

Convergence of GARCH Estimators: Theory and Empirical Evidence*

Peter Winker*

Dietmar Maringer[†]

* Faculty of Economics, Justus Liebig University Giessen,
Licher Strasse 64, D-35394 Giessen, Germany.

[†] Centre for Computational Finance and Economic Agents (CCFEA), University of Essex,
Wivenhoe Park, Colchester CO4 3SQ, United Kingdom.

Abstract

This paper combines the analysis of convergence of maximum likelihood estimators with the analysis of the convergence of stochastic optimization algorithms, e.g. threshold accepting, to the theoretical estimator. We discuss the joint convergence of estimator and algorithm in a formal framework.

An application to a GARCH model demonstrates the approach in practice by estimating actual rates of convergence through a large scale simulation study. Despite of the additional stochastic component introduced by the use of an optimization heuristic, the overall quality of the estimates turns out to be superior compared to conventional approaches.

Keywords: GARCH; Threshold Accepting; Optimization Heuristics; Convergence.

1 Introduction

The convergence of estimators, e.g. maximum likelihood (ML) estimators, for increasing sample size is well understood in many cases. However, even when the rate of convergence of the estimator is known, practical application is hampered by the fact, that the estimator cannot always be obtained at tenable computational cost. In fact, the literature mentions many estimation problems, where standard optimization methods fail to provide a reliable approximation to the theoretical estimator [3–5, 7, 10]. Often, this failure of standard methods is not due to a suboptimal implementation of the algorithms, but results from the inherent computational complexity of the problems and has to be taken as given [11, pp. 57ff].

However, if the theoretical estimator has to be replaced by some numerical approximation, the actual rate of convergence might differ from the theoretical one. In fact, if the implementation of the estimator is such that it will not

converge to the theoretical estimator with the sample size growing to infinity, the convergence properties of the estimator get lost. Unfortunately, many real life implementations of complex estimators cannot guarantee to result in the true theoretical estimator or, at least, a close approximation. Thus, if these methods work fine, the theoretical convergence results apply, if they fail, no statement on convergence can be provided.

The picture changes when the algorithm for calculating the estimator itself might be subject to a stochastic analysis. In particular, when it can be proven that the result found by the heuristic converges to the theoretical estimator with an increasing number of iterations, a joint convergence analysis becomes feasible [12]; this is the case for the threshold accepting (TA) heuristic considered for the ML estimation of the parameters of a GARCH model. The estimation problem and the application of TA to this problem is described by Maringer [9, pp. 63ff]. This contribution derives and analyzes the joint convergence properties of the optimization algorithm and the ML estimator.

2 Convergence of Optimization Based Estimators

2.1 Notation

We assume that the true model for the data generating process is known except for the values of a number of parameters collected in the true parameter vector ψ^{TR} . In particular, we will not consider issues related to model misspecification. For a given data sample consisting of T observations, let $\psi^{\text{ML},T}$ denote the value of the theoretical estimator, e.g., the ML estimator for the GARCH model. This vector cannot be observed unless a deterministic algorithm is available which provides the estimator with certainty. This condition is not fulfilled for the GARCH model when relying on standard optimization tools [3].

When a stochastic optimization heuristic like TA has to be used to obtain an approximation of the estimator, only

*We are indebted to Manfred Gilli, anonymous referees and conference and seminar participants for valuable comments on a preliminary draft of this paper.

one or several realizations of this stochastic procedure can be observed. The quality of these realizations will depend on the computational effort spent on the optimization process which, for TA, can be measured by the number of iterations, I . Thus, if the optimization is run R times with I iterations per run, we obtain R approximations of $\psi^{\text{ML},T}$, denoted as $\psi^{T,I,r}$, where $r = 1, \dots, R$.

Now the different aspects of convergence estimators can be discussed. First, asymptotic consistency of the theoretical estimator $\psi^{\text{ML},T}$ with regard to sample size T has to be established. Second, we have to demonstrate that based on the approximations $\psi^{T,I,r}$ it is possible to obtain convergence in probability towards $\psi^{\text{ML},T}$ as I goes to infinity. Finally, we have to show that both results can be combined to obtain a convergence result for the estimator found by TA. In particular, we have to provide a relationship $I(T)$ resulting in a convergence in probability of an estimate based on the $\psi^{T,I,r}$ towards the true parameter vector ψ^{TR} .

