Applications of Public Domain Global Optimization Software to Difficult Econometric Functions

Max E. Jerrell
Northern Arizona University
max.jerrell@nau.edu

1 Introduction

Determining estimates using global optimum of a maximum likelihood function or a nonlinear least squares function is very useful. The global minimum of a least squares function gives a consistent estimate [13]. While a local maximum can serve as a consistent estimate for a maximum likelihood function instead of the global maximum, the global maximum should be the most likely candidate [7].

Finding the global optimum can be challenging. The value located by a deterministic algorithm, such as Newton's method, will depend on the starting point used. Some starting points will lead to the global optimum while others will lead to a local optima. If there are a large number of local optima, finding the global optimum, in the absence of any other information, will largely be a matter of luck. Random search strategies are discussed in Sec. 2. A particular type of efficient random search, simulated annealing, is discussed in Sec. 3. A second type of efficient random search, genetic algorithms, is discussed in Sec. 4. Sec 5 considers evolutionary strategies, yet another search method and Sec 5.1 discusses a random search method combined with a deterministic search method. A list of the sources of the software used in this paper is given in Sec. 6. These algorithms are applied to difficult problems in Secs 789. Conclusions are presented in Sec. 10.

2 Random Search Strategies

Classical deterministic algorithms find minima depending on the starting point of the search. They may fail to converge or they may converge to a local rather than global minimum. Random search strategies offer methods that can avoid local minima in favor of the global minimum. Consider a search space such as the one shown in Fig. 1. A simple random search algorithm might be

1. Use random number generator to create point x in the search space.
2. Compute $f(x)$.

3. Use random number generator to create a new point x' in the search space.
4. if $f(x') < f(x)$, $x = x'$.
5. if convergence criteria met stop, else go to 3.

The problem with this search algorithm is that it retains no information except for the location of the best point ever found. If we are near the global minimum, we would like to concentrate the search in that region as shown in Fig. 2. A possibility is to reduce the search space and search near the current best point. However, this strategy will tend to become trapped in a region near a local minima. What we need is an algorithm that concentrates the search at promising locations, yet avoids being trapped at local optima.

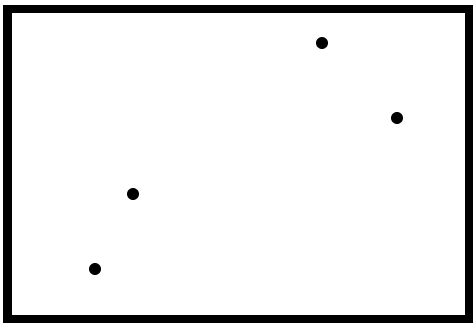

Figure 1: A Simple Search

Figure 2: A Concentrated Search

3 Simulated Annealing (SA)

Suppose we have somehow concentrated our search near the region of x_1 as shown in Fig. 3. If the search is totally restricted to be close to x_1 we will be trapped in a local minima, particularly if we adopt a rule that says only accept a movement to a point that gives a smaller function value than the current best value. Simulated Annealing (SA) avoids becoming trapped at a local minima by occasionally allowing uphill moves (to x_2 or x_3) but hopes to converge to the global optima by always allowing downhill moves. In the neighborhood of the global optimum, however, uphill moves should be discouraged. Suppose we have function values $f(x)$ and $f(x')$. Then we will move from x to x' with probability

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

where C is a normalizing constant and the parameter T is called temperature in the SA literature. SA was developed by Metropolis [19] by using an analogy to the thermodynamics of cooling and annealing metals. If heated metals are cooled slowly enough, nature finds the minimum energy state for the system. If cooled to rapidly, it will not reach this state but be trapped in a higher energy state (perhaps becoming brittle).

If T has a large value in Eq. 1 then there is a high probability of uphill moves thus escaping local minima. As T is reduced, uphill moves become less likely and the search is concentrated where it is hoped the global minimum is located. Also note that large uphill moves ($f(x') \gg f(x)$) are discouraged.

There is also the problem of how to determine new search points. A possibility is to use

$$x' = x + g(T)z \tag{2}$$

where $g(T)$ is an increasing function of T and z is some random number generator. In this case reductions in temperature also has the effect of concentrating the search near what is hoped to be the global minimum. This research uses the SA program developed by Goffe, Ferrier, and Rogers [9]. Note that a number of decisions must be made. What should the original temperature be? When should it be reduced? How much should it be reduced? What probability distribution should be used for z? How much and when do we reduce the step size control $g(T)$?

