximization (Hartley (1983b), (Bellman and Hartley (1983/84)).

These so-called neoclassical econometric methods (see Hartley (1981a, 1981b) for a discussion of the analogous partial equilibrium models) propose direct estimation of the "deep parameters" in the underlying utility or production functions, instead of their subsequent manifestation in systems of
demand/supply functions, as is customary econometric practice. In the present (general equilibrium) setting, implementation involves embedding a further set of numerical optimization problems--one for each type of economic agent (or their aggregates)--within the fixed point algorithm to solve for $Y^*_t$ (Scarf (1973)). In addition, neoclassical econometric methods can be extended to model decision-making under uncertainty (e.g., Tobin (1965), Lucas and Sargent (1981), Hartley (1983b)), and multi-stage dynamic optimization models of behavior (e.g., Lucas and Sargent (1981), Wolpin (1982), Bellman and Hartley (1983/84) and Rust (1984)).

Second, there is no reason to restrict the model specification, $f$, to algebraic, discrete-time nonlinear systems of difference equations. A continuous-time model, evaluated at discrete intervals, $t=1,2,..., T$, would also suffice (Gandolfo (1981), Wymer (1979)). Thus, e.g., in principle, one may also calibrate systems of partial differential equations, provided a numerical solution can be calculated at each discrete $t$ (Bellman and Adomian (1984)). In addition, Adomian (1983, 1984a, 1984b) has also developed a so-called decomposition method for solving general types of stochastic systems, thus extending the prototype structural form (2.1).2/

---

1/ One attractive feature of neoclassical econometric methods is that all cross-equation restrictions are automatically imposed on the latter for general, well-behaved, but otherwise arbitrary specifications of the former. We simply solve the primal optimization problem by a computer algorithm embedded within the complete system. This, in turn, suggests the use of flexible functional forms to model the primal (utility, production, anticipations formation, etc.) functions, rather than the derived (demand and supply) system's functions. See Bellman (1963, 1970b), Hartley (1981a, 1983b).

2/ The combination of the decomposition method with our systems approach to calibration/estimation will be considered elsewhere. The most interesting feature of the former is the use of "stochastic operators".
The next class of extensions involves **domain constraints**—either: (1) on the parameter vector, \( \theta \in \Theta \), or (2) on the solution vector of endogenous variables, \( y_t^* \in R_t(y_{t-1}, x_T, \theta) \), a subset of \( R_N \). E.g., if \( \Theta \) is a known convex set, we may often employ **penalty functions** to avoid infeasible parameter values. Alternatively, **shrinkage** of the step-size parameter in the parameter updating condition is often recommended (Goldfeld and Quandt (1972)). Finally, we might treat minimization of the loss function, subject to \( \theta \in \Theta \), directly as a **programming** problem.

In the case of domain restrictions on \( Y_t \), the problem is more complex. The solution algorithm for \( Y_t^* \) then also becomes a programming problem. In some cases certain elements of \( Y_t^* \) are integer-valued; in others (strong or weak) inequality restrictions may be imposed upon certain elements; etc. These **domain constraints** occur, e.g., in binary/multinomial choice problems (McFadden (1974), Amemiya (1981)), or, e.g., as non-negativity (Tobin (1958), Bellman (1961), Hartley (1976), Hartley and Swanson (1980), Amemiya (1984)) or rationing (Deaton (1981)) restrictions on consumer demand systems. In many problems, a mathematical programming formulation will suffice (e.g., Drud (1985) or Wets (1983)). A more general approach is to employ deterministic/stochastic **dynamic programming** (Bellman (1957) or Bellman and Dreyfus (1962)) as a **general purpose tool** to calculate optimal and/or feasible (sub-optimal) \( Y_t^* \) values, \(^1/\) relative to any \( \theta \in \Theta \). Finally, we may employ a suitable definition of choice (4) for the weighting matrices, \( W_t \), in the quadratic loss function, \( Q(\theta) \), of (3.5), varying with \( t \) and

---

\(^1/\) Bellman (1966a).
depending on the programming problem at hand.\footnote{1}

An economic example of a restriction in which the domain, $r_{ti}$, of a particular variable, $y_{ti}^*$, is dependent on $X_{t-1}$, $X_t$, and $\theta$ occurs in so-called switching-regression or markets-in-disequilibrium models (see, e.g., Goldfeld and Quandt (1974) or Hartley (1977a, 1977b, 1978)). E.g., in the latter, prices are determined by a non-Walrasian (possibly exogenous) process, and the quantity transacted ($y_{ti}^*$) is the minimum of quantities demanded ($y_{tk}^*$) and supplied ($y_{tk}^*$) -- each separately determined, but not necessarily observed.\footnote{2}

Finally, and of considerable practical importance in realistic macro-economic models that are to be used for policy analysis, we note that an extension of the model specification, (3.1), to

$$f_{t}(y_{t}^{*},X_{t-1},X_{t},\theta) = 0$$  \hspace{1cm} (6.1)

where the function, $f_{t}$, may now change with each period, $t$, while preserving the same parameter vector, $\theta$, in each period, involves no

\footnote{1} We shall discuss the problem of multiple domain constraints in more detail elsewhere. Again, the multiple numerical integrations required in the general implementation of general stochastic models is, at present, problematic (Hartley (1983b)), and suggests, for the present, a deterministic analogue.

