Nonlinear Regression, Forecasting, and Optimization

Hsieh Hsih-chia

謝　世　佳

（作者為本校經濟研究所兼任副教授）

摘　要

對於非線型聯立方程式，本文提供新的估計、預測與非線型最適控制方法。本文考慮餘差項所引起的估計誤差，並使用高斯西德反覆法與預測與糾正方法，所以可以預測總體經濟的最適政策指標。

ABSTRACT

This study provides a new algorithm for estimation, forecasting, and nonlinear optimal control with nonlinear simultaneous equations. This algorithm can be used to forecast the optimal macroeconomic policies by considering the unprecise estimates due to noises and by combining the Gauss-Seidel algorithm and the predictor-corrector method for both short-term and long-term forecasting.

1. Introduction

The two stage least squares method is often used to estimate the simultaneous equations and the iterative two stage least squares method is used to estimate for the forecasting period, since the lagged dependent variables will become endogenous variables even though they are predetermined variables within the sampling period (Johnston, 1973) by using the Gauss-Seidel algorithm. When the linear equations are ill-conditioned, we need to introduce the pivotal scaled Gauss elimination. In addition, we need to take care of the possible misspecification of the model. Due to omission of residuals, the regression equations could become inequality constraints. Although the linear equations can be solved as an unconstrained optimization, the system needs to be solved as a nonlinear optimization if some of the equations are nonlinear (Powell, 1981). The Gauss-Seidel algorithm, however, is a form of the first derivative. It neglects the impact of the second or higher order derivatives of nonlinear equations.

In previous studies on the accuracy of forecasts (e.g. Makridakis, 1983), by

using adaptive exponential smoothing and giving more weights to recent observations, Lewandowski's (1982) simultaneous equations are by far the best for longer forecasting horizons (six periods ahead or longer). The exponential smoothing method, however, suffers from the lack of an objective procedure for model identification (Gardner, 1985). For one period's forecasting, Lewandowski's method does clearly worse than most single equations, including combined forecasts and the Box-Jenkins time series model.

This study proposes an alternative method of estimating and forecasting. Section 2 presents a rationale for a new simulation technique and forecasting algorithm because by combining the structural model and the equations of growth rates of the same variables, the results of the simulations and forecasts are more robust and able to take into account the possible missispecifications of the structural model. In section 3, we present the nonlinear programming and nonlinear regression. Concluding remarks are in Section 4.

2. A New Modelling Algorithm

Although the Stein-rule estimator minimizes the predictive error, its coefficient is fixed and does not improve the forecasts of turning points. Since the residuals are often not independently and randomly distributed, if the residuals are regarded as a missing variable and forecasted, we can predict the turning points more accurately since the residuals may reflect both endogenous and exogenous business cycles and model misspecifications. The forecasts of the residuals of regression equations imply the forecasts of business cycles and other missing variables which have not been captured in the structural model.

Suppose a vector of dependent variables will be fitted with a system of equations, $f(b,x)$. We will estimate the coefficient matrix, b^* which minimizes the sum of the squares of the errors

$$\min \quad S(b) = \frac{1}{2}\Sigma\,(f(b,x) - y_i)^2$$

In this case we will approximately fit the data with a system of linear equations such as by using Taylor's expansion series or a linear regression. The structural model is assumed to be

$$Y = XA + e \qquad \text{in stead of} \qquad y = f(b,x) + e \tag{2.1}$$

where Y is an $(n \times 1)$ vector of observations of the dependent variable, X is an $(n \times k)$

matrix of observations for the k independent variables, A is a (k x 1) vector of the unobserved parameters, and e is an (n x 1) vector of the unobserved disturbance terms. The elements e(i) of e are assumed to be independently and normally distributed with a zero mean and a constant variance, var(e). The partitioned matrices of X and B can be written as

$$X = |X_1 X_2| \quad \text{and} \quad B = |B_1 B_2|$$

where X_1 is an $(n \times k_1)$ matrix; and X_2 is an $(n \times k_2)$ matrix; B_1 is a $(k_1 \times 1)$ vector and B_2 is a $(k_2 \times 1)$ vector; and $k_1 + k_2 = k$. However, due to ignorance, equation (2.1) cannot be perfectly specified and may be estimated as

$$Y = X_1 B_1 + v$$

where $v = X_2 B_2 + e$, i.e., the residuals, v, involve the missing variable X_2 and its coefficient B_2.