3.1 Numerical Recipes (NR)

Press et al. [22] have combined a SA algorithm with the the Nelder–Mead simplex [21] algorithm. A simplex in an n dimensional parameter consists of $(n+1)$ points which 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. 4. 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

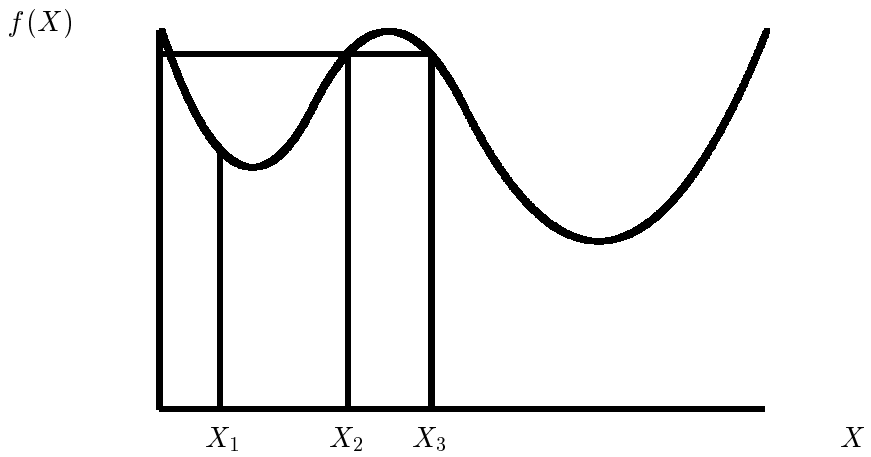

Figure 3: Uphill Moves

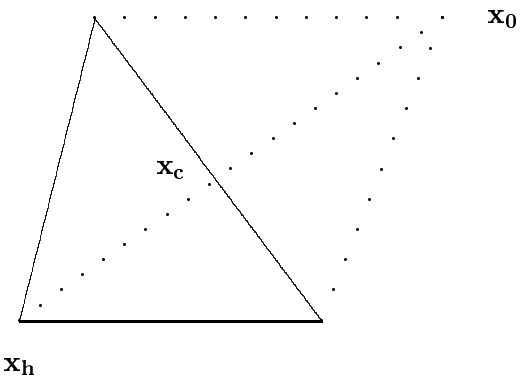

Figure 4: The Nelder–Mead simplex in two dimensions

$\mathbf{x_h}$ 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 $f(\mathbf{x_l}) \leq f(\mathbf{x_0}) \leq f(\mathbf{x_h})$ 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

$$\mathbf{x_{00}} = \lambda\mathbf{x_0} + (1 - \lambda)\mathbf{x_h}.$$

(a) If $f(\mathbf{x_{00}}) < f(\mathbf{x_l})$, replace $\mathbf{x_h}$ with $\mathbf{x_{00}}$ and return to step 1.

(b) If $f(\mathbf{x_{00}}) > f(\mathbf{x_l})$, replace $\mathbf{x_h}$ with $\mathbf{x_0}$ and return to step 1.

4. If If $f(\mathbf{x_0}) > f(\mathbf{x_s})$ contract the simplex. Let β have some value between 0 and 1.

(a) If $f(\mathbf{x_0}) < f(\mathbf{x_h})$, compute

$$\mathbf{x_{00}} = \beta\mathbf{x_0} + (1-\beta)\mathbf{x_c}.$$

(b) If $f(\mathbf{x_0}) > f(\mathbf{x_h})$, compute

$$\mathbf{x_{00}} = \lambda\mathbf{x_h} + (1-\lambda)\mathbf{x_h}.$$

(c) 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.

(d) 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 that 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.

4 Genetic Algorithms (GAs)

Genetic algorithms (GA's) mimic evolutionary processes. A point in the parameter space is called an *individual*. Each individual has a *fitness value* that is the value of the function at the point. Individuals produce offspring (new individuals at different points in the space). The most successful of these individuals, those with the best fitness values, survive and produce yet other offspring. It is hoped that offspring in a region near a local minima will eventually die out and that the surviving offspring will group near the global minimum.