2.2 Convergence of the Estimator

The ML estimator of the GARCH model $\psi^{\text{ML},T}$ converges with the standard rate \sqrt{T} to the true parameter vector ψ^{TR} and is asymptotically normally distributed if the usual regularity conditions are satisfied [8, p. 202]. Equivalently, for any given probability $\delta > 0$ and level of accuracy $\varepsilon > 0$, there exists a sample size $T(\delta, \varepsilon)$ such that for any $T \geq T(\delta, \varepsilon)$ we find $\text{P}(|\psi^{\text{ML},T} - \psi^{\text{TR}}| < \varepsilon) > 1 - \delta$. In fact, for given δ , asymptotically, T has to be chosen proportional to $1/\varepsilon^2$ to obtain this result.

2.3 Convergence of Threshold Accepting

Threshold accepting (TA) is a heuristic optimization method where a solution is repeatedly modified and updated in a stochastic fashion [6]. Consequently, repeated runs of the optimization heuristic on a single problem instance will result in a distribution of results $\psi^{T,I,r}$.

For the TA implementation there exist suitable parameters such that the global optimum of the objective function can be approximated at arbitrary accuracy with any fixed probability close to one by increasing the number of iterations [1]. If the search of parameters $\psi^{T,I,r}$ is restricted to a compact set, the continuity of the likelihood function implies that for any given $\delta > 0$ and $\varepsilon > 0$, there exists a number of iterations $I(\delta, \varepsilon)$ such that $\text{P}(|\psi^{T,I,r} - \psi^{\text{ML},T}| < \varepsilon) > 1 - \delta$ for any $r = 1, \dots, R$. Obviously, the convergence of the first order statistic of $\psi^{T,I,r}$, $r = 1, \dots, R$ will also satisfy this condition – potentially for a smaller value of I . While this results allows no general derivation of the required number of iterations $I(\delta, \varepsilon)$, it will be left to the analysis of our empirical implementation to demonstrate that $I(\delta, \varepsilon)$ can be chosen to be a function of $T(\delta, \varepsilon)$ growing at a less than linear rate.

2.4 Joint Convergence

Though the stochastic feature of TA might appear like a drawback on first sight, it actually overcomes the limitations of traditional deterministic methods: even if the latter would report inferior solutions, heuristics still have a fair chance of finding the optimal solution. Based on the previous definitions and if $I(T(\delta, \varepsilon))$ meets certain criteria, we find $\text{P}(|\psi^{T,I,R} - \psi^{\text{TR}}| < \varepsilon) > 1 - \delta$ where $\psi^{T,I,R}$ is the best result out of R replications, i.e., convergence of the heuristic optimization based estimator to the true parameter value for T going to infinity and I going to infinity as a function of T .

3 Application to GARCH Model

3.1 Model and Data for the Computational Study

As a benchmark implementation for assessing the performance of the estimation method in practice, we consider the basic GARCH(1,1) model $r_t = \psi_0 + e_t$ with $e_t \sim N(0, \sigma_t^2)$ where $\sigma_t^2 = \psi_1 + \psi_2 e_{t-1}^2 + \psi_3 \sigma_{t-1}^2$. For the empirical application, we refer to the estimates obtained by [2] based on 1974 daily observations for the changes in the German mark / British pound exchange rate. Their estimates of the parameters of the GARCH(1,1) are the following (using our notation):

$$\psi^* = [-0.00619041; 0.0107613; 0.153134; 0.805974]. \quad (1)$$

We use this model and parameters for a data generating process (i.e., $\psi^{\text{TR}} = \psi^*$) and produce 100 time series each consisting of 2100 observations. For the computational study, we then removed the first 100 observations. From the remaining series, we analyzed the first T observations with $T = 50, 100, 200, 400, 1000, \text{ and } 2000$.

3.2 The Optimization Heuristic

For finding the parameters that maximize the loglikelihood function of the GARCH model, $L = -\frac{T}{2} \ln(2\pi) - \frac{1}{2} \sum_{t=1}^T (\ln(\sigma_t^2) + e_t^2/\sigma_t^2)$, we use the TA implementation suggested in Maringer [9, chapter 2]. Starting with a random initial solution, in each of the I iterations, one of the parameters is randomly picked and slightly changed by adding a small random value. If this change improves the solution (i.e., increases the likelihood function) or does not impair it more than a given threshold, the new solution replaces the previous one; if the impairment exceeds the threshold, the change is rejected. While rather tolerant in early iterations, the threshold is gradually lowered towards zero; similarly, u is gradually reduced.