The correct estimate of coefficient B under ordinary least squares (OLS) is

$$B = (X'X)^{-1} X'Y$$

In the case of misspecification and multicollinearity over the sample period, we will use the Ridge estimator (Oman, 1984):

$$B^* = (X'X + k\,I)^{-1} X'Y \quad \text{where} \quad k \leqslant \frac{\text{var}(v)}{B'B} \text{ is a scalar.}$$

A large class of ridge estimators do not dominate the OLS estimator $\hat{B}$ uniformly in the mean square error sense, $\text{Min}(\hat{B}-\hat{B})'(B-B)$, provided $(X'X)$ is sufficiently ill-conditioned. Alternatively, if we intend to minimize the prediction error $\text{Min}(B^*-B)'X'X(B^*-B)$, we can use the Stein-rule estimator (Efron and Morris, 1973; Malinvaud, 1980; Milltelhammer, 1984)

$$B^* = (1 - \frac{k-2}{N-k+2} \frac{1-R^2}{R^2}) \hat{B} \quad \text{for } k > 2$$

While the Stein estimator has a smaller mean square error, it is not a significant improvement over the prediction of the unrestricted OLS (Judge and Bock, 1973, p. 305). The superiority of the Stein estimator is partly due to a larger number of explanatory variables, $k > 2$, and partly due to its low explanatory power, as shown in the coefficient of determination R^2 that is less than one. The coefficient estimate of B_1 may not be statistically significant since its variance $\text{var}(v)(X_1'X_1)^{-1}$ is relatively greater than the variance, $\text{var}(e)(X'X)^{-1}$, of the correct specification.

Under our new estimator, the residuals will be regarded as a missing variable and will be brought back into the equation:

$$Y = X_1 B_1 + v = X_1 B_1 + X_2 B_2 + e \tag{2.2}$$

or

$$Y = X_1 B_1 + v + u \qquad R^2 = 1.0$$

where the new residuals, u, will be close to zero. The new coefficient of determination (2.2) will be close to one, viz., $R^2 = 1.0$. Thus our new estimator (2.2) is equal to the above-mentioned Stein-rule estimator. However, the new variance, var(u) of Equation (2.2), is closer to the true variance, var(e), of the correct specification, (2.1), than that of the Stein-rule estimator. Under our new estimator, the Stein-rule estimator is equal to that derived by ordinary least squares (OLS):

$$B^* = B \text{ if } R^2 = 1$$

In addition, with $R^2 = 1.0$, ignorance of causality conditions does not damage the equality:

$$Y = XA \quad \text{and} \quad X = YB \quad \text{where} \quad A = \frac{1}{B}.$$

With this proof, the regression residuals can be forecasted with numerical algorithms since the residuals are often not independently and randomly distributed (i.e. not i.i.d.). This is a treatment of the decomposition of a function into the linear component and the nonlinear component (Zirrili, 1982).

After using least squares to estimate the coefficient values, B, we will use estimates B to forecast X by using the iterative recursive Gauss-Seidel algorithm as follows:

$$x_i^{(k+1)} = \frac{c_i}{a_{ii}} - \sum_{j=1}^{i-1} \frac{a_{ij}}{a_{ii}} x_j^{(k+1)} - \sum_{j=i+1}^{g} \frac{a_{ij}}{a_{ii}} x_j^{(k)} \tag{2.3}$$

where the upper subscript, k, denotes the k-th iteration, k = 0, 1, 2, 3, . . . On the basis of some initial fixed values $x_i^{(0)}$, the values of $x_i^{(k)}$ at the k-th iteration will be immediately substituted by the values of $x_i^{(k+1)}$ obtained at the k+1-th iteration from all preceding equations.

