

ECONOMIC RESEARCH REPORTS

ON THE COMPUTATION OF ESTIMATORS
IN SYSTEMS WITH
IMPLICITLY DEFINED VARIABLES

by

Ingmar R. Prucha

and

M. Ishaq Nadiri

R.R. #87-40

November 1987

**C. V. STARR CENTER
FOR APPLIED ECONOMICS**



**NEW YORK UNIVERSITY
FACULTY OF ARTS AND SCIENCE
DEPARTMENT OF ECONOMICS
WASHINGTON SQUARE
NEW YORK, N.Y. 10003**

On the Computation of Estimators in Systems with
Implicitly Defined Variables

Ingmar R. Prucha* and M. Ishaq Nadiri**

September 1987

* Department of Economics, University of Maryland, College Park, MD 20742; National Bureau of Economic Research, Cambridge, MA 02138; and C.V. Starr Center for Applied Economic Research, New York University, New York, NY 10003.

** Department of Economics, New York University, New York, NY 10003; National Bureau of Economic Research, Cambridge, MA 02138; and C.V. Starr Center for Applied Economic Research, New York University, New York, NY 10003.

Abstract

Estimators are often defined as the maximizing values of some objective function. This note introduces an algorithm for the computation of such estimators for the parameters of a system of equations where some of the variables are implicitly defined by an auxiliary set of equations. Systems of equations of this kind have been considered in the recent factor demand literature as well as in other areas. The algorithm is a gradient procedure. To keep the computational burden manageable it calculates the gradient from analytic expressions rather than by numerical differentiation.

1. Introduction¹

Let y_1, \dots, y_T be some data process whose conditional distribution on x_1, \dots, x_T is characterized by the parameter vector θ . Estimators for θ are often defined as the maximizing values of some objective function, say $\mathcal{L}(\theta) = \ell(\theta, y_1, \dots, y_T, x_1, \dots, x_T)$. Those estimators are generally referred to as M-estimators (or maximum likelihood type estimators.) If an explicit analytic solution is not available, the maximizing value has to be found iteratively by numerical methods. For gradient methods the (trial) parameter value at the $i+1$ -th iteration, say θ^{i+1} , is typically calculated as follows:

$$(1) \quad \theta^{i+1} = \theta^i + \lambda^i [R(\theta^i)]^{-1} d(\theta^i),$$

where R^{-1} is a positive definite direction matrix, $d = \partial \mathcal{L} / \partial \theta$ is the gradient of the objective function and λ is the step length. Various alternative methods for determining R and λ have been suggested.²

The gradient d can, in principle, be calculated by numerically differentiating the objective function. Suppose θ is of dimension $1 \times P$, then the (approximate) computation of the gradient (as a vector of difference quotients) requires at each iteration step at least $P+1$ evaluations of the objective function. If the evaluation of the objective function is numerically expensive it seems desirable to provide an analytic expression for

¹We would like to thank Gary Anderson, Ernst Berndt, Bronwyn Hall, Harry Kelejian and Sam Ouliaris for helpful comments and discussions. We assume responsibility, however, for any remaining errors. This research was supported by the C.V. Starr Center's Focus Program for Capital Formation, Technological Change, Financial Structure and Tax Policy.

²For a nice review and discussion of numerical optimization routines from an econometric perspective see e.g. Berndt et al. (1974), Judge et al. (1985) and Quandt (1983).

the gradient so that the number of evaluations of the objective function can be kept to a minimum.

The numerical evaluation of the objective function of the full information maximum likelihood (FIML) or of generalized moment estimators is typically quite complex even for relatively small and standard systems of equations. In the following we consider a system of (possibly) nonlinear equations where some of the variables are unobserved. Those unobserved variables are defined as solution values of an auxiliary set of implicit equations which may themselves depend on unknown parameters. For this model each evaluation of the objective function of the FIML or generalized moment estimator requires a numerical solution of the auxiliary set of implicit equations in addition to the computations required for a standard model. Therefore, to keep the computational cost manageable, it seems essential to construct an algorithm where all derivatives can be calculated from analytic expressions. In the following we exemplarily introduce such an algorithm for the FIML estimator taking the algorithm put forward by Berndt, Hall, Hall and Hausman (1974) as a starting point.³

2. Model Specification and Recent Examples

Consider the following (possibly) nonlinear simultaneous equation system:

$$(2a) \quad g(y_t, z_t, w_t, w_{t-1}, \theta) = \epsilon_t, \quad t=1, \dots, T,$$

$$(2b) \quad s(y_t, z_t, w_t, w_{t-1}, \theta) = 0,$$

where g and s are $1 \times M$ and $1 \times L$ vectors of twice continuously differentiable

