Journal of Global Optimization **20:** 273–295, 2001.

Global Optimization of Econometric Functions

MAX E. JERRELL and WENDY A. CAMPIONE
Northern Arizona University, 70 Mc Connell Circle, Flagstaff, AZ 86011-5066, USA
(e-mail: max.jerrell@nau.edu and *wendy.campione@nau.edu)*

Abstract. Estimating the values of the parameter estimates of econometric functions (maximum likelihood functions or nonlinear least squares functions) are often challenging global optimization problems. Determining the global optimum for these functions is necessary to understand economic behavior and to develop effective economic policies. These functions often have flat surfaces or surfaces characterized by many local optima. Classical deterministic optimization methods often do not yield successful results. For that reason, stochastic optimization methods are becoming widely used in econometrics. Selected stochastic methods are applied to two difficult econometric functions to determine if they might be useful in estimating the parameters of these functions.

Key words: Econometrics, Maximum likelihood estimation, Stochastic optimization methods

Abbreviations: GA, genetic algorithm; SA, simulated annealing; NR, numerical recipes; ES, evolutionary strategies; AS, applied statistics

1. Introduction

Least-squares and maximum likelihood are the two most frequently used methods of estimating the parameters of econometric models. It is necessary to find the global optima of these functions to successfully meet the least squares or maximum likelihood criteria.

This paper briefly discusses some examples of the problems that arise in nonlinear estimation in Section 2. These difficulties have encouraged economists to seek optimization techniques other than deterministic methods. One set of techniques which promise to mitigate some of these difficulties are random search methods. The random search methods used in this research are discussed in Section 3. The methods are applied to two econometric models which are known to be difficult to estimate. These applicatons are discussed in Sections 4–6. Conclusions are presented in Section 7.

2. Why global optimization is useful

Estimating economic models requires finding the global optimum of a nonlinear function. Suppose the economic model can be represented by

$$y_t = f(\boldsymbol{x_t}, \boldsymbol{\beta}) + e_t$$

where y_t is an observation of a dependent variable, $\mathbf{x_t}$ is a vector of explanatory variables, and $\boldsymbol{\beta}$ is a vector of parameters. The term e_t represents the sampling error for observation t. The problem is to obtain estimates of $\boldsymbol{\beta}$ so that the resulting model best predicts the behavior of whatever entity the model represents. The two estimation methods most frequently used are least-squares and maximum likelihood. The least squares estimator $\mathbf{b}$ is chosen to minimize

$$S(\boldsymbol{\beta}) = [\mathbf{y} - \mathbf{f}(\mathbf{X}, \mathbf{b})]'[\mathbf{y} - \mathbf{f}(\mathbf{X}, \mathbf{b})]$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_T)'$, $\mathbf{X}' = (\mathbf{x_1}, \mathbf{x_2}, \ldots, \mathbf{x_T})$, $\mathbf{f}(\mathbf{X}, \mathbf{b}) = \Big(f(\mathbf{x_1}, \mathbf{b}), f(\mathbf{x_2}, \mathbf{b}) \ldots f(\mathbf{x_T}, \mathbf{b})\Big)$, and the data consists of T observations.

The purpose of maximum likelihood estimation is to find the parameter estimates that give the highest probability of generating the observed sample. Generally it is assumed that the error terms are normally distributed with mean zero and variance σ^2 such that the vector of error terms $\mathbf{e} = (e_1, e_2, \ldots, e_T)' \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ where $\mathbf{I}$ is an identity matrix. The likelihood function is given by

$$l(\mathbf{b}, \hat{\sigma}^2) = \frac{1}{(2\pi\hat{\sigma}^2)^{T/2}} \exp\left\{-\frac{[\mathbf{y} - \mathbf{f}(\mathbf{X}, \mathbf{b})]'[\mathbf{y} - \mathbf{f}(\mathbf{X}, \mathbf{b})]}{2\hat{\sigma}^2}\right\}$$

and the estimation problem is to find values of $\mathbf{b}$ and $\hat{\boldsymbol{\sigma}}^2$ that maximize $l(\mathbf{b}, \hat{\boldsymbol{\sigma}}^2)$. In either the least squares case or the maximum likelihood case, the global optimization value in required to meet the stated optimization criterion.

Nonlinear econometric functions often have features that make it difficult to find the global optimum. Multiple optima or flat surfaces are common. The derivatives of such functions can often be complex and difficult to derive. Numerical derivative methods can fail due to truncation and round–off errors arising from the use of floating point arithmetic. In some cases the function may not differentiable. Deterministic optimization methods will not be very effective in such situations and can fail completely. Random search methods do not require derivative information and are often used to replace deterministic methods in these cases. In this research, we investigate how well random search methods can estimate two difficult econometric functions.

3. The algorithms

This section presents the basic algorithms used in this research. The problem will be to find the parameter values, $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ that minimize a function $f(\mathbf{x})$. Each of the algorithms uses a random process to search in the parameter space. The search is not totally random, however. Each algorithm is designed to concentrate the search in promising regions where it seems likely that a minimum exists. Portions of the search will be conducted outside those regions to try to avoid becoming trapped at a local rather than global maximum.

Two of the methods used, a genetic algorithm and an evolutionary strategy, are based on a biological model. A third method, simulated annealing, is based on an analogy from physics. These are discussed next, beginning with the genetic algorithm.

3.1. GENETIC ALGORITHMS (GAS)

The description of a basic GA follows Schwefel (1995). The steps of the algorithm are briefly discussed first and then are expanded immediately afterwards.

Step 0: (Initialization)

A given population consists of λ individuals. Each individual consists of n genes, that determine the vitality, or fitness, for survival. Each gene is represented by a (binary) bit string that can be decoded into a coordinate value of a parameter. Fitness is the value of the function at the point represented by the gene.

Step 1: (Selection)

Two individuals are selected for reproduction with probabilities proportional to their relative fitness in the current population.

Step 2: (Crossover)

Two offspring are produced by combining the genes of the parents. One of these offspring will be chosen (at random) to join the next generation.
Steps 1 and 2 are repeated until λ individuals represent the next generation.

Step 3: (Mutation)

The individual bits of the genes may undergo further modification. Each bit is assigned a small probability of reversing value.

Steps 1–3 are repeated for a given number of generations. The process is terminated at that time unless some prior termination criteria has been met. One prior termination criterion often used is to terminate if the improvement in the best fitness value between successive generations is less than a specified value.