4.1 Representation of points in GA's

It is customary to represent a point in the parameter space with binary vectors (called chromosomes). The value of x_1 could be represented by

$$v_1 = (b_{n-1}, b_{n-2}, \cdots, b_1, b_0)$$

where b_i (called a gene) can take on values of zero or one. If the search space for x_1 is restricted to the interval $[a, b]$ the conversion from the binary vector to a floating point value is given by

$$x_1 = a + \left(\sum_{i=0}^{n-1} b_i 2^i \right) \frac{b-a}{2^n - 1}$$

For example, suppose the search is to cover the interval $[1, 8]$ along x_1 using an 8 bit binary vector, then

$$v_1 = (00000000) \quad \text{gives} \quad x_1 = 1.000000$$

and

$$v_1 = (11111111) \quad \text{gives} \quad x_1 = 8.000000$$

while

$$v_1 = (01111110) \quad \text{gives} \quad x_1 = 4.486275$$

and

$$v_1 = (11111110) \quad \text{gives} \quad x_1 = 7.972549 \ .$$

4.2 GA's and GA Operators

Broadly speaking, GA's operate in three stages – selection, crossover, and mutation [20]. A GA begins with an initial population of individuals with different fitness values. Some of these individuals are selected for reproduction. Those individuals with good fitness values are given a large probability of reproduction, those with average fitness values an average probability of reproduction, and the rest a small probability of reproduction. Not all of the individuals with good fitness values are selected nor are all of those with poor fitness values ignored.

The individuals which are selected for reproduction produce offspring using the GA operators of crossover and mutation. The offspring then constitute the new population and the process is repeated. While mutation follows crossover in the process, it is easier to understand and will be discussed first. Each bit in the binary vector is given a probability changing value. Suppose we have (using the examples in Sec. 4)

$$v_1 = (11000000) \quad \text{gives} \quad x_1 = 6.270588 \ .$$

Now suppose b_6 is meets the criterion for change. The operation of mutation changes the value of b_6 producing

$$v_1' = (10000000) \quad \text{gives} \quad x_1 = 4.513725 \ .$$

Crossover combines the chromosomes of two individuals to produce two new individuals. Suppose the representation of the parents are

$$v_1 = (b_7, b_6, b_5, b_4, b_3, b_2, b_1, b_0)$$

and

$$v_2 = (c_7, c_6, c_5, c_4, c_3, c_2, c_1, c_0)$$

A position is selected at random in the binary vector, say $i = 3$, then the genes with indices $i <= 3$ are interchanged to produce

$$v_1' = (b_7, b_6, b_5, b_4, \mid c_3, c_2, c_1, c_0)$$

and

$$v_2' = (c_7, c_6, c_5, c_4, \mid b_3, b_2, b_1, b_0)$$

5 Evolutionary Strategies (ESs)

GAs originated in the United States while the development of another biological paradigm , Evolutionary Strategies, occurred in Europe. GAs were developed to be domain independent while ESs were developed with a continuous domain in mind [20]. GAs seem to have been most successful in discrete optimization problems. ESs use floating point vectors while classical GAs use binary vectors. Recent work with floating point GAs suggest that a floating point representation may be faster, provides higher precision, and greater consistency from run to run [20].

Classical GAs represent, say, a problems in a three dimensional space (x_a, x_b, x_c) with a binary representation

$$v = (a_{n-1}, \cdots, a_0, b_{n-1}, \cdots, b_0, c_{n-1}, \cdots, c_0).$$

We confess, as a matter of taste, that we feel somewhat uncomfortable with this representation and prefer, for continuous problems one that more nearly matches that domain. So, for this research, we have used an ES developed by Schwefel [23].

ESs use the same concepts of selection, cross-over, and mutation that are used in GAs. The ordering is different. The ordering in GAs is selection, cross-over, and mutation. In ESs it is mutation, cross-over, and then selection. In ESs, an initial population is created, which then generates offspring. In a (μ, λ) strategy μ parents produce λ children of which the best μ are selected to produce the next generation. In a $(\mu + \lambda)$ strategy μ of the $\mu + \lambda$ individuals are selected. The $(\mu + \lambda)$ strategy retains the current best values of the population. Unlike GAs, all individuals in the parent population, μ, are selected for reproduction.

ESs represent individuals as (x, σ) where $x = (x_1, x_2, \cdots, x_3)$ is a vector describing the location of the individual in the search space and $\sigma = \sigma_1, \sigma_2, \cdots, \sigma_n)$ is a vector used to control movement in the space. Mutations are given by

$$x' = x + N(0, \sigma)$$

where $N(0, \sigma)$ is a vector of independent, Gaussian distributed random numbers. Mutation is also applied to the standard deviation by, say,

$$\sigma' = \sigma \mathsf{LN}(0, \theta(\sigma))$$

where $\theta(\sigma)$ is a controlling parameter that becomes smaller in the vicinity of a minimum and LN() is the lognormal distribution.

