13. REPETITIVE STOCHASTIC GUESSTIMATION FOR ESTIMATING PARAMETERS IN A GARCH(1,1) MODEL

Adriana AGAPIE[*]
Constantin BRATIANU[**]

Abstract

A behavioral algorithm for optimization - Repetitive Stochastic Guesstimation (RSG) - is adapted, with complete proofs for its global convergence, for estimating parameters in a GARCH(1,1) model, based on a very small number of observations. Estimators delivered by this algorithm for the example of a GARCH(1,1) model are dependent on some computational capabilities - namely number of iterations and replications performed. In this context, the Large Numbers Law might be applied in a completely different dimension. An alternative toward waiting until the historical data series are recorded (while the underling process may change several times) is to use computers for correctly extracting information from the most recent data. Given the existent computational support, it is also possible to determine estimates for the rates of convergence. As a result, potential benefits of this econometric technique can be gained in case of very young financial markets from Eastern European countries. Also, prediction and political decisions based on these estimations are properly grounded.

Keywords: RSG, GARCH Model, financial markets

JEL Classification: C22, G10

1. Theoretical Results regarding Global Convergence for Evolutionary Algorithms

Evolutionary Algorithms (EAs) - mentioned before in the context of the analysis of the emergent markets - are probabilistic search algorithms, which start with an initial population of likely problem solutions and then evolve towards better solutions. They are based on the mechanics of natural genetics and natural selection. A simple EA

[*] *UNESCO Chair for Business Administration, Academy of Economic Studies, Bucharest, Romania, agapie@clicknet.ro.*
[**] UNESCO Chair for Business Administration, Academy of Economic Studies, Bucharest, Romania, cbratianu@yahoo.com.

requires the definition of the following components:

- a solution representation;
- a function verifying the fitness of solutions - called *fitness function*;
- some operators to carry on the evolution of a population of potential solutions from a generation to another.

Some examples of distinctive EA's are: Genetic Algorithms (GA), Simulated Annealing (SA) and Repetitive Stochastic Guesstimation (RSG). Since GA and SA have been studied for quite long time and they are also present in the economic literature [see (Goffe, Ferrier and Rogers, 1994),(Przechlewski and Strzala,1996)] we will briefly review the literature regarding results for global convergence. The intention here is to explain why it is not satisfactory to rely on these stochastic algorithms for optimization with the purpose of estimating parameters in some econometric models, when only a small sample size is available. This will clear the way in presenting the advantages of some 'adapted to the problem in work' versions of the RSG algorithm for the task of optimization, under non-linearity and small sample size assumptions. Theoretical proofs for the global convergence, indications about a correct choice of the initial values and some measure for the convergence rate will be indicated.

Simulated annealing and the genetic algorithm are stochastic relaxation search techniques suitable for application to a wide variety of combinatorial complexity non-convex optimization problems [T.E. Davis, 1991]. Each produces a sequence of candidate solutions (or populations of candidate solutions) to the underlying optimization problem, and the purpose of both algorithms is to generate sequences biased toward solutions, which optimize the objective function.

The appeal of SA is that it provides asymptotic convergence to a globally optimal solution. A substantially body of knowledge exists concerning the algorithm convergence behavior, based upon a Markov chain model. The essence of SA theory is the demonstration of (1) existence of a unique asymptotic probability distribution (stationary distribution) for the stationary Markov chain corresponding to every strictly positive constant value of an algorithm control parameter (absolute temperature), (2) existence of a stationary distribution limit as the control parameter approaches zero, (3) the desired behavior of the stationary distribution limit, i.e., optimal solution with probability one, and finally (4) sufficient conditions on the algorithm control parameter to ensure that the non-stationary algorithm achieves the limiting distribution. In the most general form SA theory is presented in [Haario, Saksman 1991].

