The behavior of trust-region methods in FIML estimation

Weihs, Claus and Calzolari, Giorgio and Panattoni, Lorenzo

IBM Scientific Center, Pisa, Italy

1986

The Behavior of Trust-Region Methods in FIML-Estimation*

C. Weihs, Basel, G. Calzolari, and L. Panattoni, Pisa

Abstract — Zusammenfassung

The Behavior of Trust-Region Methods in FIML-Estimation. This paper presents a Monte-Carlo study on the practical reliability of numerical algorithms for FIML-estimation in nonlinear econometric models. The performance of different techniques of Hessian approximation in trust-region algorithms is compared regarding their "robustness" against "bad" starting points and their "global" and "local" convergence speed, i.e. the gain in the objective function, caused by individual iteration steps far off from and near to the optimum.

Concerning robustness and global convergence speed the crude GLS-type Hessian approximations performed best, efficiently exploiting the special structure of the likelihood function. But, concerning local speed, general purpose techniques were strongly superior. So, some appropriate mixtures of these two types of approximations turned out to be the only techniques to be recommended.

AMS Subject Classifications: 62P20, 65C05, 65U05.

Key words: Econometrics, Monte-Carlo methods, numerical methods, trust-region methods, FIML-estimation.

Das Verhalten von Trust-Region-Algorithmen zur FIML-Schätzung. Diese Arbeit beschreibt eine Monte-Carlo-Studie über die praktische Verläßlichkeit numerischer Algorithmen zur FIML-Schätzung in nichtlinearen ökonometrischen Modellen. Dabei wird die Güte verschiedener Hessematrixnäherungen in trust-region Algorithmen verglichen hinsichtlich der „Robustheit" gegenüber „schlechten" Startwerten und hinsichtlich „globaler" und „lokaler" Konvergenzgeschwindigkeit, d.h. der Größe der Verbesserung der Zielfunktion bei Iterationsschritten weit entfernt bzw. in der Nähe vom Optimum.

Während sich GLS-Typ Näherungen der Hessematrix hinsichtlich Robustheit und globaler Konvergenz als deutlich überlegen erweisen wegen ihrer effizienten Ausnutzung der speziellen Struktur der Likelihood-Funktion, konvergieren Verfahren, die für allgemeine Zielfunktionen entwickelt wurden, wesentlich schneller in der Nähe des Optimums. Für die praktische Anwendung erweisen sich daher lediglich geeignete „Mischungen" dieser beiden Näherungstypen als empfehlenswert.

0. Introduction

Recently more and more attention was paid to the development of numerical algorithms for the computation of full-information-maximum-likelihood (FIML) estimates in nonlinear interdependent econometric models. Several optimization techniques have been proposed in the last few years, utilizing all the well-known tools of numerical analysis. Search algorithms were proposed (see e.g. Parke (1982)), which do not make use of any derivatives. At the same time, Newton-like methods

* This work was mainly carried out during a research visit of C. Weihs at the Scientific Center of IBM Italia in Pisa.

were suggested (see e.g. Belsley (1980)), using not only first but also second order derivatives. Moreover, stressing the excessive cost required to calculate the Hessian, quasi-Newton methods were recommended, approximating the Hessian by various techniques. Some of these approximations were motivated by the special structure of the likelihood function (see e.g. Berndt, Hall, Hall and Hausman (1974), Amemiya (1977) or Dagenais (1978)), others exclusively by their numerical properties for general objective functions (see e.g. Belsley (1979)). Also combinations of such techniques were suggested (see e.g. Weihs (1985)).

Ten years ago the econometric model builder was lucky to have any algorithm to compute FIML-estimates. In the meantime, some guideline seems to be necessary to get an idea, which of all these algorithms might be reliable in practice. A first step into this direction was done in Calzolari, Panattoni (1985), where the performance of some of those algorithms was tested by means of a large number of Monte-Carlo experiments. Considering the "robustness" of the algorithms against "bad" starting points and the "gain" in the objective function, caused by individual iteration steps, this study supports the idea of a mixture of algorithms.

Motivated this way, the authors decided to investigate the relative performance of such mixed algorithms in more detail. This resulted in two papers. In Calzolari, Panattoni and Weihs (1985) different techniques of Hessian approximation were studied in the classical framework of line search algorithms (extending the work in Calzolari, Panattoni (1985)). In the present paper a modern "trust-region" algorithm serves as the basis for the implementation of approximation techniques (extending the work in Weihs (1985)). Objects of investigation are some *quasi-Newton approximations* to the Hessian and their "mixtures". Their performance will be judged on the basis of different Monte-Carlo statistics, characterizing both the "global" and the "local" convergence of the iteration process. "Global" convergence is related to the capability and speed of an algorithm to reach a neighborhood of the optimum, whereas "local" convergence is related to the speed near to the optimum. Surely an econometric model builder would at once agree to the relevance of "global" convergence. But "local" convergence is certainly just as important, since if an algorithm converges slowly near to optimum "we may never be able to see it converge" (cp. Dennis, Moré (1977), 50).

The plan of this paper is the following. In sections 1 and 2 the theoretical econometric model and the FIML-estimator are defined. In sections 3 and 4 the "trust-region" algorithm and the techniques of Hessian approximation are introduced. In sections 5 and 6 the Monte-Carlo experiment and the tested empirical econometric models are described. In section 7 the results of the Monte-Carlo study are discussed and in section 8 the paper is completed by a conclusion.