Unlike GAs, every member of the population has an equal probability of being selected to produce offspring by crossover. Consider a population with μ parents

$$(x^1, \sigma^1), \cdots, (x^\mu, \sigma^\mu)$$

which produce an offspring

$$((x_1^{q_1}, \cdots, x_n^{q_n}), (\sigma_1^{q_1}, \cdots, \sigma_n^{q_n}))$$

with $q_j = 1, q_j = 2, \cdots, q_j = \mu$ with probability $1/\mu$. Another possibility is to assign the average values of the elements of pairs of parents to the offspring

$$((x_1^1 + x_1^2)/2, \cdots, (x_n^1 + x_n^2)/2), (\sigma_1^1 + \sigma_1^2)/2, \cdots, (x_n^1 + x_n^2)/2)).$$

Schwefel [23] devised a means of correlating the mutation parameters, σ_j, so that mutations need not be made along coordinate directions, but can occur in the most promising direction. This allows mutations to proceed along valleys which are not orthogonal with respect to the coordinates.

5.1 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 algorithm in its journal Applied Statistics. Algorithm AS 298 [2] 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 [14] served as the user supplied minimization routine in this research. Algorithm AS 319 is a variable metric quasi-Newton method with approximate gradients computed using forward differences. It is claimed 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 [14].

6 Software sources

Dorsey and Mayer examined the performance of several optimization methods on a number of functions, including econometric functions [3]. 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. [10] which is discussed below.

Dorsey and Mayer's code is based on an schema developed by Holland [12]. Dorsey and Mayer conclude that stochastic methods are more reliable that deterministic methods except for small problems that have no more than four dimensions. They also conclude that their version of the genetic algorithm is more reliable than the simulated annealing method of Goffe et al.

Goffe et al. [10] have coded a basic simulated annealing algorithm and applied it to a simple econometric model. This version of the algorithm is not elaborate, but does have a very nice output interface. This 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`

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

The AS algorithms can be located in the Applied Statistics folder at the Statlib web site `http://lib.stat.cmu.edu/`.

7 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 [5] and refined by Fair and Kelejian [6], Hartley and Mallela [11], and Mayer [18]. Here we concentrate on the formulation by Maddala and Nelson [17] because it was selected by Dorsey and Mayer [3] as an item in a test suite of particularly difficult optimization problems faced by econometricians.

Suppose that the demand equation is written as

$$D_t = \beta_1 X_{1t} + u_t,$$

while the supply equation is

$$S_t = \beta_2 X_{2t} + v_t,$$

where D_t is the quantity demanded, S_t is the quantity supplied, X_{1t} and X_{2t} are matrices of independent variables, β_1 and β_2 are vectors of parameters to be estimated, and u_t and v_t are error terms. 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 β_1 and β_2 be estimated? One of the models considered by Maddala and Nelson [17] and Maddala [16] and used by Dorsey and Mayer [3] is considered 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

$$\begin{aligned} \pi_t &= P(D_t < S_t) \\ &= P(\beta_1 X_{1t} + u_t < \beta_2 X_{2t} + v_t) \\ &= P(u_t - v_t < \beta_2 X_{2t} - \beta_1 X_{1t}). \end{aligned}$$

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

$$\pi_t = \int_{-\infty}^{(\beta_2 X_{2t} - \beta_1 X_{1t})/\sigma} \frac{1}{\sqrt{2\pi}} \exp(-u^2/2) du,$$

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

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 [16]

$$\begin{aligned} g(Q_t \mid Q_t = D_t) &= \frac{\int_{Q_t}^{\infty} g(Q_t, S_t) dS_t}{P(Q_t = D_t)} \\ &= \frac{1}{\pi_t} \int_{Q_t}^{\infty} g(Q_t, S_t) dS_t, \end{aligned}$$

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

$$\begin{aligned} g(Q_t \mid Q_t = S_t) &= \frac{\int_{Q_t}^{\infty} g(D_t, Q_t) dD_t}{P(Q_t = S_t)} \\ &= \frac{1}{1 - \pi_t} \int_{Q_t}^{\infty} g(D_t, Q_t) dD_t. \end{aligned}$$

The unconditional density of Q_t is

$$\begin{aligned} G(Q_t) &= \pi_t g(Q_t \mid Q_t = D_t) + (1 - \pi_t) g(Q_t \mid Q_t = S_t) \\ &= \int_{Q_t}^{\infty} g(Q_t, S_t) dS_t + \int_{Q_t}^{\infty} g(D_t, Q_t) dD_t. \end{aligned}$$