Attempting to copy this theory onto GA was only partially successful, resulting in complicated - and rather intractable - definitions for the genetic operators, and questionable convergence proofs. That occurs because GA does not rely on the Gibbs distribution, the magic key to global convergence engaged by SA. Instead, GA took advantage of the possibility of using more than one individual per iteration, achieving simple homogeneous Markov chain convergence for the elitist algorithm [Rudolph, 1996]. Summing up, most of the SA and GA convergence results are for general fitness function and infinite time convergence, which is of academic use only. Practical relevance could be achieved by analyzing particular convergence rates of particular problems and algorithms, but this has not been done yet, except a few (simple) cases.

Repetitive Stochastic Guesstimation (RSG) is a probabilistic algorithm introduced in [Charemza, 2002] which mimics the usual guessing of the parameters involved in a complex, generally large, empirically oriented macroeconomic model. The main intention of this paper is to show how an empirical algorithm like RSG can be adjusted for estimating parameters in a non-linear model based on a minimal set of observations. This adjustment is possible to be performed in such a way that proofs for global convergence should not rely on the data set, but on the number of iterations in the algorithm. Building on the idea in [Agapie, 20091] a formal representation of the RSG for estimating parameters in a GARCH (1,1) model will be presented. Sufficient conditions for determining the 'true parameters' values in the model will be indicated. Before all these, a very brief review of this algorithm is required.

First, three points should be mentioned where RSG differs from other evolutionary algorithms:

- At the initial stage, by making use of the prior beliefs concerning the parameters to be guessed (according to the economist's expertise and intuition);
- By successively restricting the search space from one iteration to another, providing an asymptotic convergence of the algorithm to some extreme point;
- By using two objective functions, instead of one.

In terms of their learning and convergence ability, the difference between RSG and GA and SA is that the last ones, even if able to estimate some parameters based on a very few number of observations, rely on infinite time convergence results only.

The RSG procedure is recalled, as presented in [Agapie,2009].

Procedure Repetitive Stochastic Guesstimation

1. Set the iteration index to zero: i=0
2. Choose some initial values and intervals for the parameters to be optimized
3. Choose/compute the initial value for the learning rate l_0
4. Randomly generate (guess) a new candidate solution, inside the current intervals
5. Compare the candidate solution vs. the current one-w.r.t. both criterions-and decide: accept or reject
6. If accepted, it becomes the current solution; otherwise, keep the old one.
7. Repeat 4-6 several times, until a better solution is obtained
8. i=i+1, decrease the learning rate, decrease the intervals' lengths and go to 4
9. Repeat 8, until STOP

2. Adjusting the RSG algorithm for achieving global convergence on the problem of estimating a GARCH (1,1) model

The next definitions and theorems are needed as a base for the coming formalization.

***Definition* 1**. Let $(\Omega, \Im, P)$ be a probability space and $\Im_0 \subseteq \Im_1 \subseteq \ldots \subseteq \Im$ be an increasing family of σ sub-algebras of $\Im$ and let $\Im_\infty = \sigma(\cup_t^\infty \Im_t) \subseteq \Im$. A stochastic process (X_t) that is

$\Im_t$-measurable is termed *supermartingale* if E(| X_t|)$\rightarrow\infty$ and E(X_{t+1}/ $\Im_t$)$\leq$ X_t for each positive integer *t*.

Condition (X_t) is $\Im_t$-measurable is fulfilled for example-and this is the case in this paper-if we consider for each *t*, $\Im_t=\sigma\{X_i / 0\leq i \leq t\}$, the smallest σ-algebra that makes all random variables up to X_t measurable.

***Definition* 2**. Let *f*: $\boldsymbol{R}^n\rightarrow\boldsymbol{R}$ and $a\in\mathbf{R}$. We denote the lower level set of *f*( )at level *a* by $L(a)=\{x\in\boldsymbol{R}^n / f(x)<a\}$.

***Definition* 3** Let *X* be a random variable and (X_n) a sequence of random variables defined on a probability space (Ω, $\Im$, P). Then (X_n) is said to converge *completely* to *X* if for any $\varepsilon>0$ $\lim_{n\rightarrow\infty}\sum_{i=1}^{n} P\left\{\left|X_i - X\right| > \varepsilon\right\} < \infty$. ($X_n$) is said to converge *in mean* to X if $\lim_{n\rightarrow\infty} E\left[\left|X_n - X\right|\right] = 0$.