The individual steps are now discussed in greater detail. A point $\mathbf{x} = (x_1, x_2, \cdots, x_n)$ is represented by a binary bit vector in a GA called an individual. Each individual is characterized by a number of genes where each gene corresponds to a coordinate value of the point. The bit vector representation of the coordinate values requires that the coordinate values be constrained to a range

$$u_i \leqslant x_i \leqslant v_i.$$

A gene is represented by a bit vector of length l so that

$$a_i = (a_{i1}, a_{i2}, \cdots, a_{il})$$

represents the i–th gene. The coordinate value, x_i, of a parameter is determined by a decoding such as

$$x_i = u_i + \frac{v_i - u_i}{2^k - 1} \sum_{j=1}^{l} a_{i,j} 2^{j-1}.$$

The individual is constructed by concatinating the genes into a single bit vector

$$a_1 a_2 \cdots a_n$$

of length nl.

3.1.1. *Selection*

This section discusses how individuals are selected to become members of the next generation.

1. Compute the fitness of each individual k to be $f_k = -f(x_1, x_2, \cdots, x_n)$. (assume for convenience that all fitness values f_k are positive, else the fitness values must be scaled).
2. Compute a fitness value for the entire population
$$F = \sum_{j=1}^{\lambda} f_j.$$
3. Determine the probability that individual k will be selected for breeding
$$p_k = f_k / F$$
4. Determine the cumulative probability for each individual in the population $(j = 1, \cdots, \lambda)$.
$$q_j = \sum_{i=1}^{j} p_i$$
5. Select a random variable, r, from a uniform distribution $U(0, 1)$.
6. If $r < q_1$ select individual 1, else select the individual for which $q_{k-1} < r < q_k$ where $2 \leqslant k \leqslant \lambda$. This process gives a high probability of selection for individuals with good fitness values, a moderate probability of selection of those individuals with average fitness values, and a low probability of selection for individuals with poor fitness values. Note that this will allow an individual to be selected more than once. Successful individuals will tend to survive and reproduce while others tend to die out.

3.1.2. *Crossover*

Certain individuals of the new generation are selected for reproduction (crossover). The crossover process changes the bit vectors for the individuals chosen for mating. This process moves points through the parameter space. A description of the crossover process follows:

1. For each individual in the population, do the following.
2. Randomly determine if a individual will be selected for mating. Let the probability of crossover be p_c. Select a random value r from a uniform distribution $U(0, 1)$. If $r < p_c$, the individual is selected for mating.

Randomly pair the individuals selected for mating (if a odd number have been selected either drop one of those selected or add an additional individual to the list of selected individuals). Suppose that one of these pairs is

$$(a_1, a_2, \cdots, a_{nl})$$

$$(b_1, b_2, \cdots, b_{nl})$$

Select a random integer *pos* from the integers $[1..nl - 1]$ and form two new individuals

$$(a_1, a_2, \cdots, a_{pos}, b_{pos+1}, \cdots, b_{nl})$$

$$(b_1, b_2, \cdots, b_{pos}, a_{pos+1}, \cdots, a_{nl})$$

3.1.3. *Mutation*

Mutation is the other operation that causes changes in an individuals bit pattern. Mutation may occur in individuals not selected for mating.

1. For every individual in the population do the following.
2. Every bit in an individual has a probability p_m, usually very small, of mutating. Select a random value r from a uniform distribution $U(0, 1)$. If $r < p_m$, switch the value of the bit. Repeat for every bit in the individual.

Dorsey and Mayer (1995) examined the performance of several optimization methods on a number of functions, including econometric functions. They conclude that stochastic methods tend to perform better than deterministic methods on difficult econometric functions. They considered a GA and a simulated annealing algorithm in their research. One of the econometric functions they examined, a disequilibrium model, is also considered in this paper. The data for the disequilibrium model and the code for the GA can be found at
http://www.bus.olemiss.edu/dorsey/dorsey.htm. The simulated annealing algorithm they used is that of Goffe et al. (1994) which is discussed in Section 3.3.

3.2. EVOLUTION STRATEGIES (ESS)

The simplest ES strategy is one where a single parent produces a single offspring and the individual with the best fitness value is the one that survives to the next generation. This is designated as a $(1 + 1)$ strategy. More elaborate strategies use μ parents to produce λ offspring. The $(\mu + \lambda)$ strategy creates a new population of $(\mu + \lambda)$ individuals from which μ individuals with the best fitness values survive. In this case it is possible for an individual to survive for a number of generations. In the (μ, λ) strategy $(\lambda > \mu)$ offspring are produced but the next generation of μ individuals is selected from the λ offspring with the best fitness values. In this case no individual survives to the next generation. Unlike the GA, each of the μ parents is given an equal probability of mating. The selection mechanism of the ES is in the selection process of the individuals that survive to the next generation not in the selection process for mating. This research used $(\mu + \lambda)$ and (μ, λ) strategies.

The evolutionary strategy represents an individual as a pair of floating point vectors $(\boldsymbol{x}, \boldsymbol{\sigma})$. The vector $\mathbf{x}$ is a point in the search space and $\boldsymbol{\sigma}$ is a vector of standard deviations. Movement from point $\mathbf{x}^{t+1}$ to $\mathbf{x}^t$ is generated by

$$\mathbf{x^{t+1}} = \mathbf{x}^t + \mathrm{N}\,(\mathbf{0}, \boldsymbol{\sigma})$$

where $\mathrm{N}\,(\mathbf{0}, \boldsymbol{\sigma})$ is a vector of independent normally distributed random numbers with standard deviations $\boldsymbol{\sigma}$. Like GAs, ESs also incorporate crossover and mutation.

3.2.1. *Crossover*

Two individuals are randomly selected for mating

$$(\mathbf{x^1},\ \boldsymbol{\sigma^1}) = \left((x_1^1, \cdots, x_n^1)(\sigma_1^1, \cdots, \sigma_n^1)\right) \text{ and}$$
$$(\mathbf{x^2},\ \boldsymbol{\sigma^2}) = \left((x_1^2, \cdots, x_n^2)(\sigma_1^2, \cdots, \sigma_n^2)\right)$$

There are two types of crossover

- create a new offspring

$$(\mathbf{x},\ \boldsymbol{\sigma}) = \left((x_1^{q_1}, \cdots, x_n^{q_n})(\sigma_1^{q_1}, \cdots, \sigma_n^{q_n})\right)$$

  where $q_i = 1$ or 2 with equal probability, or
- Create a new offspring with

$$(\mathbf{x},\ \boldsymbol{\sigma}) = \Big(((x_1^1 + x_1^2)/2, \cdots, (x_n^1 + x_n^2)/2),$$
$$((\sigma_1^1 + \sigma_1^2)/2, \cdots, (\sigma_n^1 + \sigma_n^2)/2))\Big)$$

3.2.2. *Mutation*

The offspring undergo mutation through

$$\boldsymbol{\sigma}' = \boldsymbol{\sigma} \exp\Big(N(0, \Delta\Big)$$
$$\mathbf{x}' = \mathbf{x} + N(0, \boldsymbol{\sigma}')$$

where Δ is a factor that is adjusted during execution to expand or contract the effective search region.