Suppose that the error terms u_t and v_t are independent so that $g(D_t, S_t) = g_1(D_t)g_2(S_t)$. In this case the unconditional density simplifies to

$$\begin{aligned} G(Q_t) &= g_1(Q_t)\int_{Q_t}^{\infty} g_2(S_t)dS_t + g_2(Q_t)\int_{Q_t}^{\infty} g_1(D_t)dD_t \\ &= g_1(Q_t)G_2(Q_t) + g_2(Q_t)G_1(Q_t), \end{aligned}$$

where g_1 and g_2 are density functions and G_1 and G_2 are distribution functions. The log-likelihood function is

$$L = \sum_{t=1}^{n} \log\ G(Q_t),$$

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

$$\begin{aligned} h_{1t} &= \frac{Q_t - \beta_1 X_{1t}}{\sigma_1}, \\ h_{2t} &= \frac{Q_t - \beta_2 X_{2t}}{\sigma_2}, \\ g_{1t} &= \frac{1}{\sqrt{2\pi}\sigma_1}\exp(-h_{1t}^2/2), \\ g_{2t} &= \frac{1}{\sqrt{2\pi}\sigma_2}\exp(-h_{2t}^2/2), \\ G_{1t} &= \int_{h_{1t}}^{\infty} \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)du, \\ G_{2t} &= \int_{h_{2t}}^{\infty} \frac{1}{\sqrt{2\pi}}\exp(-u^2/2)du, \\ G_t(Q_t) &= g_{1t}G_{2t} + g_{2t}G_{1t}. \end{aligned}$$

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. Also note that the variance terms are parameters that must be estimated.

Table 7 gives parameter estimates and the value of the log likelihood function for the disequilibrium model of housing starts considered by Maddala and Nelson [17]. The estimates for the parameters 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. The dependent variable Q is the observed number of housing starts. Of particular interest is the sign of the mortgage interest rate coefficient in the demand equation. The *original solution* in Table 7 is the best result obtained by Maddala and Nelson. *GA solution 1* is the best result obtained by

Dorsey and Mayer using the GA. *GA solution 2* is the result Dorsey and Mayer found using a quadratic hill climbing method with *GA solution 1* as a starting point. They note change in sign on the coefficient and that the sign should be negative, the mortgage rate being a price variable in the demand equation. Their conclusion is that the demand equation is mis–specified. The research presented here suggests that Dorsey and Mayer have located a local rather than the global maximum.

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

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 1 and GA 2 are solutions computed by two genetic algorithms reported by Dorsey and Mayer

Table 2 shows the results for the disequilibrium model using the evolutionary strategy with three different conditions. ES solution 1 is the solution given for an arbitrary starting point. ES solution 2 and ES solution 3 both use the the best value found using the SA routine of Goffe et al. (shown in Table 3). ES solution 3 employed a $\mu + \lambda$ strategy while ES solution 2 employed a (μ, λ) strategy. ES solution 3 did not improve on the results of Goffe et al., but ES solution 2 wandered from this point. This suggests that retaining the best individuals, letting them reproduce, and living until a better individual replaces them can be a successful strategy. Note that this depended on a very good starting point, however.

Table 3 shows the results for SA, GA, and NR for the disequilibrium model. As mentioned in the previous paragraph, ES did not find a particularly good optimum until it was provided with the best result from the SA routine of Goffe et al. The GA algorithm is that of Dorsey and Mayer which used a larger parameter search space that Dorsey and Mayer used. The expanded search space produced better results than they found using the GA in their research. The NR routine used the starting point suggested by SA. While this may suggest that the SA routine tended to produce better results than the other techniques, this is not necessarily so. SA did not give particularly good results until using a starting point suggested by NR. But then NR gave better results later using the new results from SA as a starting point. This, in turn, suggested the expanded search space for GA. Note that in all cases, the mortgage rate variable now has the expected sign. This suggests that Dorsey and Mayer located a local rather than global variable and the conclusion that the model is mis–specified is too hasty. Finally it must be mentioned that AS never

Variable	ES solution 1	ES solution 2	ES solution 3
Demand constant	374.680	445.874	1159.725
TT	0.755	0.240	7.737
SH	-0.007	-0.003	-0.066
MR(-2)	-0.435	-0.549	-1.792
Supply Constant	-38.341	0.013	-20.895
TT	-0.055	0.075	-0.168
PDF(-1)	0.062	0.055	0.043
BG(-2)	0.072	0.061	0.049
MR(-1)	0.127	0.061	0.152
σ_1^2	9.285	8.466	5.925
σ_2^2	9.925	10.117	9.789
Log-likelihood	-449.605	-446.718	-437.155

Table 2: Results for the disequilibrium model of housing starts using Evolutionary Strategies

really was successful in any attempts to estimate the parameters of the disequilibrium model.

8 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 in changes in the variance of such economics variables as the inflation rate, exchange rates, and equity price variability for example.