According to [Neveu, 1975] , non-negative supermartingales (i.e., satisfying in addition $X_t \geq 0$ for all *t*) have the following remarkable property.

***Theorem* 1**. Let $(X_t)_t \geq 0$ be a non-negative supermartingale. Then $X_t\rightarrow X< \infty$ almost sure.□

Another important property of non-negative supermartingale follows from [Rudolph, 1997].

***Theorem* 2**. Let $(X_t)_t \geq 0$ be a non-negative supermartingale satisfying $E(X_{t+1}/\Im_t)\leq c_t X_t$ almost sure for all $t\geq 0$ with $c_t\geq 0$ and $\Sigma_{t=1..\infty}(\prod_{k=0..t-1} c_k)<\infty$. Then $X_t\rightarrow 0$ in mean and completely.□

The next concept was introduced in [Rudloph,1997] for characterizing real valued objective functions.

Derived from the previous two theorems, in [Agapie,2009] was proved the following result.

***Proposition* 1**. Let $(C_i)_{i=1..niter}$ be the sequence of coefficients generated by RSG at successive iterations and assume that $OF_1^i = OF_1^i (C_i)$ has bounded lower level sets and let OF_1* define the overall minimal value. Then the sequence $X_t=OF_1^i - OF_1$* defines a non-negative uniformly integrable supermartingale.□

Remark: If in Theorem 2 one takes $c_i=(1-l_i)^2$ (with l_i = sqrt(1-(*i*-1)/*niter*)) and defines replications in the RSG algorithm such that $E(OF_1^{i+1}/ \Im_i) \leq c_i\ OF_1^i$ then by Theorem2 and Proposition1 it comes that $OF_1^i \rightarrow OF_1^*$ when $i \rightarrow \infty$.

Consider $y_t \in N(0,h_t) \quad \left(\Rightarrow y_t = \sqrt{h_t}n_t\ , y_{t-1} = \sqrt{h_{t-1}}n_{t-1}, n_t, n_{t-1}\ IIDN(0,1)\right)$ (1)

with $h_t = \alpha_0 + \alpha_1\ y_{t-1}^2 + \beta_1\ h_{t-1}$ and

$$\alpha_0, \alpha_1\ , \beta_1\ > 0 \tag{2}$$

For every iteration i (Step 8 in previous *Procedure*), $i = \overline{1, niter}$, assume that replications are performed indexed by $r, r = \overline{1, nrepl}$. Step 4 in previous *Procedure* can be formalized as follows: for every iteration *i* and replication *r*, the current values for the coefficients are computed according to:

$$\alpha_{0;r}^{i+1} = \alpha_0^i + l_0^{\alpha_0} l_i\ u_r \tag{3}$$

$$\alpha_{1;r}^{i+1} = \alpha_1^i + l_0^{\alpha_1} l_i\ u_r \tag{4}$$

$$\beta_{1;r}^{i+1} = \beta_1^i + l_0^{\beta_1} l_i\ u_r \tag{5}$$

with u_r being a realization at the replication *r* of a uniformly distributed random variable ***u*** in the (-0.5; 0.5) interval, l_i a real non-negative number decreasing to zero when $i \rightarrow niter$ and $niter \rightarrow \infty$ (usually $l_i = \sqrt{1-\frac{i-1}{niter}}$). Also, α_0^i, α_1^i and β_1^i in (3)-(5) are the current values of the parameters at iteration *i* while $l_0^{\alpha_0}$, $l_0^{\alpha_1}$, $l_0^{\beta_1}$ are the initial interval lengths around the correspondent parameters.

Adapting RSG for estimating a GARCH (1,1) model will consist in dropping the assumption about the uniform distribution of the random variable ***u*** in the (-0.5, 0.5) interval and determine instead sufficient conditions regarding the distribution of the random variable **u**, the starting points $\alpha_0^0, \alpha_1^0, \beta_1^0$ as well as for corresponding intervals length around these values for achieving global convergence when number of iterations $niter \rightarrow \infty$.