These steps are repeated for a given number of generations or until a termination criterion is met. A possible termination criterion could be the following: suppose that $F(w)$ is the worst fitness value in a generation and $F(b)$ is the best value. Terminate the process if $|F(b) - F(w)| < \epsilon$ for a minimization problem.

One global optimization method used in this study was the evolutionary strategy of Schwefel (1995). This optimization methodology is similar to GAs but does not seem to be as well known, particularly in the United States. It was developed as a method for continuous optimization problems while GAs developed as a general optimization technique Michalewicz (1996). An immediate distinction between the two is that ESs use floating point vectors where classical GAs use binary vectors.

The code for this algorithm is contained in a disk accompanying Schwefel's book and a hardcopy version of the code is presented in the book itself.

3.3. SIMULATED ANNEALING (SA)

This optimization technique relies on analogy from physics rather than the biological sciences. The atoms of liquid metal move about freely but tend to lose this mobility as the metal is cooled. If the cooling process occurs slowly the atoms will tend to align with one another producing a stable piece of solid metal (the metal will be in a minimum energy state). If the cooling takes place too quickly the result will be either an amorphous or a brittle piece of metal (it will be trapped at a energy state higher than the minimum state). Annealing is the process of slowly reducing the temperature of the metal to achieve the minimum energy state. In terms of a global optimization algorithm the local minima (high energy states) are avoided by slow cooling to reach the global minimum (the minimum energy state).

SAs differ from GAs and ESs in how search points are generated and in the mechanism of deciding whether to move to a new point or not. The heart of the simulated annealing algorithm lies in the decision of whether to move from one point to another in the parameter space. Suppose the search has lead to a point $\mathbf{x}$ and the decision is whether to stay at that point or to move to $\mathbf{x}'$. The decision will be to move to $\mathbf{x}'$ with probability

$$\begin{cases} 1 & \text{if } f(\mathbf{x}') < f(\mathbf{x}) \\ \frac{1}{C} \exp\left(-\frac{f(\mathbf{x}') - f(\mathbf{x})}{T}\right) & \text{otherwise} \end{cases}$$

where C is a normalizing constant and the parameter T represents temperature. In the SA terminology these are called uphill and downhill moves. Downhill moves are always allowed while uphill moves are occasionally permitted. The latter feature will hopefully allow the algorithm to avoid becoming trapped at a local minumum. Note that when the temperature is high, as it is at the beginning of the algorithm, the probability of an uphill move will be large. So initially the algorithm will search widely throughout the parameter space. The temperature is decreased periodically so that when the temperature is very low the probability of an uphill move will be very small. This should occur at the end of the algorithm when it is hoped the search will be concentrated at the global minimum.

New points in the search space are generated by $\mathbf{x}' = \mathbf{x} + d \cdot \mathbf{z}$ where $\mathbf{z}$ is a vector symmetric random variables with mean zero. The factor d is used to control the step sizes of movements in the parameter space. This factor will be reduced during the course of the algorithm, so that the step sizes will tend to be very small at the end where the search will be concentrated at what is hoped to be the global minimum.

Schwefel (1995) gives a nice summary of a basic SA. The algorithm consists of an inner and an outer loop. The system is allowed to achieve thermal equilibrium in the inner loop. After equilibrium has been obtained in the inner loop, the temperature and stepsize parameters are reduced in the outer loop.

Step 0: (Initialization)

Choose a starting position $\mathbf{x}^{(\mathbf{0},\mathbf{0})}$,
an initial temperature T^0, and
an initial stepsize parameter d^0.
Set $\mathbf{x}^* = \hat{\mathbf{x}} = \mathbf{x}^{(\mathbf{0},\mathbf{0})}$, $k = 0$, and $l = 0$.

Step 1: (inner loop)

$\mathbf{x}^{(\mathbf{k},\mathbf{l})} = \hat{\mathbf{x}} + d^{(k)}\mathbf{z}$
If $f(\mathbf{x}^{(\mathbf{k},\mathbf{l})}) < f(\mathbf{x}^*)$ set $\mathbf{x}^* = \mathbf{x}^{(\mathbf{k},\mathbf{l})}$
If $f(\mathbf{x}^{(\mathbf{k},\mathbf{l})}) < f(\hat{\mathbf{x}})$, go to Step 3, else draw a uniform random number, χ from the interval [0, 1],
If $\chi < \frac{1}{c}\, exp\left(\frac{f(\mathbf{x}^{(\mathbf{k},\mathbf{l})}) - f(\hat{\mathbf{x}})}{T^{(k)}}\right)$, go to Step 3.

Step 2: (check for equilibrium)

If $f(\mathbf{x}^*)$ has not been improved within the last N trials, go to step 4.

Step 3: (end inner loop)

Set $\hat{\mathbf{x}} = \mathbf{x}(\mathbf{k}, \mathbf{l})$, set $l = l + 1$, go to Step 1.

Step 4: (Termination check)

If $(T^{(k)} \leqslant \epsilon)$, end the search with result $\mathbf{x}^*$

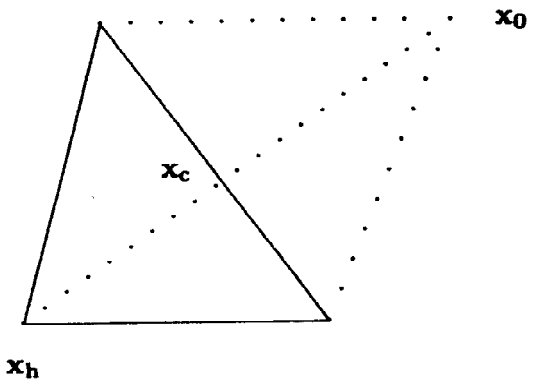

Figure 1. The Nelder–Mead simplex in two dimensions

Step 5: (Cooling, outer loop)
Set $\mathbf{x}^{(\mathbf{k},\mathbf{l})} = \mathbf{x}^*$, $\hat{\mathbf{x}} = \mathbf{x}^*$.
Set $T^{(k+1)} = \alpha T^{(k)}$, $0 < \alpha < 1$.
Set $d^{(k+1)} = \beta d^{(k)}$, $0 < \beta < 1$.
Set $l = 1$, $k = k + 1$, go to Step 1.