The necessary and sufficient condition for absolute convergence under the Gauss-Seidel algorithm is that its first derivative should be less than 1:

$$\frac{dX^{(k+1)}}{dX^{(k)}} = \left| \frac{\sum_{j=1}^{g} a_{ij}}{a_{ii}} \right| < 1 \tag{2.4}$$

or

$$| a_{ii} | \geqslant \sum_{j=1}^{g} | a_{ij} |$$

The system (2.3) can be solved or expressed in terms of eigenvalues:

$$X^{(k+1)} = MX^{(k)} + C \tag{2.5}$$

where M is the eigenvalue matrix of $AX - C = 0$ and C is a constant matrix. Alternatively, the condition of a stationary iteration method is that the matrix M of eigenvalues (or spectral radius) should be smaller than 1:

$$| M | \leqslant 1 \tag{2.6}$$

Thus the error of the solution X in the k+1-th iteration is equal to

$$dX^{(k+1)} = M\, dX^{(k)} \tag{2.7}$$

The error should become smaller and smaller as the iteration proceeds. In other words, equation (2.5) can be expressed in terms of previous iterative values:

$$X^{(k+1)} = M(MX^{(k-1)} + B) + B$$

$$X^{(k+1)} = M^k X^{(0)} + \frac{(1-M^k)}{1-M} B$$

since $M < 1$, as k approaches infinity, M^k becomes close to zero, and by the fixed point theorem, $X = (1-M)^{-1} B$, i.e., the value of x will stabilize itself toward a fixed point. This is called the stationary state.

The maximal error after the k-th iteration is close to zero

$$\left| X^{(k+1)} - X^{(k)} \right| < \frac{M^{k+1}}{1-M} \left| X^{(1)} - X^{(0)} \right| \tag{2.8}$$

At the infinite iteration, our estimate $X^{(k)}$ will approach a unique fixed point $X = X^{(k+1)}$.

As shown in equations (2.4)-(2.5), the Gauss-Seidel algorithm is based on the first-order derivative similar to Newton's, Euler's, or the Newton-Raphson equations:

$$X^{(k+1)} = X^{(k)} + h \frac{df(X)}{dX}$$

where k denotes the k-th iteration. The advantage of the quasi-Newton algorithm is that $x^{(k+1)}$ depends on only $x^{(k)}$ and F(x) so that bad effects from the previous iterations are not carried along. The disadvantage is that $x^{(k)}$ requires a good initial approximation to the correct solution, x^*. The Newton method has a super-linear rate of convergence. Its recent improvements, including Broyden's quasi-Newton method (Powell, 1983), concern the quadratic rate of convergence. They often lead to only the local minimization, and are applicable only to a smooth minimization function. When the function is not smooth, the solution is often not globally optimal.

Newton's method is written as

$$X^{(k+1)} = x^{(k)} + \frac{f(x)}{f'(x)} \tag{2.10}$$

or

$$x^{(k+1)} = x^{(k)} + \text{error correction}$$

where f'(x) can be approximated by a definite difference.

By further estimating the growth rates (or derivatives) of all variables, we can gradually improve the estimates by using a several-stage Gauss-Seidel procedure. For example:

Stage 1: estimate the second-order derivative of x, $\frac{d\dot{x}(t)}{dt}$

$$\dot{x} = (1 + \ddot{x})\,\dot{x}$$

where $\dot{x} = \frac{dx}{dt}$ and $\ddot{x} = \frac{d\dot{x}}{dt}$. If x is a logarithmic value, log x, their derivative, (d log x/d t), is a growth rate of x.