Multiplying equation (2) by n_t^2 we get:

$$y_t^2 = \alpha_0 n_t^2 + \alpha_1 y_{t-1}^2 n_t^2 + \beta_1 y_{t-1} \frac{n_t^2}{n_{t-1}^2} \tag{6}$$

At iteration *i*+1 the objective function OF^{i+1} is computed according to:

$$OF^{i+1} = \min_{r=1,nrepl} \left(y_t^2 - \alpha_{0;r}^{i+1} n_t^2 - \alpha_{1;r}^{i+1} y_{t-1}^2 n_t^2 - \beta_{1;r}^{i+1} y_{t-1} \frac{n_t^2}{n_{t-1}^2} \right)^2$$

$$= \left(y_t^2 - \alpha_0^{i+1} n_t^2 - \alpha_1^{i+1} y_{t-1}^2 n_t^2 - \beta_1^{i+1} y_{t-1} \frac{n_t^2}{n_{t-1}^2} \right)^2 \quad (7)$$

α_0^{i+1}, α_1^{i+1}, β_1^{i+1} in (7) are coefficients at a certain replication *r* for which the minimum was achieved and the correspondent u_r in (3)-(5) will be simply denoted with *u*.

Even if the original version of RSG is considering two objective functions-since it can be checked that the weighted objective function asymptotically converges (when number of iterations goes to infinity) to the un-weighted one (see [Agapie, 20091])-the following considerations will rely only on the objective function indicated in (7).

Assuming that it would be possible to write down such an objective function, by knowing the numerical values corresponding to the standard normal variables entering y_t and y_{t-1},-this is n_t and n_{t-1}-its overall minimal value (denoted by OF^*) is equal to zero and it is achieved in the case of a perfect guess for the parameters $\alpha_0, \alpha_1, \beta_1$.

The sequence $\left(OF^i\right)_{i=\overline{1,niter}}$ can be regarded as a stochastic process indexed by iterations, assumed to be performed an infinite time (*niter*$\to\infty$). Global convergence of the sequence $\left(OF^i\right)_{i=\overline{1,niter}}$ is achieved by deriving sufficient conditions for $\left(OF^i\right)_{i\to\infty}$ to be a supermartingale (see the *Definition* 1) satisfying

$$E\left[OF^{i+1} / \Im_i\right] < c_i \, OF^i \quad (8)$$

$$\text{with } \prod_{i=1}^{\infty} \left(\sum_{k=0}^{i-1} c_k \right) < \infty \quad (9)$$

(take *t*=*i*, X_t=OF_i, *X*=*OF** in *Theorems* 1,2 and *Proposition*1 above) and $\Im_i$ the smallest σ-algebra that makes all random variables up to OF^i measurable.

Perhaps it is worth noticing that by "global convergence" is meant the convergence in mean and completely (see for convenience the definitions in the Appendix) of the sequence $\left(OF^i\right)_{i=\overline{1,niter}}$ to its overall minimal value *OF**.

Assuming that $l_0^{\alpha_0} = \alpha_0^i$, $l_0^{\alpha_1} = \alpha_1^i$, $l_0^{\beta_1} = \beta_1^i$ in (3)-(5), then (7) can be written in the equivalent form:

$$OF^{i+1} = \left[\sqrt{OF^i} \left(1 + l_i \, u\right) - l_i \, u \, y_t^2 \right]^2 \quad (10)$$

Then

$$E\left[OF^{i+1} / \Im_i\right] = OF^i\, E\left[\left(1 + l_i\, u\right)^2\right] - 2\sqrt{OF^i}\, l_i\, y_t^2\, E\left[u\left(1 + l_i\, u\right)\right] + l_i^2\, y_t^4\, E\left[u^2\right] \quad (11)$$

Next step is to determine the random variable *u* in (11) such that inequality (8) and (9) are satisfied.

According to [Agapie, 2009], this will lead to the conclusion that OF^i converges in *mean* and *completely* to zero, when number of iterations *i* is increasing to infinity.