Goffe et al. (1994) have coded a simulated annealing algorithm and applied it to a simple econometric model. This version of the algorithm has a very nice output interface that allows a user to view the progress of the algorithm. If the progress does not seem satisfactory the user can stop the program and change the parameters of the algorithm to see if the change improves performance. This feature is most useful. The code for this implementation can be found at
`http://netlib2.cs.utk.edu/opt/simann.f`

3.3.1. *The Numerical Recipies SA*

Press et al. (1994) have combined a SA algorithm with the the Nelder–Mead simplex (1965) algorithm. A simplex in an n dimensional parameter consists of $(n+1)$ points that do not lie on a hyperplane, together with every possible convex combination of these points. An example of a simplex for a two dimensional parameter space is shown in Fig. 1. We are only interested in the vertices of the simplex. The Nelder–Mead algorithm is described next and will be followed by the a description of the modifications of Press et al.

Let

- $\mathbf{x_h}$ be the vertex with the highest function value,
- $\mathbf{x_s}$ be the vertex with the second highest function value,
- $\mathbf{x_l}$ be the vertex with the lowest function value function value,
- $\mathbf{x_c}$ be the centroid of all vertices except $\mathbf{x_h}$,

The location of a new vertex is determined as follows:

1. Reflect $\mathbf{x_h}$ through the centroid using some reflection factor $\alpha > 0$, that is compute
$$\mathbf{x_0} = (1 + \alpha)\mathbf{x_c} - \alpha\mathbf{x_h}.$$
2. If $f(\mathbf{x_l}) \leqslant f(\mathbf{x_0}) \leqslant f(\mathbf{x_h})$ then replace $\mathbf{x_h}$ with $\mathbf{x_0}$ and return to step 1.
3. If $f(\mathbf{x_0}) < f(\mathbf{x_l})$ expand the simplex using an expansion factor $\lambda > 1$ and compute
$$\mathbf{x_{00}} = \lambda\mathbf{x_0} + (1 - \lambda)\mathbf{x_h}.$$
    1. If $f(\mathbf{x_{00}}) < f(\mathbf{x_l})$, replace $\mathbf{x_h}$ with $\mathbf{x_{00}}$ and return to step 1.
    2. If $f(\mathbf{x_{00}}) > f(\mathbf{x_l})$, replace $\mathbf{x_h}$ with $\mathbf{x_0}$ and return to step 1.
3. If $f(\mathbf{x_0}) > f(\mathbf{x_s})$ contract the simplex. Let β have some value between 0 and 1.
    1. If $f(\mathbf{x_0}) < f(\mathbf{x_h})$, compute
    $$\mathbf{x_{00}} = \beta\mathbf{x_0} + (1 - \beta)\mathbf{x_c}.$$
    2. If $f(\mathbf{x_0}) > f(\mathbf{x_h})$, compute
    $$\mathbf{x_{00}} = \lambda\mathbf{x_h} + (1 - \lambda)\mathbf{x_h}.$$
    3. If $f(\mathbf{x_{00}}) < f(\mathbf{x_h})$ and $f(\mathbf{x_{00}}) < f(\mathbf{x_0})$, replace $\mathbf{x_h}$ with $\mathbf{x_{00}}$ and return to step 1.
    4. If $f(\mathbf{x_{00}}) > f(\mathbf{x_h})$ or $f(\mathbf{x_{00}}) < f(\mathbf{x_0})$, reduce the size of the simplex and return to step 1.

Press et al. consider the standard SA algorithm to be inefficient in that it may allow uphill moves when perfectly good downhill moves are available. This situation can arise when the algorithm is trying to maneuver through a narrow valley or when it is near a minimum. Consider any step in the Nelder–Mead algorithm where a comparison is made. At such a point, a positive random perturbation proportional to the temperature is added to the function value of any previously computed vertex and a positive random perturbation proportional to the temperature is subtracted from the function value computed at the proposed new vertex. Note that when the temperature is small, as it would be near the end of the algorithm, that the perturbations will have almost no effect and the Nelder–Mead algorithm will predominate. Adding and subtracting positive quantities from the function values computed at the vertices assures that downhill moves are always allowed and that uphill moves are occasionally allowed hence the similarity to SA.

3.4. A HYBRID METHOD (AS)

Stochastic optimization methods can serve as front ends for traditional deterministic optimization methods. In this case the stochastic method serves to provide the deterministic methods with, hopefully, good starting points.

The Royal Statistical Society at one time published a series of statistical algorithms in its journal Applied Statistics. Algorithm AS 298 (Brooks, 1995) is a hybrid minimization technique using simulated annealing and any minimization routine supplied by the user. AS 298 uses all of the points visited at the final temperature plus the best point found as starting points for the traditional minimization routine. Algorithm AS 319 (Koval, 1997) serves as a user supplied minimization routine in this research. Algorithm AS 319 is a variable metric quasi-Newton method using approximate gradients computed using forward differences. The author claims that the algorithm is designed to permit maximum likelihood estimation of functions of complex form and has a reasonable chance of obtaining the global optimum in every trial of a Monte Carlo simulation (Koval, 1997). The AS algorithms can be located in the Applied Statistics folder at the Statlib web site `http://lib.stat.cmu.edu/`.

4. A disequilibrium model

Economists usually assume that markets are in equilibrium, that is, that supply and demand are equal. This is probably a reasonable assumption for many goods such as commodities. However, given the time period over which much economic data is collected (often a quarter of a year or longer), it is not possible to determine whether the markets are in equilibrium. In some cases, such as the purchase of housing, the market can be in disequilibrium for long periods of time. Disequilibrium models present interesting and difficult estimation problems. The econometrics of such functions was first studied by Fair and Jaffe (1972) and refined by Fair and Kelejian (1974), Hartley and Mallela (1977), and Mayer (1989). Here we concentrate on the formulation by Maddala and Nelson (1974) because it was selected by Dorsey and Mayer (1995) as an item in a test suite of particularly difficult optimization problems faced by econometricians.

Suppose that the demand equation is written as

$$D_t = \mathbf{X}'_{\mathbf{1t}}\boldsymbol{\beta}_{\mathbf{1}} + u_t,$$

while the supply equation is

$$S_t = \mathbf{X}'_{\mathbf{2t}}\boldsymbol{\beta}_{\mathbf{2}} + v_t,$$

where D_t is the quantity demanded, S_t is the quantity supplied, and u_t and v_t are error terms. $\mathbf{X}_{1t}$ and $\mathbf{X}_{2t}$ are vectors of explanatory variables and $\boldsymbol{\beta}_{\mathbf{1}}$and $\boldsymbol{\beta}_{\mathbf{2}}$are parameters to be estimated. $\mathbf{X}_{1t}$ and $\boldsymbol{\beta}_{\mathbf{1}}$both have the same length as do $\mathbf{X}_{2t}$ and $\boldsymbol{\beta}_{\mathbf{2}}$. What makes the disequilibrium model interesting is that consumers cannot be forced to buy more than they want nor can they buy more than is offered. So the quantity actually purchased (Q_t) will be the lesser of supply or demand:

$$Q_t = \min(D_t, S_t).$$

What makes the problem even more interesting is that one may not be able to determine whether the quantity purchased is from the supply equation or from the demand equation. How, then, can the parameters $\boldsymbol{\beta_1}$and $\boldsymbol{\beta_2}$be estimated? One of the models considered by Maddala and Nelson (1974) and Maddala (1983) and used by Dorsey and Mayer (1995) is outlined in this section.