Stage 2: estimate the growth rate $\dot{x}$ of x,

$$x(t+1) = x(t)(1 + \dot{x})$$

Due to difficulty in differentiating discrete data, in stage 1 instead of deriving the first derivative of the function, we use ordinary least squares and estimate the linear equations for the growth rates for all the variables in the structural equations (2.2) and use the Gauss-Seidel algorithm to yield the initial values for the stage 2 iteration. In stage 2 we use the Gauss-Seidel algorithm to solve the structural equations per se. The number of iterations will be the same as the number of equations. If both the specifications of the structural model and the linear growth-rate equations are correct, both solutions and forecasts should be equal:

$$F(B,x) = \frac{dF(B,x)}{dx} = 0$$

Since we approximate the structural model with the linear growth-rate model, and since both models are estimated from the same feasible observations, the combined forecasts are the forecasts of the variable under the unconstrained optimization. If we wish to optimize or maximize the values of x subject to nonlinear constraints, we can estimate the optimal growth-rate equations based on the selected optimal historical experiences by means of the least squares method. They are the barrier function within the feasible field. Furthermore, we can introduce and estimate the objective function, thereby ascertaining the degree of optimization. With the last growth rates as initial values, the equations of growth rates will be simulated until the solutions converge, and these solutions will serve as the initial estimates for the forecasts of the structural model, as shown in stage 2 above. This will reduce the possibility of different initial values leading to different estimates or forecasts.

To prove that our algorithm approximates the Taylor expansion series precisely up to the second or higher order term, this several-stage algorithm can be expanded as a nested first-order recursive formula:

$$x(t+1) = x(t) + \frac{dx}{dt} = (1 + \dot{x}(t)\,(1 + r\frac{d\dot{x}}{dt}))\,x(t)$$

where r is the Newton-step size, and $r=1/2$ approximately. This algorithm reduces the high order equations into sets of first-order equations.

Furthermore, we can improve the forecasting algorithm for simultaneous equations as the predictor-corrector method does (King, 1984), as shown below, which follows stage 1 on the growth rate equations and stage 2 on the structural model:

Stage 3. $x(t+1) = x(t) + \frac{1}{2}(3\frac{dx(t)}{dt} - \frac{dx(t-1)}{dt})$

Stage 4. $x(t+1) = x(t) + \frac{1}{2}(\frac{dx(t+1)}{dt} + \frac{dx(t)}{dt})$

An advantage of this approach is that it need only one or two iterations.

3. The Constrained and Unconstrained Optimization

We will consider the optimization problem in which the nonlinear least squares are the unconstrained optimization, which minimizes errors. First, if the original problem should have a unique or finite optimal solution, it is feasible to combine the dual and primal problems into one problem, as shown below:

Max Ax
subject to $C^T u < c$

or

Min $C^T u$
subject to $Ax > b$

thus

$$A^T x - C^T u = 0 \qquad x \geqslant 0,\ u \geqslant 0$$

The system can be approximated and solved as linear equations under linear constraints by using linear programming. Zirrile (1982) combined the linear programming with the second-order Taylor expansion to solve the nonlinear optimization. A few problems remain unsolved.

Secondly, we can linearize the nonlinear programming by using Taylor series expansion about x:

Minimize $f(x)$

Subject to $q(x) = 0$

this problem will be approximated by a linear system:

$$x = x_0 + c$$

$$\text{minimize} \quad f(x) = f(x_0) + c\,\frac{df(x_0)}{dx} + s$$

$$\text{subject to} \quad g(x) = g(x_0) + c\,\frac{dg(x_0)}{dx} + s = 0$$

where s represents the error or the slack variable of the approximation to $g(x)$.

If the objective function is nonlinear and subjective to linear constraints, the cost function or the Hamiltonian can be minimized through projected Hessian updating algorithms and iteratively approximated at its local minimum. The value of Lagrangian multiplier can be eliminated in the computation process; A slack variable (or an arbitrary constant) is used to make the inequality of constraints become equal. Every time after one of the solutions is found, the values of the slack variable will be changed such that the algorithm can proceed to find the global minimal solution.