³A special case of the algorithm introduced in this note was put forward earlier in Prucha and Nadiri (1984); see also Prucha and Nadiri (1986).

real functions, y_t is the $1 \times M$ vector of observed endogenous variables, $z_t = (y_{t-1}, \dots, y_{t-k}, x_t)$ where x_t is a $1 \times K$ vector of exogenous variables, ϵ_t is a $1 \times M$ vector of i.i.d. normal disturbances with zero mean and variance covariance matrix Σ , and θ is a $1 \times P$ vector of model parameters. With w_t we denote the $1 \times L$ vector of unobserved variables implicitly defined by the auxiliary set of equations (2b). More specifically, we assume that (2b) can, in principle, be solved uniquely for w_t . Let $w_t = \phi(y_t, z_t, w_{t-1}, \theta)$ denote this solution of (2b), i.e. $s(y_t, z_t, \phi(y_t, z_t, w_{t-1}, \theta), w_{t-1}, \theta) = 0$. We assume that ϕ is twice continuously differentiable. (The algorithm will focus on the case where for practical purposes an explicit analytic expression for ϕ is not available or cumbersome to obtain). Recursive substitution yields $w_t = \psi_t(y_t, y_{t-1}, \dots, x_t, x_{t-1}, \dots, w_0, \theta)$; theoretically we can hence write model (2) in the following equivalent form:

$$(3) \quad f_t(y_t, \dots, y_1, x_t, \dots, x_1, \theta) = \epsilon_t, \quad t=1, \dots, T,$$

with $f_t(\cdot) = g(\cdot, \psi_t(\cdot), \psi_{t-1}(\cdot), \dots)$.⁴ We refer in the following to (2) and (3) as, respectively, the "unsubstituted" and the "substituted" form of the model.

For illustrative examples of the above model in the recent literature see, e.g., Epstein and Denny (1980) and Prucha and Nadiri (1986). In those papers special cases of model (2) are encountered within, respectively, the framework of a factor demand model where the utilization rate of capital is endogenous and within the framework of a dynamic factor demand model with a finite but shifting planning horizon.

⁴For ease of notation, initial values for y and x corresponding to $t \leq 0$ are incorporated in f_t . The initial value for w is assumed to be either observed or incorporated into θ .

We note that both of the above examples correspond to the case where the auxiliary system of equations (2b) is static in the sense that it does not depend on w_{t-1} . As a simple illustration of the case where (2b) is truly dynamic consider some system of factor demand equations that includes a demand equation for capital investment goods. Assume that capital evolves according to (*) $K_t - I_t - (1-\delta)K_{t-1} = 0$, where K_t and I_t denote the end of period capital stock and gross investment, respectively, and δ denotes the constant depreciation rate. Assume further that δ is unknown and is to be estimated from the data. Clearly (*) is now a special case of (2b) with, respectively, I_t , K_t and δ elements of the vectors y_t , w_t , and θ . We note further that also various rational expectations models can be viewed as examples of models of the form (2).

3. Estimation Algorithm

The algorithm introduced by Berndt et al. for the computation of the FIML estimates applies to standard nonlinear simultaneous systems of equations. It only requires first order derivatives of the likelihood function, has been implemented in the Time Series Processor (TSP) software package and performs well in applications. We note that only if the "substituted" form of the model were known could we use the Berndt et al. algorithm to compute the FIML estimates. In the following we now use the Berndt et al. algorithm as a starting point to derive an algorithm for the "unsubstituted" form of the model. More specifically, we introduce an algorithm that is expressed entirely in terms of the "unsubstituted" form of the model, but is equivalent to the Berndt et al. algorithm for the "substituted" form of the model in the sense that it computes the same sequence of (trial) parameter values.

The concentrated (conditional) log-likelihood function expressed in terms

of the "substituted" form of the model is given by:⁵

$$(4) \quad \mathcal{L}(\theta) = \sum_t \mathcal{L}_t(\theta), \quad \mathcal{L}_t(\theta) = -\frac{T}{2} \ln(2\pi) + \ln|\det[J_t(\theta)]| - \frac{1}{2} \ln \det[S(\theta)] \\ - \frac{1}{2} f_t' [S(\theta)]^{-1} f_t,$$

where $J_t(\theta) = \partial f_t' / \partial y_t'$ denotes the Jacobian of the transformation and $S(\theta) = (1/T) \sum_t f_t' f_t$. Since the last term in the expression for \mathcal{L}_t sums to TM we can write the concentrated log-likelihood function as:

$$(5) \quad \mathcal{L}(\theta) = \sum_t \mathcal{L}_t(\theta) = \text{const} + \sum_t \ln|\det[J_t(\theta)]| - (T/2) \ln \det[S(\theta)].$$