While the underlying model consists of two linear equations, the estimation problem is a difficult nonlinear one. The nonlinearity arises because we cannot determine whether a given value Q_t applies to the supply or to the demand equation. Instead we can determine only the probability that it came from one or the other of the equations. The probability that the observed value is a point on the demand equation is

$$\begin{aligned}\pi_t &= P(D_t < S_t)\\ &= P(\mathbf{X}'_{\mathbf{1t}}\boldsymbol{\beta_1} + u_t < \mathbf{X}'_{\mathbf{2t}}\boldsymbol{\beta_2} + v_t)\\ &= P(u_t - v_t < \mathbf{X}'_{\mathbf{2t}}\boldsymbol{\beta_2} - \mathbf{X}'_{\mathbf{1t}}\boldsymbol{\beta_1}).\end{aligned}$$

If the error terms are assumed to be independent and normally distributed, for example, then

$$\pi_t = \int_{-\infty}^{(\mathbf{X}'_{\mathbf{2t}}\boldsymbol{\beta_2} - \mathbf{X}'_{\mathbf{1t}}\boldsymbol{\beta_1})/\sigma} \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)\mathrm{d}u, \tag{1}$$

where $\sigma^2 = \sigma_1^2 + \sigma_2^2$. A similar relation holds for $P(S_t < D_t)$. Define

$$\begin{aligned}f_1(Q_t) &= \frac{1}{\sqrt{2\pi}\sigma_1}\exp\Big[-\frac{1}{2\sigma_1^2}(Q_t - \mathbf{X}'_{\mathbf{1t}}\boldsymbol{\beta_1})^2\Big]\\ f_2(Q_t) &= \frac{1}{\sqrt{2\pi}\sigma_2}\exp\Big[-\frac{1}{2\sigma_2^2}(Q_t - \mathbf{X}'_{\mathbf{2t}}\boldsymbol{\beta_2})^2\Big]\\ F_1(Q_t) &= \frac{1}{\sqrt{2\pi}\sigma_1}\int_{Q_t}^{\infty}\exp\Big[-\frac{1}{2\sigma_1^2}(D_t - \mathbf{X}'_{\mathbf{1t}}\boldsymbol{\beta_1})^2\Big]\mathrm{d}D_t\\ F_2(Q_t) &= \frac{1}{\sqrt{2\pi}\sigma_2}\int_{Q_t}^{\infty}\exp\Big[-\frac{1}{2\sigma_2^2}(S_t - \mathbf{X}'_{\mathbf{2t}}\boldsymbol{\beta_2})^2\Big]\mathrm{d}S_t\end{aligned}$$

Let $g(D_t, S_t)$ be the joint density of D and S. If an observation is a point on the demand equation, the conditional density of Q_t is (??)

$$\begin{aligned}g(Q_t \mid Q_t = D_t) &= \frac{f_1(Q_t)F_2(Q_t)}{P(Q_t = D_t)}\\ &= \frac{f_1(Q_t)F_2(Q_t)}{\pi_t}.\end{aligned}$$

and if the observation is a point on the supply equation,

$$g(Q_t \mid Q_t = S_t) = \frac{f_2(Q_t)F_1(Q_t)}{P(Q_t = S_t)} = \frac{f_2(Q_t)F_1(Q_t)}{1 - \pi_t}.$$

The unconditional density of Q_t is

$$\begin{aligned} f(Q_t) &= g(Q_t \mid Q_t = D_t)P(Q_t = D_t) + g(Q_t \mid Q_t)P(Q_t = S_t) \\ &= \pi_t \frac{f_1(Q_t)F_2(Q_t)}{\pi_t} + (1 - \pi_t)\frac{f_2(Q_t)F_1(Q_t)}{(1 - \pi_t)} \\ &= f_1(Q_t)F_2(Q_t) + f_2(Q_t)F_1(Q_t) \end{aligned}$$

The log-likelihood function then is

$$L = \sum_{t=1}^{T} \log[f_1(Q_t)F_2(Q_t) + f_2(Q_t)F_1(Q_t)],$$

where T is the sample size. If it is assumed that the error terms are normally distributed, then

$$\begin{aligned} h_{1t} &= \frac{Q_t - \mathbf{X}'_{1t}\boldsymbol{\beta}_1}{\sigma_1}, \\ h_{2t} &= \frac{Q_t - \mathbf{X}'_{2t}\boldsymbol{\beta}_2}{\sigma_2}, \\ f_{1t} &= \frac{1}{\sqrt{2\pi}\sigma_1}\exp(-h_{1t}^2/2), \\ f_{2t} &= \frac{1}{\sqrt{2\pi}\sigma_2}\exp(-h_{2t}^2/2), \\ F_{1t} &= \int_{h_{1t}}^{\infty} \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)\mathrm{d}u, \\ F_{2t} &= \int_{h_{2t}}^{\infty} \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)\mathrm{d}u, \\ L &= \sum_{t=1}^{T} \log(f_{1t}F_{2t} + f_{2t}F_{1t}), \end{aligned}$$

Note that while we desire to maximize the likelihood function, in practice we will minimize the negative likelihood function. This approach allows us to speak in terms of minimization, which is more common practice. The values that are reported in the subsequent tables, however, are the values of the likelihood function — in the tables larger values are preferred to smaller values. Also note that the variance

Table 1. Results for the disequilibrium model of housing starts reported by Dorsey and Mayer. The original solution is that of Maddala and Nelson. GA solution 1 and GA solution 2 are solutions reported by Dorsey and Mayer

Variable	Original solution	GA solution 1	GA solution 2
Demand constant	223.740	436.333	429.464
TT	2.520	0.457	10.615
SH	−0.002	−0.035	−0.130
MR(−2)	−0.90	0.178	0.328
Supply constant	15.550	5.058	7.788
TT	−0.153	−0.164	−0.161
PDF(-1)	0.053	0.055	0.054
BG(-2)	0.053	0.056	0.055
MR(-1)	0.093	0.108	0.014
σ_1^2	350.000	2.090	0.321
σ_2^2	80.200	88.922	88.641
Log likelihood	−459.618	−454.476	−452.449

terms are parameters that must be estimated. The estimates were constrained to lie in an interval $[-500, 500]$ for the β coefficients while the standard deviations were constrained to lie in an interval $[10^{-6}, 500]$.