\footnote{2} Quandt (1984) has provided a comprehensive bibliography of the major contributions to the theory and econometrics of this problem. See, e.g., Fair and Jaffee (1972) or Hartley and Mallela (1977) for the stochastic single-market case and Ito (1980) for a discussion of stochastic multi-market disequilibrium models, with so-called "spill-over effects" and the attendant multiple numerical integrations. The lack of differentiability at certain unknown points, created by "regime switches" in a deterministic model poses interesting computational problems which will be considered elsewhere. A circular argument should be minimally sufficient. See also Bellman (1953, 1955, 1956a), Bellman, Kashef, Vasudevan (1972), Hartley, Hartley and Bowen (1979).
substantive modification in our calibration procedures. Notationally, we must simply recast the solution vector as

$$\hat{Y}_t = g_t(Y_{t-1}, X_t, \theta),$$  \hspace{1cm} (6.2)

with appropriate resubscripting elsewhere. In short, this implies that a different model, $f_t$, containing the same underlying parameters, $\theta$, in each time period can be treated as easily as the same maintained hypothesis throughout.

As governments change their policies, possibly in each period, individual economic agents (or their aggregates) will confront a changing mix of constraints, which will, in turn, generate changing functional forms of behavioral responses. E.g., consider a mix of certain government-fixed prices, export subsidies, import duties and selective quotas, minimum wages, interest rate ceilings, food and fertilizer price subsidies, progressive income and corporate profits taxes, and a given budget composition, as constituting a possible government "policy regime" for an LDC in a given year. These imply that individual economic agents will confront a different form of constrained optimization problem in any period as, at a minimum, the domain constraints on $Y_t$, i.e., $R_t$, may change with $t$.

It will be appreciated that working out analytic, closed-form solutions, representing the response vector of each economic agent to the particular policy mix confronted in each period is a formidable task for any specific functional form—much less in general, and with incomplete data. Obtaining the appropriate expression for the p.d.f., $p_t(Y_{1t} \mid Y_{1t-1})$ of (2.4) for each $t=1, \ldots, T$, to replace (2.3), is then necessary to calculate the MLE (Bellman (1970b), Bellman, Kashef, Lee, Vasudevan (1975)).
Our (deterministic) neoclassical econometric approach (Hartley (1981a, 1981b, 1983a, 1983b, 1984a), Bellman and Hartley (1983/84)) simply redefines the particular form of programming problem confronted (say) by the various types of households and firms each period. For example, suppose we assume the same utility/production functions each period, and assume utility/profit maximization as a "behavioral hypothesis" for households/firms. Then, the "deep parameters" in the system are invariant over time, while both the actual data, \( \{Y_t\} \), and the solution vectors, \( \{y^*_t\} \), satisfy the domain constraint, \( R_t(Y_{t-1},X_t,\theta) \), in each period, \( t=1,\ldots,T \). In short, while the deep parameter vectors and the primal (utility, production, anticipations formation, etc.) functions are being viewed (here) as being within the "eye of the hurricane", with respect to changing government policies and external conditions; while the decision functions of the typical economic agent are not. See Bellman and Wing (1975). In addition, having attempted to realistically capture the effect of past policy mixes in the course of calibrating a structural model (as sketched here) against observed consequences should enable better analysis of proposed future measures.

It should be evident from the above discussion that our systems approach is computer-intensive, and that certain extensions must await a further quantum-jump in computer technology--before they are both feasible and cost-effective (Bellman (1964)). However, as data are scarce with LDCs, we must economize--in our view by: (1) selecting an appropriate level of aggregation within a taxonomy of possible deterministic models, and (2) utilizing our multi-stage "systems approach" to extract the maximum information about any such model from any given configuration of available
data, with due regard for the present state and price of computer technology and the size of the budget.\(^1\)

We believe, at the present juncture, that more realistic deterministic econometric models may be calibrated, and take better account of the extensions in specifications noted above, than can be estimated using general forms of stochastic analogues. Stochastic features should then be introduced parsimoniously, and only where important.\(^2\) The problem of constructing realistic deterministic macroeconomic models is, at present, hard enough, without attempting, as yet, to introduce general stochastic complexities (Bellman (1971), Bellman, Soong and Vasudevan (1972). Regardless, the solution to the problem of the infinite lies in the discrete.\(^3\)

---

\(^1\) We therefore treat model building as an economic issue. See Bellman (1962), Bellman, Kashef, Lee and Vasudevan (1975).


\(^3\) However, "...it is very interesting to see if one can discern a pattern and to know how large a role chance played in one's life" (Bellman (1978)).
REFERENCES


(1967), Perturbation Techniques in Mathematics, Physics and Engineering (German translation), Munchen: R. Oldenbourg Verlag.


(1984), "Bibliography of Quantity Rationing and Disequilibrium Models," Department of Economics, Princeton University (available upon written request).


A systems approach to the calibration of economy-wide models with incomplete data