A simple autoregressive model can be written as

$$y_t = \sum_{i=1}^{n} \beta_i y_{t-i} + \epsilon_t \tag{3}$$

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 [1] is a popular representation of such a situation, where the error process is modeled by

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

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

Variable	SA	GA	NR
Demand constant	1159.725	1105.39	1160.638
TT	7.737	18.1	7.754
SH	-0.066	-1.509	-0.006
MR(-2)	-1.792	-0.219	-1.794
Supply Constant	-20.895	14.629	-20.764
TT	-0.168	-0.15	-0.169
PDF(-1)	0.043	0.052	0.043
BG(-2)	0.049	0.054	0.049
MR(-1)	0.152	0.093	0.151
σ_1^2	5.925	0.072	5.911
σ_2^2	9.789	9.192	9.806
Log-likelihood	-437.155	-452.955	-437.156

Table 3: SA, GA, and NR results for the disequilibrium model of housing starts. The GA solution uses the code of Dorsey and Mayer but a different parameter space.

A general model of the variance process is

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i e_{t-i}^2 + \sum_{j=1}^{q} \gamma_j h_{t-i} + \delta z_t \tag{4}$$

where z_t is a vector of exogenous variables and

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

and $\hat{\beta}_i$ are estimates of the parameters of the autoregressive model. The log likelihood function for such a model is [1]

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

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

One of the interesting features of the optimization problems 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 [4]. This is not desirable. This would require that 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.

For this research we investigated

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \beta_3 y_{t-3}$$

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

where y_t it 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.

The short term interest rates are annualized nominal rates of interest on three month government securities. This model was suggested by Gerety and Leachman [8]. They graciously supplied the data used.

The hybrid method did not produce an acceptable result in any of 200 runs and none of these results are reported here. Table 6 give the best results for a number of runs with different starting points using the Regression Analysis of Times Series (RATS) statistical package, a widely used econometrics software package. RATSs uses deterministic maximum likelihood algorithms. There is very little consistency between the signs of the coefficient estimates produced by the different runs suggesting that the likelihood function may be characterized by multiple local maxima. This is reinforced by examining the values of the likelihood function produced by RATS and the values produced by random methods.

The best solution for all methods is that given by the SA method shown in Table 4. This result could not be produced from a random starting point, however. This result was produced using the best NR result as a starting point for the SA method.

Table 5 gives results for the ES. ESa was the best solution found in 200 runs using arbitrary starting points. ESb is the best solution found in 200 runs using the best solutions from the NR method as a starting point and letting the best individual live forever. ESc also used the NR result as a starting point, but parents always die in this version. Note that in the ESc case that ES wanders away from the best point.

Table 7 gives the results for three runs of the GA. GA 1 and GA 2 give the largest value of the likelihood function, but neither one is as good as the best result from NR, ES, or SA. While the value of the likelihood function is approximately the same for GA 1 and GA 2, hardly any of the parameters have the same sign. Examination of the signs of the estimates for all of the methods does not show any consistency. This can be taken as evidence that the likelihood function has multiple optima.

9 Kowalik's equation

Kowalik [15] function serves as an example of a nonlinear least squares estimation problem. The objective function is

Variable	NR solution	ES solution	SA solution
Constant(mean equation)	0.006	0.005	0.006
US returns(-1)	0.198	0.293	0.338
US returns(-2)	0.036	-0.110	0.014
US returns(-3)	-0.043	0.053	0.029
Constant(variance equation)	0.006	0.004	0.005
ϵ_{t-1}^2	0.133	0.174	0.062
ϵ_{t-2}^2	0.007	0.018	0.032
h_{t-1}	-0.450	-0.654	-0.376
h_{t-2}	-0.428	-0.333	-0.822
Ger/US int. rate	0.001	-0.009	0.003
Ger/US int.rate(-1)	0.002	-0.005	-0.001
UK/US int. rate	0.002	0.009	-0.003
UK/US int. rate(-1)	-0.006	-0.007	0.000
J/US int. rate	0.000	-0.002	-0.0584
J/US int. rate(-1)	0.000	0.001	0.001
Log-likelihood	724.584	710.273	726.840

Table 4: Results for the volatility model of United States equity returns by Gerety and Leachman. NR is the Numerical Recipes solution, ES the evolutionary strategy solution, SA the simulated annealing solution