Table 1 gives parameter estimates and the value of the log likelihood function for the disequilibrium model of housing starts considered by (??). The estimates for the parameters of the model that are shown in this table are TT = time trend, SH = stock of houses, MR(-2) = the mortgage rate lagged two periods, PDF(-1)= moving average of private deposit flows lagged one period, BG(-2) = moving average of borrowing by savings and loan associations from the Federal Home Loan Bank lagged two periods, and MR(-1) = the mortgage rate lagged one period. The mortgage rate serves as the price variable in this model and this choice seemingly has consequences that will be discussed later. The dependent variable Q is the observed number of housing starts. The solution labeled *Original solution* contains the best parameter estimates found by Maddala and Nelson. Maddala and Nelson do not conclude that they have located the global maximum, and only report the best solution. *GA solution 1* is the best solution found by Dorsey and Mayer using a genetic algorithm alone. *GA solution 2* uses the previous estimates as a starting point for a deterministic hill climbing routine. The hill climbing method gave a modest improvement in the value of the likelihood function.

Dorsey and Mayer report that neither GA or the SA of Goffe et al. were successful in locating the global optimum but that the GA tended to produce better likelihood values. They also note that the GA required substantially more execution time per run than did the SA. They ran 1000 SA trials in approximately the same

time that it took to run 10 GA trials. We discuss our timing findings in Section 5. Dorsey and Mayer also note the change in sign for the price variable in the demand equation, MR(-2), between their solutions and that of Maddala and Nelson.

We should mention that our computation of the likelihood function using these parameters differ from those reported in Table 1. We agree with the value of -459.618 for the original solution of Maddala and Nelson. However we find a value of -473.507 for the log likelihood function for GA solution 1 and a value of -2695.775 for GA solution 2. The results for GA solutions 1 and 2 were computed using our own code and the commercially vended statistical software package RATS (Regression Analysis for Time Series). The standard errors of the coefficients produced by the RATS program are very large at this point (sufficiently large that only one standard error led to a P value less than 0.5 and all but two P values had values greater than 0.999, using a t distribution). We found *GA solution 1* to have a poorly conditioned Hessian matrix at the point given. This makes it difficult to interpret the sign change on the price variable mentioned above. Interestingly a good result was obtained using *GA solution 2* as a starting point for the RATS maximization process. This result is shown in Table 2 in the column labeled *GA solution 2b* with the standard errors in parenthesis. The results obtained by Maddala and Nelson along with the standard errors of the parameter estimates are shown in the column labeled Original solution.

We did not find that any of the random search methods terminated at the same point more than once. This is consistent with the findings of Dorsey and Mayer. Most of the solutions found by the random search methods produced likelihood values worse than those shown in Table 1. We did find that the best solutions for ES and SA produced better likelihood values than the original solution of Maddala and Nelson. The solution produced by SA, however, also had a poorly conditioned Hessian matrix. The same was true for the best solutions using the other random search techniques except for the ES method. The best result using ES is given in Table 3. It gave better results than those found in Table 1 except for *GA solution 2* about which we have noted our suspicions. The *Best solution* found for the disequilibrium model is also given in Table 3. This solution was found by lucky accident rather than using any of the methods considered in this research. We used Maddala and Nelson's original solution as a starting point while using the RATS maximum likelihood program. No other starting point when using RATS produced such a result.

5. A test problem for the disequilibrium model using generated data

Maddala and Nelson (1974) also report difficulties in estimating this model. They computed the probability that each observation would belong to the demand side of the model using Eq. (1). They report that the "probabilities were uniformly low (less than .1) except for a few observations in 1969 which showed probabilities in the range .1 to .3. This shows that the entire period is possibly characterized by

Table 2. Results for the disequlibrium model. The original solution is that of Maddala and Nelson. GA solution 2b was found by using GA solution 2 as a starting point in the RATS maximization softwere. Standard errors are in parentheses.

Variable	Original solution		GA solution 2b	
Demand constant	223.740		458.899	
TT	2.520	(1.45)	0.618	(0.349)
SH	−0.022	(0.013)	−0.039	(0.009)
MR(-2)	−0.90	(0.062)	0.202	(0.052)
Supply constant	15.550		5.415	
TT	−0.153	(0.027)	−0.164	(0.036)
PDF(-1)	0.053	(0.006)	0.056	(0.006)
BG(-2)	0.053	(0.003)	0.055	(0.004)
MR(-1)	0.093	(0.006)	0.108	(0.023)
σ_1^2	350.000		2.002	
σ_2^2	80.200		88.906	
Log likelihood	−459.618		−454.411	

Table 3. Results for the disequlibrium model. The best result from a stochastic search method was from the ES shown in the column labeled Best ES solution. The best overall solution found is shown in column Best solution. Standard errors are in parentheses.

Variable	Best ES solution		Best solution	
Demand constant	441.658		273.597	
TT	11.834	(2.508)	2.062	(1.287)
SH	−0.145	(0.003)	−0.016	(0.012)
MR(-2)	0.395	(0.099)	−0.232	(0.022)
Supply constant	11.823		12.613	
TT	−0.156	(0.004)	−0.156	(0.036)
PDF(-1)	0.055	(0.006)	0.053	(0.006)
BG(-2)	0.054	(0.004)	0.053	(0.004)
MR(-1)	0.098	(0.004)	0.097	(0.023)
σ_1^2	0.468		0.013	
σ_2^2	89.246		90.935	
Log likelihood	−453.078		−451.741	

excess demand or that the model is poorly specified." This is consistent with our findings, also that the the probabilities show that almost all observations lie on the demand curve (124 out of 126 observations using the best set of parameters that we found).

Maddala and Nelson show particular concern with the adequacy of the price variable used. They note that it is "doubtful that mortgage rates really measure prices on the demand side, much less the supply side." The lack of a meaningful price variable also distinguishes the housing market from more conventional markets such as commodity markets. Bushels of corn are fairly homogeneous; houses are quite different. It is meaningful to talk about a price for a bushel of corn and at a particular time most bushels of corn will sell at the same price. Such cannot be said for houses. The behavior of the model at the points where we obtained good function values are typical of a mis-specified model; a poorly conditioned Hessian matrix and suspiciously good values for the standard errors of a model.