4 Results

4.1 Notation

Comparing six different values for T as well as for I and 100 data series adds up to 3 600 different optimization problems. For each of these problems, the algorithm was run approximately $R \approx 1700$ times, resulting in a total of 6 146 393 reported solutions. For evaluation purposes, we then computed the mean squared deviation between the reported parameters and the true (TR) and the ML parameters, respectively:

$$\text{MSD}_p^{\text{TR},d,T,I} = \frac{1}{R} \cdot \sum_{r=1}^R \left(\psi_p^{d,T,I,r} - \psi_p^{\text{TR}} \right)^2 \quad (2)$$

$$\text{MSD}_p^{\text{ML},d,T,I} = \frac{1}{R} \cdot \sum_{r=1}^R \left(\psi_p^{d,T,I,r} - \psi_p^{\text{ML},d,T} \right)^2 \quad (3)$$

where $\psi_p^{d,T,I,r}$ is the p -th element of the optimal parameter vector for the data series d with T observations reported in the r -th run and found within I iterations. Figure 1 illustrates the results for the MSD's for the GARCH parameter ψ_3 for one specimen data series d as a function of the sample size T and the number of iterations I used in the TA implementation.

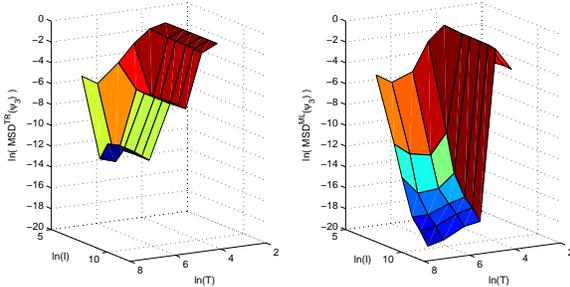


Figure 1: MSD^{TR} (left) and MSD^{ML} (right) for ψ_3 for one specimen data series

4.2 Convergence of the Estimator

Figure 2 depicts the distribution of the optimal parameters for the 100 data series in dependence of T . Though the reported values for ψ_1 and ψ_3 , are not symmetrically distributed, they converge towards their true value when increasing T .

This convergence can be expected when individual data series are considered. The left graph in Figure 1 suggests a linear relationship between the logs of T and MSD^{TR} , i.e., that the ML parameters tend to converge to the true values when a time series is prolonged. Though in particular for small data sets, this convergence is not necessarily smooth, the slope of the linear relationship provides an estimate of the rate of convergence. In order to isolate

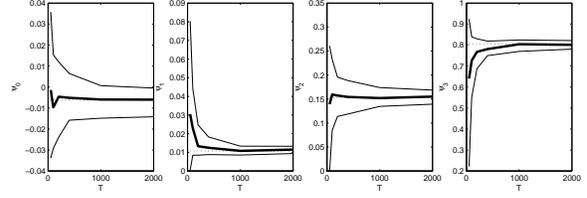


Figure 2: Median and 25% and 75% quantiles of the ML estimators of the 100 data series

effects of the optimization heuristic, we group the data by the number of the algorithm's iterations, I , and then estimate for each data series $d = 1 \dots 100$ the parameters of the model $\ln(\text{MSD}_p^{\text{TR},d,T,I}) = a_p^{d,I} + b_p^{d,I} \cdot \ln(T)$.

On average, the mean rate of convergence of $\text{MSD}_p^{\text{TR},d,T,I}$ as a function of T is found to be approximately of the order of T^{-1} (corresponding to the usual rate of convergence of $T^{-1/2}$ for the parameters) or even faster for ψ_1 and ψ_3 with a rates up to $T^{-1.9}$; in more than half of the cases, these parameters are statistically significant. The relatively large standard deviations of the parameters $b_p^{d,I}$ indicate that their values (and thus the convergence rates) can differ substantially between different realizations of the data generating process. We found two main reasons for this: Either the whole of the data series contains few or no outliers (and the MSD will be low for any T , i.e., no further convergence is possible), or short data series are extended with some extreme outliers (which first increases and eventually decreases the MSD).

On the other hand, if the extreme values of a data series are concentrated in the first observations, then for short T , ψ^{ML} will differ substantially from ψ^{TR} . Adding further observations are then likely to quickly drive the optimal parameters ψ^{ML} towards ψ^{TR} , and the convergence rate will be substantially above average. Thus, the high standard deviation is mainly due to small sample effects.

4.3 Convergence of Threshold Acceptance

Analyzing the convergence of ψ^{ML} requires reliable estimates of these parameters. In TA, an indicator for this reliability is the number of conceded iterations. Assuming a linear relationship between the logs of the MSD^{ML} and the number of iterations, I , a model of the type $\ln(\text{MSD}_p^{\text{ML},d,T,I}) = a_p^{d,T} + b_p^{d,T} \cdot \ln(I)$ can be estimated for each data series d and fixed length T . In particular for long data series, the mean rate of convergence of $\text{MSD}_p^{\text{ML},T}$ as a function of I is found to be of the order I^{-1} up to I^{-3} . Though the convergence rate differs between data series, these differences diminish the longer the time series become. In many (or, for large T , virtually all) cases, this relationship is statistically significant.