The Hamiltanian to be minimized is

Minimize $H = c(x) + af(x)$

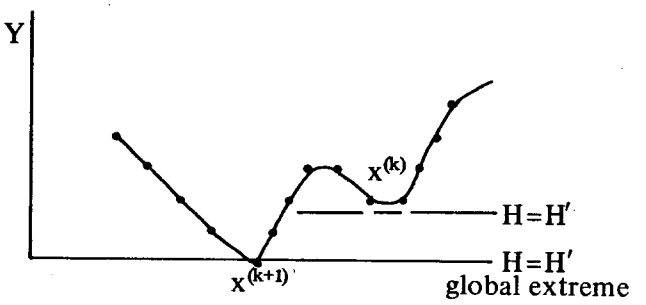

Fig. 1. Nolinear Programming with a Set of Linear Constraints

where a is a Lagrangian multiplier. If x^* is a local interior minimum for all neighboring points $x^* + dx$, then

$$H(x^*) \leqslant H(x^* + dx)$$

if H(x) is continuously differentiable, the right hand side of the last equation can be expanded in a Taylor expansion:

$$H(x^* + dx) = H(x^*) + \frac{dH(x^*)}{dx} dx + \frac{1}{2!} \frac{d^2H(x^*)}{dx^2} (dx)^2 + \ldots$$

To derive the first-order linear Euler's equation instead of the full Taylor expansion series, let

$$H(x^* + dx) = H(x^*) + \frac{H'}{1!} dx + \frac{H''}{2!} (dx)^2 + \ldots$$

at the minimum of H, $H(x+dx) = H(x)$ and both terms can be eliminated, and we have the Gauss-Newton equation. Alternatively, by optimizing and integrating over the time horizon, it becomes Euler's equation:

$$H = \frac{dH}{dx} + \ldots$$

If neglecting the second and higher order terms, it is the Euler's equation. Without integration, it becomes the Gauss-Newton equation:

$$x(t+1) = x(t) - 2\frac{H'}{H''}$$

or

$$x(t+1) = x(t) - rH'$$

where x is an optimal policy variable and r the Newton step size. Regardless of the boundary conditions of the time period and all variables, Euler's equation must be satisfied. This is similar to the first order term of the Taylor expansion series. Because the nondifferentiability of data, we use the finite difference as an approximation to differential equation:

$$\frac{df}{dx} = \frac{f(x(k+1)) - f(x(k))}{x(k+1) - x(k)}$$

when $x(k+1)$ is the solution of $f(x) = 0$, it becomes the Newton-Raphason equation:

$$x(k+1) = x(k) + f(x(k))/(df(x(k))/dx(k))$$

where k is an integer, $k = 1, 2, \ldots$.

Similarly, a linear feedback control function is

$$x(t) = gy(t-1) + q \tag{3.1}$$

where $x(t)$ is a vector of optimal policy variables, $y(t)$ is a vector of state variables; g is a vector of constant output feedback coefficients or the gains of control, and q is a vector of constant intercept.

The linear feedback function has the advantage of possible stability. It also has several weaknesses such that we need the use of both linear and nonlinear feedback functions. First, this feedback control equation is based on linearization about the optimal path. Different starting values used to perform the linearization will yield different feedback coefficients, g. Second, we may not know the controllability of $x(t)$ since we do not know whether the second derivative of H or y with respect to $x(t)$ is negative or positive or zero, i.e., we do not know the absolute concavity of $H(x)$. That is, there may be several local convergent solutions. Euler's equation with higher than first order derivatives may yield globally convergent solutions (Zirilli, 1982). Changes in the constraints of economic resources will change the parameters of the equations of motion through which the control variables could affect the growth rates of the state variables.

In nonlinear least squares or unconstrained optimization, Quasi-Newton methods and Marquadt's method have been used to minimize the square errors, R:

$$\text{minimize } H = \frac{1}{2} E\,(f(b,x) - y)^2 = \frac{1}{2} R(x)^T R(x)$$

The Gauss-Newton equation is

$$x^{(k+1)} = x^{(k)} - (J(x)^T J(x) + u)^{-1} J(x)^T R(x) \quad \text{for} \quad \frac{dH}{dx} = J(x)R(x)$$

where $\frac{d^2H(x)}{dx^2} = J(x)^T J(x) + R(x)\frac{d^2R(x)}{dx^2}$; the latter term is denoted by u

If the nonlinear function is not smooth, however, there could be multiple solutions and the solution yielded here may not be the global solution.