Maddala and Nelson propose the following test problem using artificial data to determine whether it is the data or the model that is causing the problem:

$$
\begin{aligned}
&D_t = \alpha_0 + \alpha_1 X_{1t} + \alpha_2 P_{t-1} + u_t,\\
&S_t = \beta_0 + \beta_1 X_{2t} + \beta_2 P_{t-1} + v_t,\\
&\frac{P_t - P_{t-1}}{P_{t-1}} = 0.5\frac{D_t - S_t}{D_t},\\
&Q_t = \min(D_t, S_t),
\end{aligned}
$$

$$
X_1 \sim U(0,1) \quad X_2 \sim U(0,1) \quad u_t \sim N(0, \sigma_u^2) \quad v_t \sim N(0, \sigma_v^2)
$$

and

$$
\begin{array}{lllllll}
P_0 &=& 1.0 & \sigma_u^2 &=& 0.0625 & \sigma_v^2 = 0.01,\\
\alpha_0 &=& 2.0 & \alpha_1 &=& 1.0 & \alpha_2 = -0.5,\\
\beta_0 &=& 1.4 & \beta_1 &=& 1.0 & \beta_2 = 0.1,
\end{array}
$$

where $U()$ indicates the uniform distribution and $N()$ the normal distribution. The created sample size is $n = 100$.

The search space was constrained as follows: $-50 < \alpha_1 < 50$, $-50 < \beta_i < 50$, $10^{-6} < \sigma_i < 50$. All of the random search methods terminated at the same point most of the time. ES and SA terminated at the same point all of the time. We strongly suspect that this is the global optimum for the disequilibrium model using generated data. The parameter estimates and standard errors are shown in Table 4. Timing information and the percentage of successful runs are shown in Table 5. We also computed the probability that an observed value would be on the demand curve using Eq. (1). The estimated model predicts that 38 out of 100 observations should be on the demand equation. The generated data actually had 45 out of 100 observation on the demand equation. These results suggest that Maddala

Table 4. Results for the disequlibrium model using generated data. Standard errors are in parentheses.

Variable	True value	Estimates	
α_0	2.0	1.858	
α_1	1.0	1.198	(0.171)
α_2	−0.5	−0.488	(0.123)
β_0	1.4	1.342	
β_1	1.0	1.098	(0.060)
β_2	0.1	0.138	(0.083)
σ_1^2	0.0625	0.0548	
σ_2^2	0.01	0.01	
Log-likelihood		50.612	

Table 5. The number of successful trials and total time for the trials for each method in estimating the disequlibrium model using generated data.

Method	Successful trials	Total time (s)
ES	100 out of 100	2802
GA	18 out of 20	262559
SA	100 out of 100	25159
NR	61 out of 100	16151
AS	70 out of 100	2504

and Nelson's procedure is useful in estimating a disequilibrium model and that mis-specification is the chief source of difficulty with the estimation process.

The results in Table 5 suggest that ES, GA, and SA all were very successful in finding the global optimum. The GA method required substantially more execution time than did any of the other methods and did not seem to be more reliable than ES or SA. The large execution time required by the GA may be the result of the choice of parameters used, but as noted previously, Dorsey and Mayer also found that GA required far more time than SA. These estimates agree with those produced by the RATS program. RATS was sensitive to the choice of starting point and these results give some confidence that the global optimum has been found. It seems, then, that Maddala and Nelson's method successfully models disequilibrium phenomena and that the techniques used in this research have most likely found the global maximum.

6. Garch models

Lately much effort has been devoted to the estimation of nonlinear time series models. A class of models that has seen a great deal of use are those that model changes in the variance of a time series. These models are useful in studying the determinants of changes in the variance of such economic variables as the inflation rate, exchange rates, and equity price variability for example.

A simple autoregressive model can be written as

$$y_t = \sum_{i=1}^{k} \beta_i y_{t-i} + \epsilon_t \tag{2}$$

where y is the dependent variable, β a vector of parameters, and ϵ_t a normally distributed error term with zero mean and constant variance. We consider, however, the situation where the variance of the error term is not constant. Such may be the case for inflation rates, exchange rates, and stock market indexes. The GARCH model (Bollerslev, 1986) is a popular representation of such a situation, where the error process is modeled by

$$\epsilon_t = v_t \sqrt{h_t}$$

and v_t is a random process with zero mean and variance equal to one.

A general model of the variance process is

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i e_{t-i}^2 + \sum_{j=1}^{q} \gamma_j h_{t-i} + \delta z_t \tag{3}$$

where z_t is a vector of exogenous variables and

$$e_t = y_t - \sum_{i=1}^{k} \hat{\beta}_i y_{t-i}$$

and $\hat{\beta}_i$ are estimates of the parameters of the autoregressive model. The log likelihood function for such a model is (Bollerslev, 1986)

$$L = -\frac{T-1}{2} \log(2\pi) - \frac{1}{2} \sum_{t=k}^{T} \left(\log h_t + \frac{e_t^2}{h_t} \right)$$

where k is a value equal to the maximum lag in the system.

One of the interesting features of this optimization problem is that h_t must be positive definite. Negative values are not sensible mathematically or statistically. One approach is to constrain all the parameters of the variance equation to be positive (Enders, 1995). This is not desirable. This would require an increase in the

value of any of the variables in the variance equation to always cause an increase in the value of the variance. This places unreasonable restrictions on the behavior of economic and financial systems.

There are other restrictions that the estimates must have so that the model will be stable. The condition for this is that the roots of the characteristic equation

$$1 - \gamma_1 z - \gamma_2 z^2 - \cdots - \gamma_p z^p = 0$$

have modulus greater than one, where γ_i are the coefficients of the lagged variance terms in Equation (3).

For this research we investigated

$$\begin{aligned} y_t &= \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3} \\ h_t &= \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2 + \gamma_1 h_{t-1} + \gamma_2 h_{t-2} \\ &\quad + \delta_1 z_{1,t} + \delta_2 z_{1,t-1} + \delta_3 z_{2,t} + \delta_4 z_{2,t-1} + \delta_5 z_{3,t} + \delta_6 z_{3,t-1} \end{aligned}$$

where y_t is the monthly return for the S&P 500 stock market index, $z_{1,t}$ is the ratio of German to US short-term interest rates at time t, $z_{2,t}$ the ratio of UK to US short-term interest rates at time t, and $z_{3,t}$ the ratio of Japanese to US short-term interest rates. The short term interest rates are annualized nominal rates of interest on three-month government securities. This model was suggested by Gerety and Learchman (1996). They graciously supplied the data used. The data consisted of 251 observations.

We can apply some restriction on the estimates for this model using the stability criteria on the roots of the characteristic equation. The roots will have modulus greater than one if $| \gamma_2 | < 1$, $\gamma_1 + \gamma_2 < 1$, and $\gamma_1 - \gamma_2 > -1$.

The hybrid method did not produce an acceptable result in any of 200 runs and none of the results are reported here. The best results from RATS using random starting points is shown in Table 6. GA did not produce any results better than the RATS solution although we only ran 20 trials because of the amount of time required (257 700 s). Further GA did not terminate at the same solution for any of the 20 trials. Because GA did not produce better results than RATS GA output is not shown here. NR(73 491 s for 200 runs), SA (122 553 s), and ES (67 896 s) did not converge to the same point in any of 200 trials. The best result for NR was marginally better than the RATS solution but had a set of parameter estimates that produced a poorly conditioned Hessian matrix. It did prove useful in providing a starting point for SA and ES. The best SA and ES estimates are shown in Table 7. These values are the best overall and did not produce a poorly conditioned Hessian matrix.