Its gradient is given by:

$$(6) \quad d(\theta) = \sum_t a_t(\theta), \quad a_t(\theta) = \frac{\partial \mathcal{L}_t(\theta)}{\partial \theta'} = \frac{\partial \text{vec}(J_t)}{\partial \theta'} \text{vec}(J_t^{-1})' - \frac{1}{T} \sum_t \frac{\partial f_t'}{\partial \theta'} S^{-1} f_t' \\ - \frac{\partial f_t'}{\partial \theta'} S^{-1} f_t' + \left\{ \frac{1}{T} \sum_t \frac{\partial f_t'}{\partial \theta'} [f_t' S^{-1} \otimes S^{-1} + S^{-1} \otimes f_t' S^{-1}] \right\} \frac{1}{2} \text{vec}(f_t' f_t).$$

We define furthermore:

$$(7) \quad R(\theta) = (1/T) \sum_t a_t(\theta) a_t(\theta)'$$

It then follows from Berndt et al. (1974) that the iteration scheme (1) with d defined by (6) and R defined by (7) converges to a stationary point, given the step length λ is computed in a certain reasonable way. For starting values reasonably close to the maximum of the likelihood function this iterative algorithm will hence converge to the maximum of the likelihood function. Let $\hat{\theta}$ be the FIML estimator. It then follows from Gallant (1987), ch. 6, that under appropriate regularity conditions $T^{1/2}(\hat{\theta} - \theta) \xrightarrow{i.d.} N(0, \Omega)$ and that upon convergence of the estimation algorithm R^{-1} is a consistent estimator for the

⁵Unless stated otherwise sums are always taken over $t=1, \dots, T$.

asymptotic variance covariance matrix Ω .⁶

Consider now the "unsubstituted" form of the model. In terms of this model the concentrated log-likelihood function is given by:

$$(8) \quad \mathcal{L}_*(\theta, w_*) = \text{const} + \sum_t \ln |\det[J_{*t}(\theta, w_t)]| - (T/2) \ln \det[S_*(\theta, w_*)],$$

where $J_{*t}(\theta, w_t) = \partial g'_t / \partial y'_t + [\partial w'_t / \partial y'_t][\partial g'_t / \partial w'_t]$, $S_*(\theta, w_*) = (1/T) \sum_t g'_t g'_t$ and $g_t = g(y_t, z_t, w_t, w_{t-1}, \theta)$ with $w_* = [w_1, \dots, w_T]$. The gradient is given by:

$$(9) \quad d_*(\theta, w_*) = \sum_t a_{*t}(\theta, w_*),$$

$$a_{*t}(\theta, w_*) = \left[\frac{\partial \text{vec}(J_{*t})}{\partial \theta'} + \frac{\partial w'_t}{\partial \theta'} \frac{\partial \text{vec}(J_{*t})}{\partial w'_t} + \frac{\partial w'_{t-1}}{\partial \theta'} \frac{\partial \text{vec}(J_{*t})}{\partial w'_{t-1}} \right] \text{vec}(J_{*t}^{-1})' -$$

$$\frac{1}{T} \sum_t \left[\frac{\partial g'_t}{\partial \theta'} + \frac{\partial w'_t}{\partial \theta'} \frac{\partial g'_t}{\partial w'_t} + \frac{\partial w'_{t-1}}{\partial \theta'} \frac{\partial g'_t}{\partial w'_{t-1}} \right] S^{-1} g'_t -$$

$$\left[\frac{\partial g'_t}{\partial \theta'} + \frac{\partial w'_t}{\partial \theta'} \frac{\partial g'_t}{\partial w'_t} + \frac{\partial w'_{t-1}}{\partial \theta'} \frac{\partial g'_t}{\partial w'_{t-1}} \right] S^{-1} g'_t +$$

$$\left\{ \frac{1}{T} \sum_t \left[\frac{\partial g'_t}{\partial \theta'} + \frac{\partial w'_t}{\partial \theta'} \frac{\partial g'_t}{\partial w'_t} + \frac{\partial w'_{t-1}}{\partial \theta'} \frac{\partial g'_t}{\partial w'_{t-1}} \right] [g'_t S^{-1} \otimes S^{-1} + S^{-1} \otimes g'_t S^{-1}] \right\} \frac{1}{2} \text{vec}(g'_t g'_t).$$

The derivatives of w_t with respect to y_t can be calculated as:

$$(10a) \quad \partial w'_t / \partial y'_t = - [\partial s'_t / \partial y'_t][\partial s'_t / \partial w'_t]^{-1}, \quad t=1, \dots, T,$$