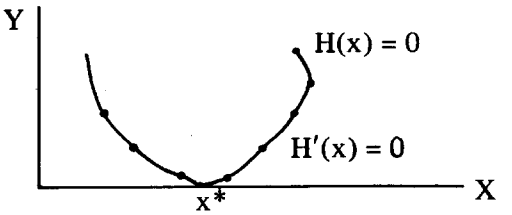

Fig. 2. The Unconstrained Optimization with an Unique Solution

For a nonlinear function with a nonlinear constraint, the constrained function can be approximated by linear programming (Karmarker, 1985), or by the projected Hessian updating algorithm and linear constraints (Nocedal, 1985), or by nonlinear programming (Mottl et al., 1978). If a nonlinear function is to be minimized subject to constraints, f, some of which are not actively constraining. We will use nonlinear programming, estimate the optimal feedback control function by the least squares method, and through the iterative solutions, we approximate the global optimum (including maximization or minimization) since the historical optimal experiences on the average are estimated.

Policy feedback control functions have several required properties: 1) passiveness: the policy passively and optimally reacts and gears the state variable to the target; 2) globally optimal and robust to changes in situations; 3) causality of the policy impact on the state variable; 4) existence and uniqueness of the optimal policy. During economic modelling, if solutions do not converge or are not unique, it is necessary to reduce the control cost by eliminating unimportant feedback lines.

Third, we deal with the nonlinear equations subject to nonlinear constraints

minimize $c(x)$ subject to $f(x) \leqslant 0$

Here, as an additional constraint, we add the first derivative of $c(x)$ as a barrier function:

$c'(x) + f(x) + s = 0$ where s is a slack variable.

subject to the feasibility surface:

$F(x) = e^{c(x)} + e^{c'(x)}$

where both the original surface, c(x), and the mean surface of the first derivative, c'(x), should lie on the feasibility surface and both can be estimated by least squares. After one intersection of c(x) and c'(x) has been found, say x, we take another derivative of c(x),

$$c''(x) = \frac{c'(x)}{(X - X^{(k)})} + s$$

where s is a remaining constant (King, 1984, p. 80-81), and try to find another intersection $x^{(k+1)}$ and gradually migrate to the global extreme. There is no line or grid search involved. Thus the problem is reduced to a sequence of problems, reestimate the approximate optimal feedback function each time when the new experimental result enters, and step by step reach the feasible and optimal domain, as shown in Fig. 3.

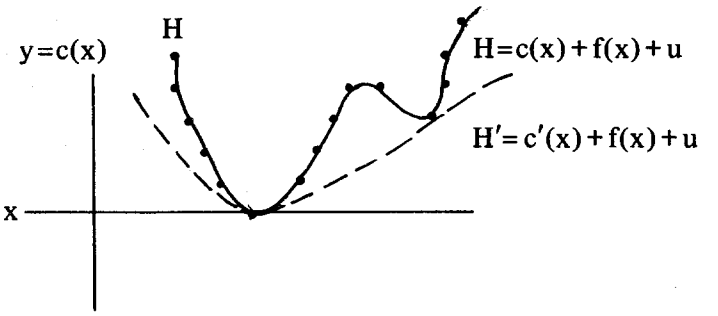

Fig. 3. Nonlinear Equations with Nonlinear Contraints

Euler's equation is a feedback control function and also called the law of motion through which the control variables can affect the state variables. The coefficient values of the control variables with favorable impacts should be increased and those with adverse impacts reduced.

We can rearrange the Euler equation into an optimal feedback control equation where the control variable is the dependent variable which can be estimated while the system parameters serve as fixed independent variables.