The two sufficient conditions for (8) and (11) to hold together are:

$$E\left[\left(1 + l_i\, u\right)^2\right] - c_i \le 0 \quad (12)$$

and

$$E^2\left[u\left(1 + l_i\, u\right)\right] \le E\left[\left(1 + l_i\, u\right)^2\right] - c_i \quad (13)$$

Conditions (12) and (13) uniquely determine c_i as being equal to $E\left[\left(1 + l_i\, u\right)^2\right]$ if and only if the random variable *u* is chosen in such a way that $E\left[u\left(1 + l_i\, u\right)\right]$=0.

If the following notations are to be used, namely $a = E[u]$ and $b = Var[b]$ the a sufficient condition for u to satisfy (12) and (13) is

$$a + l_i\left(b^2 + a^2\right) = 0 \quad (14)$$

and, therefore $c_i = E\left[\left(1 + l_i\, u\right)^2\right] = 1 + l_i^2\left(a^2 + b^2\right) + 2\, l_i\, a = 1 + l_i\, a$ (15)

Values of c_i are also an indicator of the rate of convergence from iteration *i* to iteration *i*+1.As a consequence, a choice of $a = E[u]$ closer to $-\frac{1}{l_i}$ and negative will assure that (14) is fulfilled and the rate of convergence c_i (besides the fact it satisfies condition (9)) is increasing when $i \to \infty$.

There is no unique determination for a random variable *u* such as (8), (9), (14) and (15) are satisfied. One, for example, could sample *u* from a normal N(*a*,*b*) distribution with $a = -\frac{1}{2 l_i}$, $b = \frac{-a - l_i\, a^2}{l_i} = \frac{1}{2 l_i}$ and get an algorithm decreasing exponentially to its overall minimal value at a rate of $\frac{1}{2}$. It is also worth noting that parameters *a* and *b* can also be chosen such that the nonnegative condition for the parameters $\alpha_0, \alpha_1, \beta_1$ is assured.

For completing this discussion, previous calculations show that if one would know precisely the realization of a standard normal entering in y_t (this is n_t), then one could write the previous 'perfect' objective function OF^{i+1} in (7), converging to the overall minimal value *OF**-zero. This leads to an asymptotically convergence to the '*true*' coefficients of the model.

To get closer to this desirable case, the practitioner can sample T (T→∞) times from the 'ideal' distribution of *OF**, simply computing for every *mc* ∈{1,…T} and for a realization $n_{t;mc}$ of a standard normal

$$OF_{mc}^{i+1} = \min_{r=1,nrepl} \left(y_t^2 - \alpha_{0;r}^{i+1} n_{t;mc}^2 - \alpha_{1;r}^{i+1} y_{t-1}^2 n_{t;mc}^2 - \beta_{1;r}^{i+1} y_{t-1} \frac{n_{t'mc}^2}{n_{t-1;mc}^2} \right)^2 \tag{16}$$

$$= \left(y_t^2 - \alpha_0^{i+1} n_{t;mc}^2 - \alpha_1^{i+1} y_{t-1}^2 n_{t;mc}^2 - \beta_1^{i+1} y_{t-1} \frac{n_{t;mc}^2}{n_{t-1;mc}^2} \right)^2$$

finds its correspondent minimal value, say OF_{mc}^* and then consider, as the best estimates for the coefficients in the GARCH model, those corresponding to the minimal value in the row $(OF_{mc}^*)_{mc=\overline{1,T}}$. Of course, all the 'partial' overall minimal should be computed using the same initial conditions in the RSG algorithm.

3. Conclusions and Further Research Directions

Economists attempting to build econometric or forecasting models are frequently restricted due to data scarcity in terms of short time series of data and also of parameter non-constancy and under specification. In this case, a realistic alternative is often to guess rather than to estimate parameters of such models.