4.4 Joint Convergence

The empirical results on the convergence of the ML estimator (subsection 4.2), in particular the estimated rate of convergence confirm the asymptotic theory. Furthermore, when I is chosen at least proportional to T , the TA approximation of this estimator (subsection 4.3) will converge at the same rate to the ML estimator. As indicated in subsection 2.4, the following joint convergence property results: There exists a constant λ_I such that if I is chosen to be $\lambda_I T$, the TA approximation $\psi_p^{d,T,I,r}$ to ψ_p^{TR} satisfies the convergence condition for any given probability $1 - \delta$ and any $\varepsilon > 0$ when T grows at a rate proportional to $1/\varepsilon^2$. In other words, the TA based ML estimator of the GARCH model parameters converges in probability to the parameters of the data generating process at the same rate as the theoretical ML estimator. Thus, the additional stochastic component introduced by the use of a stochastic search heuristic does not destroy the convergence properties of the ML estimator.

5 Conclusions

For many estimation tasks, deterministic algorithms will not always provide the theoretical estimator. In this case, optimization heuristics might be an adequate solution – although (or because) the stochastic features of these algorithms introduce an additional source of uncertainty to the estimator. If the search heuristic converges to the theoretical estimator for the number of iterations going to infinity, it is possible to derive a joint convergence result. We introduce such a convergence result for TA applied to ML estimation. Unfortunately, so far, no distributional results are available for the approximation by TA. Nevertheless, convergence in probability is a strong result as compared to standard algorithms.

We find for the GARCH model that the theoretical joint convergence result holds for the application already when setting the number of iterations of the algorithm proportional to the sample size. Thus, the TA based estimator has superior convergence properties compared to standard approaches. It generates better and more robust results already for small samples.

Overall we conclude that the use of stochastic optimization tools in econometrics is indicated whenever standard tools fail to generate reliable results. These tools might provide benchmarks when the quality of standard methods is unknown. Our empirical results rest on a single data generating process, hence further evidence is required to assess the robustness of our findings. It would be highly interesting to derive the distribution of the results obtained by the optimization tool instead of a convergence in probability result. These extensions are left for future research.

References

- [1] I. Althöfer and K.-U. Koschnik. On the convergence of threshold accepting. *Applied Mathematics and Optimization*, 24:183–195, 1991.
- [2] T. Bollerslev and E. Ghysels. Periodic autoregressive conditional heteroscedasticity. *Journal of Business and Economic Statistics*, 14(2):139–151, 1996.
- [3] C. Brooks, S. P. Burke, and G. Persaud. Benchmark and the accuracy of GARCH model estimation. *International Journal of Forecasting*, 17:45–56, 2001.
- [4] M. P. Clements and H.-M. Krolzig. A comparison of the forecast performance of Markov-switching and threshold autoregressive models. *Econometrics Journal*, 1:C47–C75, 1998.
- [5] B. Dorsey and W. J. Mayer. Genetic algorithms for estimation problems with multiple optima, nondifferentiability and other irregular features. *Journal of Business and Economic Statistics*, 13:53–66, 1995.
- [6] G. Dueck and T. Scheuer. Threshold accepting: A general purpose algorithm appearing superior to simulated annealing. *Journal of Computational Physics*, 90:161–175, 1990.
- [7] B. Fitzenberger and P. Winker. Using threshold accepting to improve the computation of censored quantile regression. In R. Payne and P. Green, editors, *COMPSTAT 1998, Proceedings in Computational Statistics*, pages 311–316, Heidelberg, 1998. Physica.
- [8] H. Herwartz. Conditional heteroskedasticity. In H. Lütkepohl and M. Krätzig, editors, *Applied Time Series Econometrics*, pages 197–221. Cambridge University Press, Cambridge, 2004.
- [9] D. Maringer. *Portfolio Management with Heuristic Optimization*. Springer, Berlin, 2005.
- [10] B.D. McCullough and Berry Wilson. On the accuracy of statistical procedures in Microsoft Excel 2003. *Computational Statistics & Data Analysis*, 49(4):1244–1252, June 2005.
- [11] P. Winker. *Optimization Heuristics in Econometrics: Applications of Threshold Accepting*. Wiley, Chichester, 2001.
- [12] P. Winker. The stochastics of threshold accepting: Analysis of an application to the uniform design problem. working paper 2005-003E, University of Erfurt, 2005.