Variable	ESa	ESb	ESc
Constant(mean equation)	0.005	0.006	0.007
US returns(-1)	0.293	0.198	0.184
US returns(-2)	-0.110	0.036	0.244
US returns(-3)	0.053	-0.043	-0.353
Constant(variance equation)	0.004	0.006	0.006
ϵ_{t-1}^2	0.174	0.133	0.073
ϵ_{t-2}^2	0.018	0.007	0.062
h_{t-1}	-0.654	-0.450	-0.845
h_{t-2}	-0.333	-0.428	-0.566
Ger/US int. rate	-0.009	-0.001	-0.005
Ger/US int.rate(-1)	0.005	0.002	0.005
UK/US int. rate	0.009	0.002	0.005
UK/US int. rate(-1)	-0.007	-0.006	-0.007
J/US int. rate	-0.002	0.000	-0.001
J/US int. rate(-1)	0.001	0.000	0.001
Log-likelihood	710.273	721.945	682.712

Table 5: Results for the volatility model of United States equity returns by Gerety and Leachman. ESa is the evolutionary strategy solution found by starting with an arbitrary point retaining the best solution, ESb uses the NR solution as a starting point also retaining the best solution, and ESc uses the NR solution as a starting point but does not retain the best solution.

Variable	RATS1	RATS2	RATS3
Constant(mean equation)	0.004	0.000	0.006614
US returns(-1)	0.245	0.353	0.213129
US returns(-2)	-0.066	0.088	-0.13232
US returns(-3)	-0.009	0.056	0.036214
Constant(variance equation)	0.003	0.002	0.000563
ϵ^2_{t-1}	0.211	0.083	-0.04118
ϵ^2_{t-2}	0.024	0.034	-0.03066
h_{t-1}	-0.681	0.111	0.374663
h_{t-2}	0.019	0.269	-0.53744
Ger/US int. rate	0.000	0.005	0.000151
Ger/US int.rate(-1)	0.001	-0.006	-0.00148
UK/US int. rate	0.000	-0.007	0.001189
UK/US int. rate(-1)	-0.001	0.007	0.000687
J/US int. rate	-0.004	-0.003	-0.0006
J/US int. rate(-1)	0.003	0.003	0.000561
Log-likelihood	676.465	664.966	646.8787

Table 6: Results for the volatility model of United States equity returns by Gerety and Leachman. Results from the RATS package with different starting points

Variable	GA 1	GA 2	GA 3
Constant(mean equation)	0.002	0.006	0.002
US returns(-1)	0.275	0.263	0.303
US returns(-2)	-0.093	0.000	0.015
US returns(-3)	0.001	-0.001	0.076
Constant(variance equation)	0.002	0.006	0.001
ϵ^2_{t-1}	0.173	0.000	0.052
ϵ^2_{t-2}	-0.063	0.003	-0.154
h_{t-1}	0.000	-0.281	-0.082
h_{t-2}	0.045	-0.806	0.036
Ger/US int. rate	0.002	0.000	-0.018
Ger/US int.rate(-1)	-0.003	0.000	0.015
UK/US int. rate	0.000	0.000	0.026
UK/US int. rate(-1)	0.000	-0.003	-0.021
J/US int. rate	0.000	-0.001	-0.006
J/US int. rate(-1)	0.000	0.001	0.006
Log-likelihood	715.759	714.713	695.160

Table 7: Results for the volatility model of United States equity returns by Gerety and Leachman. Results from the genetic algorithm of Dorsey and Mayer

i	a_i	b_i^{-1}
1	0.1957	0.25
2	0.1947	0.5
3	0.1735	1
4	0.1600	2
5	0.0844	4
6	0.0627	6
7	0.0456	8
8	0.0342	10
9	0.0323	12
10	0.0235	14
11	0.0246	16

Table 8: Constants for Kowalik's function.

Estimate	RATS 1	RATS 2	RATS 3
x_1	0.1941	0.2069	0.2104
x_2	0.1601	-0.0707	-5.5319
x_3	0.1231	0.2509	-3.8819
x_4	0.1202	-0.0225	-2.3026

Table 9: Results from the RATS package for Kowalik's function

$$f(x) = \sum_{i=1}^{11} \left(a_i - \frac{x_1(b_i^2 + b_i x_2)}{b_i^2 + b_i x_3 + x_4} \right)^2$$

with values for a and b given in Table 8.

The global minimum for this function is $x = (0.1928, 0.1908, 0.1231, 0.1358)$. This function has some features which make it difficult for traditional minimization methods. It tends to suffer greatly from round-off error in the vicinity of the global minima [23].

This function was estimated using the nonlinear least squares program in RATS. If starting points were chosen close to the global minimum it consistently converged to an improper value. When given a starting point equal to the global minimum, Rats produced the the results shown in column RATS 1 in Table 9. At a starting point such as $x = (0, 0, 0, 0)$ RATS produced the results shown in column RATS 2. A starting point of $x = (1, 1, 1, 1)$ produced the results in column RATS 3.