An algorithm of repetitive guessing (drawing) parameters from iteratively changing distributions, for minimizing the squares of ex post predictions errors, weighted by penalty weights and subject to a learning process, has recently been introduced (see Charemza, 2002). Despite obvious advantages, especially when applied to undersized empirical models with a large number of parameters, applications of Repetitive Stochastic Guesstimation (RSG) have been, so far, limited. This has presumably been caused by the lack of rigorous proof of its convergence. Such proof for a class of linear models, both identifiable (in the economic sense) and not, is provided in Agapie, Ad (2009). Another proof of convergence for the parameters in a GARCH (1,1) model is provided in this paper. Both proofs for convergence show that an adjustment for estimating parameters in a model (linear or nonlinear) based on a minimal set of observations is possible to be performed in such a way that proofs for global convergence should not rely on the data set but on the number of iterations of the algorithm. By being dependent on some computational capabilities - namely, number of iterations and replications performed - the Large Numbers Law might be applied in a different dimension. An alternative towards waiting until historical data series are recorded (while the underlying process may change several times) - for estimating a regression line or a GARCH(1,1) model, is to use computers, RSG, and experts for correctly extracting information from the most recent data.

It is worth noticing that the way this algorithm works resembles the Bayesian framework. The difference is that, instead of expressing priors on coefficients in terms of distribution functions, only initial point values are requested for start. Every iteration, the distributions over coefficients are determined according to the same sufficient conditions for the global convergence.

There is a need to develop careful Monte Carlo experiments to test the rates of convergence for the RSG parameters' estimators, design experiments to test the dependence of the final estimations on the starting points and also on the initial interval considered. These experiments will be first done on a minimal set of observations (two observations) and then, the dependence of the RSG's estimators on the sample size will be considered empirically. This dependence was not considered from a theoretical point of view, yet that is necessary in order to compare this technique with the traditional methods of estimating parameters in GARCH models (maximum likelihood methods).

Acknowledgements. We owe much gratitude to Wojciech Charemza for providing a consistent stream of scientific support and encouragement during this research.

References

Agapie, A. (2009), "Convergence of the Guesstimation Algorithm", *Communications in Statistics: Theory and Methods*, 38(5): 711-718.

Agapie, A. (2009), "*Computational Intelligence Techniques in Econometric Modeling - The Case of Romania"*, LAP Lambert Academic Publishing AG&Co.

Agapie, A., Badin, L. (2009), "Confidence Intervals for the Guesstimation Algorithm: A Bootstrap Approach", *Economic Computation and Economic Cybernetics Studies and Research*, 43(3): 167-178.

Agapie, A. (1999), "Stochastic Optimization in Econometric Models - A Comparison of GA, SA and RSG", *Research Memorandum ACE Project-MEET IV*, University of Leicester, no. 99/1

Charemza, W. (2002), "Guesstimation", *Journal of Forecasting,* 21: 417:433.

Davis, T.E. (1991), "Toward an extrapolation of the simulated annealing convergence theory onto the simple genetic algorithm", *Ph.D. Dissertation*, University of Florida.

Goffe, W. Ferrier, G.D. and Rogers, J. (1994), "Global optimization of statistical functions with simulated annealing", *Journal of Econometrics*, 60.

Haario, H. and Saksman, E. (1991), "Simulated annealing process in general state space", *Advances in Applied Probability*, 23.

Mills, T.C. (1999), "*The Econometric Modeling of Financial Time Series"*, Second Edition, Cambridge University Press.

Neveu, J. (1975), "*Discrete-Parameter Martingales"*, Amsterdam and Oxford: North Holland.

Przechlewski, T. and Strzala, K. (1996), "Evaluation and Comparison of the RSG and RSF-GA", ACE Project Research memoranda, 17.

Rudolph, G. (1996), "*Convergence Properties of Evolutionary Algorithms:*, Verlag Dr. Kovac.

Sargent, T. J. (1995), "*Bounded Rationality in Macroeconomics"*, Oxford University Press.

Simon, H.A. (1957), "*Models of Man: Social and Rational; Mathematical Essays on Rational Human Behavior in Society Setting"* , New York: John Wiley.

Weigend, A.S. and Gershenfeld, N. A. (eds.) (1994), "*Time series prediction. Forecasting the future and understanding the past"*, Addison Wesley.