A fair amount of conflict exists between the signs and statistical significance of the coefficients of SA and ES. The SA results indicate that the coefficient on h_{t-2} is not significant while the ES results suggest that it is. The sign differs for the coefficient on h_{t-1} between ES and SA while both are significant (although the the SA coefficient is only marginally so). Likewise both models disagree about the

Table 6. Results for the volatility model of United States equity returns by Gerety and Learchman. ES is the evolutionary strategy and SA the simulated annealing method. Both used a point found by NR as as starting point. Standard errors are in parentheses.

Variable	ES		SA	
Constant(mean equation)	0.007		0.005	
US returns(−1)	0.243	(0.066)	0.305	(0.069)
US returns(−2)	−0.112	(0.066)	−0.152	(0.038)
US returns(−3)	0.021	(0.044)	−0.050	(0.065)
Constant(variance equation)	0.001		0.001	
ϵ^2_{t-1}	−0.015	(0.053)	0.078	(0.134)
ϵ^2_{t-2}	0.011	(0.037)	−0.109	(0.023)
h_{t-1}	1.323	(0.087)	−0.048	(0.229)
h_{t-2}	−0.601	(0.044)	0.534	(0.303)
Ger/US int. rate	0.004	(0.000)	−0.001	(0.001)
Ger/US int.rate(−1)	−0.005	(0.000)	0.000	(0.001)
UK/US int. rate	−0.005	(0.001)	0.004	(0.000)
UK/US int. rate(−1)	0.005	(0.001)	−0.003	(0.001)
J/US int. rate	−0.001	(0.001)	−0.002	(0.001)
J/US int. rate(−1)	0.001	(0.001)	0.002	(0.001)
Log likelihood	722.077		722.488	

significance of the coefficient of the e_{t-2} term. Also there is no agreement about the significance of the coefficient of several of the interest rate terms.

7. Conclusions

We examined two difficult economic estimation problems. None of the results for either model using actual data give much confidence that the global optimum has been located for either problem. There is fairly strong evidence that the data used to estimate the disequilibrium model has created a specification error. While it is possible to find good function values in some cases it is not clear what this really means. If the Hessian matrix is poorly conditioned, then the standard errors can be quite unreliable and it may not be clear which coefficients are statistically significant. A disequilibrium model using generated data does suggest that the estimation problems might well vanish with a properly specified model.

The GARCH model also proved difficult to estimate in the sense that the random search methods did not terminate at the same point. We were able to establish some points that gave function values superior to the deterministic method used in RATS. Only two of these points were not associated with a poorly conditioned Hessian matrix, however. This research suggests that it can be quite difficult to find the

Table 7. Results for the volatility model of United States equity returns by Gerety and Learchman using the RATS package. Standard errors are in parentheses.

Variable	RATS	
Constant(mean equation)	0.006	
US returns(−1)	0.240	(0.054)
US returns(−2)	−0.112	(0.071)
US returns(−3)	0.025	(0.057)
Constant(variance equation)	0.000	
ϵ^2_{t-1}	−0.014	(0.022)
ϵ^2_{t-2}	0.010	(0.036)
h_{t-1}	1.326	(0.248)
h_{t-2}	−0.602	(0.233)
Ger/US int. rate	0.003	(0.001)
Ger/US int.rate(−1)	−0.004	(0.001)
UK/US int. rate	−0.003	(0.001)
UK/US int. rate(−1)	0.004	(0.001)
J/US int. rate	−0.001	(0.001)
J/US int. rate(−1)	0.001	(0.001)
Log likelihood	719.622	

global optimum of certain econometric functions and that such a task can provide a fruitful and useful research agenda. The two best estimation methods used in this research were the simulated annealing method and the evolutionary strategy. The evolutionary strategy used less processor time than the simulated annealing technique and seemed to provide results of equal quality.

References

Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroscedasticity. *Journal of Econometrics* **31**, 307–327.

Brooks, S. P. (1995), Algorithm AS 298: A Hybrid Optimization Algorithm. *Applied Statistics* **44**(4), 530–552.

Dorsey, R. E. and Mayer W. J. (1995), Genetic Algorithms for Estimation Problems with Multiple Optima, Nondifferentiability, and Other Irregular Features. *Journal of Business & Economic Statistics* **13**(1), 53–66.

Enders, W. (1995), *Applied Econometric Time Series*, Wiley Series in Probability and Mathematical Statistics. Wiley, New York.

Fair, R. C. and Jaffee, D. M. (1972), Methods of estimation for markets in disequilibrium. *Econometrica* **40**, 497–514.

Fair, R. C. and Kelejian, H. (1974), Methods of Estimation for Markets in Disequilibrium: a Further Study. *Econometrica* **42**, 177–190.

Gerety, M. and Learchman, L. (1996), Assessing Interest Rate and Exchange Rate Effects on Equity Market Volatility: The Case of the US, UK, and Germany. *Journal of Economics* **XXII**(2), 25–33.

Goffe, W. L., Ferrier, G. D. and Rogers, J. (1994), Global Optimization of Statistical Functions with Simulated Annealing. *Journal of Econometrics* **60**, 65–99.

Hartley, M. J. and Mallela, P. (1977), The Asymptotic Properties of a Maximum Likelihood Estimator for a Market in Disequilibrium. *Econometrica* **46**, 1251–1257.

Koval, J. J. (1997), Algorithm AS 319: Variable Metric Function Minimization. *Applied Statistics* **46**(4), 515–540.

Maddala, G. S. (1983), *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Maddala, G. S. and Nelson, F. D. (1974), Maximum Likelihood Methods for Models of Markets in Disequilibrium. *Econometrica* **42**(6), 1013–1030.

Mayer, W. J. (1989), Estimating Disequilibrium Models with Limited a Priori Price-Adjustment Information. *Journal of Econometrics* **41**, 303–317.

Michalewicz, Z. (1996), *Genetic Algorithms + Data Structures = Evolution Programs*. 3rd edition Springer, Berlin.

Nelder, J. A. and Mead, R. (1965), A simplex method for function minimization. *The Computer Journal* **7**, 308–313.

Press, W. H., Teukolsky, S. A., Vettering, W. T. and Flannery, B. P. (1994), *Numerical Recipes in C - The Art of Scientific Computing*. 2nd edition Cambridge University Press, Cambridge.

Schwefel, H.-P. (1995), *Evolution and Optimum Seeking*. Wiley, New York.