By a feedback control function, it means that the current policy, X(t), is a function of a series of past policies, X(t-i) and interacts with state variables, Y(t-i), where t=1, 2, . . . , n. By approximating the differential equations of motion with finite difference equations, we obtain

$$A(x) = B(y)$$

or

$$x(t) + a_1 x(t-1) + \ldots + a_m x(t-m) = y(t) + b_1 y(t-1) + \ldots + b_n y(t-n)$$

$$x(t) = G(x(t-1), x(t-2), \ldots, x(t-m), y(t), y(t-1), y(t-2), \ldots, y(t-n))$$

where x is the control variable, y the state variable or output variable, and G is a polynomial of the gains of control. $n \geqslant m$ is proper or controllable in the sense that the policy variable, x, can lead the state variable from y(t-n) to the target y(t).

In contrast to the traditional approach of the assumed linear feedback function, this study may directly estimate third-order feedback control functions and approximate the differential equations with difference equations and the intergral value by the sum over the time horizon. For example, let x, the control variable, be the growth rate of the money supply; y be the state variable which depends on its previous values y(t-i) and may not be observable at time t; and p be the price inflation rate which is called the output variable because it is observable at time t or because p is one of the leading indicators. Our feedback control functions are

$$x(t) = a_1 y(t) + a_2 y(t-1) + a_3 x(t-1) + \ldots .$$

and

$$x(t) = b_1 p(t) + b_2 p(t+1) + b_3 x(t-1) + \ldots .$$

where a and b are constant parameters, denoting gains of control.

By combining the above two equations, we can directly estimate feedback control functions in which those annual growth rates of real GNP selected here should exceed the average growth rate within the sample period. In order to estimate the optimal policy variable X(t), the optimal combined feedback function here is

$$x(t) = a_1 p(t) + a_2 p(t-1) + a_3 p(t-2) + a_4 x(t-1) + a_5 x(t-2) \ldots \text{for } y \geqslant \bar{y} \qquad (3.2)$$

where a_i, i=1, 2, . . . , 5 are constant; y is the actual annual growth rates of real GNP; and $\bar{y}$ is the average growth rate of real GNP in the sample period.

Lyapunov's stability condition of feedback control is that the sum of coefficients, c, for all lagged dependent variables, x(t-i), should be less than unity, since equation (2.2) can be rewritten as

$$\frac{dX(t)}{dt} = x(t) \quad \text{and} \quad X(t) = e^{Ax(t)}$$

where $|A| < 1$ denotes the sum of the constant coefficients, and p(t), t = 1, 2, and 3 are assumed given. However, according to our control theory, an unstable economic

system should be stabilized (i.e. offset or divided) by unstable control policy functions in such a way that after the resulting system becomes stabilized, the control policy will stabilize itself.

According to the values of the second-derivative in feedback functions, we can detect the controllability of the policies, as shown in the strict concavity of the state variable with respect to the control variable. That is, the first derivative should be zero and the second derivative negative:

$$\frac{dy}{dx} = 0 \quad \text{and} \quad \frac{d^2y}{d^2x} \leqslant 0$$

It is equivalent to Pontryagin's minimal principle, which is as follows:

$$H(x^*, p^*, y^*, t) \leqslant H(x, p, y, t) \qquad \text{for } y^* \geqslant \bar{y},\ x^* \text{ is optimal}$$

when there are multivariate state and policy variables, the elements of the matrix need to be in accord with the above-mentioned criteria.

Schittkowski (1982) suggests that a rough ranking of algorithms with respect to their efficiency is as follows: 1. quadratic approximation, quasi-Newton method; 2. generalized reduced gradient methods; 3. multipliers or Lagrangian methods; 4. penalty methods. This study combines these methods into a system of simultaneous equations and solves it by using both the line search and the gradient and minimizing the penalty objective function.