⁶We note that model (3) is quite general in that the argument list of f_t may depend on t . We note further that the above algorithm differs slightly from that introduced by Berndt et al.: Consistent with the observation that the last term in the expression for \mathcal{L}_t sums to TM it is readily seen that the last two terms in the expression for a_t sum to zero. The gradient can hence also be calculated as: $d(\theta) = \sum_t b_t(\theta)$ with $b_t(\theta) = \{\partial \text{vec}(J_t) / \partial \theta'\} \text{vec}(J_t^{-1})' - (1/T) \sum_t \{\partial f'_t / \partial \theta'\} S^{-1} f'_t$. In the Berndt et al. algorithm R is essentially defined as $R(\theta) = (1/T) \sum_t b_t(\theta) b_t(\theta)'$. However under this definition R^{-1} is generally only a consistent estimator for the asymptotic variance covariance matrix of the FIML estimator if Σ is known; see Hatanaka (1978) and Prucha (1984) for a further discussion.

where $s_t = s(y_t, z_t, w_t, w_{t-1}, \theta)$. The derivatives of w_t with respect to θ can be calculated recursively as:

$$(10b) \quad \partial w'_t / \partial \theta' = - [\partial s'_t / \partial \theta' + (\partial w'_{t-1} / \partial \theta') (\partial s'_t / \partial w'_{t-1})] [\partial s'_t / \partial w'_t]^{-1}, \quad t=1, \dots, T,$$

with $\partial w'_0 / \partial \theta' = 0$. The above expressions are readily obtained upon differentiating (2b) implicitly. We define furthermore:

$$(11) \quad R_*(\theta, w_*) = (1/T) \sum_t a_{*t}(\theta, w_*) a_{*t}(\theta, w_*)'$$

Let w_t^i denote the solution values of (2b) for w_t corresponding to θ^i and let $w_*^i = [w_1^i, \dots, w_T^i]$. Then clearly $\mathcal{L}(\theta^i) = \mathcal{L}_*(\theta^i, w_*^i)$, $d(\theta^i) = d_*(\theta^i, w_*^i)$ and $R(\theta^i) = R_*(\theta^i, w_*^i)$. Consequently, we can express the above iterative algorithm for the "substituted" model equivalently in terms of the "unsubstituted" model. The i -th iteration step now consists of the following sub-steps:

Sub-step 1: Corresponding to θ^i solve $s(y_t, z_t, w_t^i, w_{t-1}^i, \theta^i) = 0$, $t=1, \dots, T$, recursively for w_t^i .

Sub-step 2: Corresponding to θ^i and w_t^i calculate $\partial w'_t / \partial y'_t$ and $\partial w'_t / \partial \theta'$, $t=1, \dots, T$, from (10). Subsequently calculate $\mathcal{L}_*(\theta^i, w_*^i)$, $d_*(\theta^i, w_*^i)$ and $R_*(\theta^i, w_*^i)$ from (8), (9) and (11).

Sub-step 3: Calculate $\theta^{i+1} = \theta^i + \lambda^i [R_*(\theta^i, w_*^i)]^{-1} d_*(\theta^i, w_*^i)$.

As a by-product of the algorithm we obtain again in R_*^{-1} a consistent estimator for the asymptotic variance covariance matrix of the FIML estimator.

Clearly, the above methodology of modifying an existing estimation algorithm for the "substituted" form of the model such that it can be evaluated from the "unsubstituted" form of the model extends to other objective functions and hence to other estimators.

References

- Berndt, E.K., B.H. Hall, R.E. Hall, and J.A. Hausman, 1974, "Estimation and inference in nonlinear structural models", *Annals of Economic and Social Measurement* 3, 653-665.
- Epstein, L.G, and M. Denny, 1980, "Endogenous capital utilization in a short-run production model: Theory and empirical application," *Journal of Econometrics* 12, 189-207.
- Gallant, A.R., 1987, *Nonlinear Statistical Models*, New York: Wiley.
- Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lütkepohl and T.-C. Lee, 1985, *The Theory and Practice of Econometrics*, New York: Wiley.
- Prucha, I.R., 1984, "On the estimation of the variance covariance matrix of maximum likelihood estimators in nonlinear simultaneous equation systems: A Monte Carlo study," Working Paper No. 84-14, Department of Economics, University of Maryland, September, mimeo.
- Prucha, I.R., and M.I. Nadiri, 1984, "Formulation and estimation of dynamic factor demand equations under non-static expectations: A finite horizon model," Technical Paper No. 26, National Bureau of Economic Research, Cambridge, revised March, mimeo.
- Prucha, I.R., and M.I. Nadiri, 1986, "Comparison of alternative methods for the estimation of dynamic factor demand models under nonstatic expectations," *Journal of Econometrics* 33, 187-211.
- Quandt, R.E., 1983, "Computational problems and methods," in Z. Griliches and M. Intriligator, eds., *Handbook of Econometrics*, Vol. I, Amsterdam: North Holland.