The hybrid method did not converge in any of 200 runs. This is not too surprising given the round–off error near the minimum and is consistent with the results for RATS 1 in Table 9. NR did not converge

Estimate	GA	ES	SA
x_1	0.1928	0.1928	0.1928
x_2	0.1909	0.1924	0.1907
x_3	0.1232	0.1235	0.1231
x_4	0.1358	0.1365	0.1357
Number converged	200	194	200

Table 10: Results for Kowalik's function for GA, ES and SA. Number converged is the number of runs that converges out of 200 total runs. The values shown for the estimates are for the best function value for the 200 runs.

in any of 200 runs either. NR, unlike the other methods, is not confined to search in a particular region. This led to NR wandering very far from the global minimum. One such result was (approximately) $x = (.1, 1.0e34, -1.0e34, 1.0e20)$ NR never converged to the same location twice if there was any change in the starting position.

The other techniques performed quite well on Kowalik's function as shown in Table 10. GA and SA both converged to the same location in each of 200 test runs. ES converged in 194 out of 200 runs. The values shown for the parameter estimates are those that gave the minimum function value for the 200 runs for each technique. SA and GA gave values closer to the value given by Kowalik.

10 Conclusions

No single method used in this research proved clearly superior. One recommendation frequently made is to use the results of a random search method for a starting point in a deterministic method. This research suggests that this is not always sound. The hybrid method did not perform well in this research and failed completely for the Garch model and Kowalik's function.

This research also suggests that it is worthwhile to use more than one method. An experimental approach, where different combinations of techniques are tried, produced the best results.

References

[1] T. Bollerslev. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, 31:307–327, 1986.

[2] Stephen P. Brooks. Algorithm as 298: A hybrid optimization algorithm. *Applied Statistics*, 44(4):530–552, 1995.

[3] R. E. Dorsey and W. J. Mayer. Genetic algorithms for estimation problems with multiple optima, nondifferentiability, and other irregular features. *Journal of Business & Economic Statistics*, 13(1):53–66, Jan 1995.

[4] Walter Enders. *Applied Econometric Time Series*. Wiley Series in Probability and Mathematical Statistics. Wiley, New York, 1995.

[5] R. C. Fair and D. M. Jaffee. Methods of estimation for markets in disequilibrium. *Econometrica*, 40:497–514, 1972.

[6] R. C. Fair and H. Kelejian. Methods of estimation for markets in disequilibrium: a further study. *Econometrica*, 42:177–190, 1974.

[7] Li Gan and Jiming Jiang. A test for global maximum. *Journal of the American Statistical Association*, 94(447):847–854, 1999.

[8] Mason Gerety and Lori Learchman. 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, 1996.

[9] W. L. Goffe, G. D. Ferrier, and J. Rogers. Simulated annealing: An initial application in econometrics. *Computer Science in Economics and Management*, 5:133–146, 1992.

[10] W. L. Goffe, G. D. Ferrier, and J. Rogers. Global optimization of statistical functions with simulated annealing. *Journal of Econometrics*, 60:65–99, 1994.

[11] M. J. Hartley and P. Mallela. The asymptotic properties of a maximum likelihood estimator for a market in disequilibrium. *Econometrica*, 46:1251–1257, 1977.

[12] J. H. Holland. *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI, 1975.

[13] G. G. Judge, R. C. Hill, W. E. Griffiths, H. Lütkepohl, and T. Lee. *Introduction to the Theory and Practice of Econometrics*. John Wiley & Sons, New York, 1982.

[14] John J. Koval. Algorithm as 319: Variable metric function minimization. *Applied Statistics*, 46(4):515–540, 1997.

[15] J. Kowalik and J. F. Morrison. Analysis of kinetic data for allosteric enzyme reactions as a nonlinear regression problem. *Math. Biosci.*, 2:57–66, 1968.

[16] G. S. Maddala. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge University Press, Cambridge, 1983.

[17] G. S. Maddala and F. D. Nelson. Maximum likelihood methods for models of markets in disequilibrium. *Econometrica*, 42(6):1013–1030, Nov 1974.

[18] W. J. Mayer. Estimating disequilibrium models with limited a priori price-adjustment information. *Journal of Econometrics*, 41:303–317, 1989.

[19] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

[20] Z. Michalewicz. *Genetic Algorithms + Data Structures = Evolution Programs*. Springer, Berlin, 3rd edition, 1996.

[21] J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7:308–313, 1965.

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

[23] Hans-Paul Schwefel. *Evolution and Optimum Seeking*. John Wiley & Sons, Inc., New York, 1995.