This study improves Mottl and Mottlova's (1978) work by linearizing the penalty function. The constrained nonlinear stochastic programming is transformed into an unconstrained one. Linear programming, especially Euler's method, could become an unstabilizer due to the error of the first-order derivative (Mottl, 1978). To solve the simultaneous equations, the identification condition required here is that the rank of the coefficients of endogenous and exogenous variables should be no less than the number of equations since the number of unknowns in the sampling period equals the rank of the regression coefficients, B, and in the forecasting period, it equals the rank of forecasted observations, X.

4. Concluding Remarks

Due to the impact of noises or regression residuals, we cannot estimate the coefficients or variables precisely, and we improve the estimates through forecasts of noises. Next we improve the Gauss-Seidel algorithm for both the structural model and the growth rate equations, and use the predictor-corrector method to improve

forecasts. The accuracy of forecasts is much more reliable. Finally we present the nonlinear regression and optimization with the barrier function as the growth-rate equations, which is estimated with regression equations or based on selected optimal experience after reducing the adverse impact on penalty. The solution will be an approximation to the global optimal solution; this solution should be iterated.

References

Efron, B. and C. Morris, March 1973, "Stein's Estimation Rule and its Competitors-An Empirical Bayes Approach", *Journal of the American Statistical Association*, 68(341), p. 117-130.

Gardner, E.S., 1985, "Exponential Smoothing: The State of the Art", *Journal of Forecasting*, 4(1), 1-28.

Gupta, G.K., R. Sacks-Davis and P.E. Tischer, March 1985, "A Review of Recent Developments in Solving ODEs", *Computing Surveys*, 17(1), 5-47.

Hald, J. and K. Madsen, February 1985, Combined LP and Quasi-Newton Methods for Nonlinear Optimization, *SIAM J. Numerical Anal.*, 22(1), 68-80.

Hall, J. 1979, "The Gauss-Seidel Algorithm and its application under International Price Uncertainty: The Case of Senegal", *American Agricultural Economics Association*, 188-198.

Johnston, H.N., L.R. Klein and K. Shinjo, 1974, "Estimation and Prediction in Dynamic Econometric Models", in W. Sellekaets edit., *Econometrics and Economic Theory*, (New York: International Arts and Sciences Press., Inc.), 27-56.

Judge, G.G. and M.E. Bock, 1978, *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*, (New York, North-Holland Publishing Company).

King, J.T., 1984, *Introduction to Numerical Computation*, (New York, The Southeast Book Company), 80-81.

Lewandowski, R., 1982, "Sales Forecasting by FORSYS", *The Journal of Forecasting*, 1, 205-214.

Makridakis, S., 1983, "Forecasting: State of the Art", in Makridakis S. edit., *The Forecasting Accuracy of Major Time Series Methods*, (New York: John Wiley & Sons, Co.,), 1-17.

Malinvaud, E., 1980, *Statistical Methods of Econometrics* (Oxford: North-Holland Publishing Company), 3rd edition, p. 244.

Milltelhammer, R.C., 1984, "Restricted Least Squares, Pre-Test, OLS, and Stein-Rule Estimator", *Journal of Econometrics*, 25, 151-164.

Mottl, J. and L. Mottlova, August 1978, "Description of a program for solving problems nonlinear programming", *The Computer Journal*, 21(3), 263-269.

Oman, S.D., 1984, "A Different Empirical Bayes Interpretation of Ridge and Stein Estimators", *J. R. Statis. Soc. B*, 46, No. 3, 544-557.

Powell, M.J.D., 1981, *Nonlinear Optimization*, (New York: Academic Press).

Peck, E. Anne, 1977, *Selected Writings on Future Markets*, Vol. II, Chicago Board of Trade.

Schittkowski, Klaus, 1981, "The Current State of Constrained Optimization Software", in Powell (1981) edit., *"Nonlinear Optimization"* (New York: Academic Press).

Zirilli, F., August 1982, The Solution of Nonlinear Systems of Equations by Second Order Systems of O.D.E. and Linearly Implicit A-Stable Techniques", *SIAM J. Numer. Anal.*, 19(4), 800-815.