1. The Theoretical Econometric Model

Consider a multi-equation nonlinear interdependent econometric model which can be represented by:

$$\begin{aligned} Y_{it} &= f_i(Y_t, X_t, a^*) + \varepsilon_{it}, \quad i = 1, \ldots, m, \\ Y_{it} &= f_i(Y_t, X_t), \quad i = m+1, \ldots, M, \; t = 1, \ldots, T, \end{aligned} \tag{1}$$

where $Y_t := (Y_{1t} \dots Y_{Mt})$ is the vector of jointly dependent variables, the $X_{jt}, j = 1, \dots, P$, are predetermined variables (possibly including lagged jointly dependent variables), a^* is the (unknown) vector of "true" (functional) parameters and the ε_{it} are error terms assumed to be independently and identically $N(0, \Sigma)$-distributed with Σ unknown but assumed to be positive definite. The first m equations are called behavioral equations, the others identities.

2. The FIML-Estimator

If observations y_{it}, $i = 1, \dots, M$, x_{jt}, $j = 1, \dots, P$, $t = 1, \dots, T$, are fixed for the model variables, then the *model error*

$$e_{it} := y_{it} - f_i(y_t, x_t, a), \quad i = 1, \dots, m, \tag{2}$$

gives a measure for the goodness of fit corresponding to a parameter vector a (usually $\neq a^*$). The *FIML-estimator* $\hat{a}$ for a^* can be computed by minimizing the (negative) logarithmic concentrated likelihood function

$$K(a) := \frac{T}{2} \log\left(\det \hat{\Sigma}(a)\right) - \sum_{t=1}^{T} \log\left|\det\left(J_t(a)\right)\right|, \tag{3}$$

where

$$\hat{\Sigma}(\hat{a}) := \frac{1}{T} \sum_{t=1}^{T} e_t(\hat{a})\, e_t(\hat{a})'$$

is the FIML-estimator for the covariance matrix Σ, if $\hat{a}$ is the FIML-estimator for a^* (cp. 1.), and where $J_t(a)$ is the Jacobian of e_t with regard to the variables $y_{1t}, \dots, y_{mt}$ (taking into account the identities, if any, see Dagenais (1978), 1354 – 5).

3. The Minimizing Algorithm

Since the minimization problem is extremely non-linear, the algorithm used is iterative, generating in every step an approximation a_k to the optimum $\hat{a}$, starting from a user supplied starting vector a_0. The algorithm is part of a modern class of numerical algorithms, the so-called trust-region algorithms. Such algorithms were first discussed by Powell (1975).

They fit a quadratic

$$Q_k(a_k + s) := K(a_k) + g_k' s + \frac{1}{2} s' H_k s \tag{4}$$

to the likelihood function K in the approximation a_k, where g_k is the gradient vector of K in a_k and H_k is an approximation of the Hessian of K in a_k. This quadratic is then minimized in a neighbourhood N_k of a_k, the so-called trust-region. This generates an updating step s_k. Obviously, the usefulness of s_k depends on the goodness of the fit of the quadratic Q_k to the objective K in the trust-region N_k. Intending to obtain the maximum gain in the objective K, the trust-region N_k possibly should not exclude any area, where K is reasonably approximated by the

quadratic Q_k. Therefore, starting from some initial region N_0, the size of N_k is flexibly updated depending on the correspondence of the reductions in Q_k and K caused by step s_k. If the reduction in K is "much smaller" than expected by the reduction in Q_k, s_k is rejected and the trust-region is reduced because of bad fit, hoping for a more trustworthy approximation in the new region. In the case of reasonable fit the trust-region is extended for a trial, and

$$a_{k+1} := a_k + s_k \tag{5}$$

is taken to be the new approximation to the optimal parameter vector. Note that line searches are avoided in this way. Further note that the updating step s_k is either equal to the Newton-like step

$$s_k^N := -H_k^{-1} g_k, \tag{6}$$

or, if H_k is not invertible or s_k^N is not a feasible point, s_k is a Levenberg-Marquardt step such that a_{k+1} is placed near to the boundary of the trust-region. A special feature of the implemented trust-region algorithm is its automatic scaling procedure, which allows econometric models to be not well scaled.[1] For the selection of H_k various approaches have been tested.

4. Approximations to the Hessian

First and second derivatives of K have the following representations.

$$\frac{\partial K}{\partial a}(a)' = G(a)' S(a)^{-1} e(a), \text{ where}$$

$$G(a) := (G_1(a) \ldots G_n(a)),$$

$$G_i(a) := \frac{\partial e}{\partial a_i}(a) - \frac{1}{T} R_i(a)' e(a), e(a) := (e_1(a)' \ldots e_T(a)')', \tag{7}$$

$$R_i(a) := B_i(a) \otimes I, \; B_i(a) := \sum_{t=1}^{T} J_t(a)^{-1} \frac{\partial J_t}{\partial a_i}(a),$$

$$\text{and } S(a) := \hat{\Sigma}(a) \otimes I;$$

$$\begin{aligned} \frac{\partial^2 K}{\partial a_j \partial a_i}(a) = {} & \frac{\partial e}{\partial a_i}(a)' S(a)^{-1} \frac{\partial e}{\partial a_j}(a) \\ & - \frac{\partial e}{\partial a_i}(a)' S(a)^{-1} \frac{\partial S}{\partial a_j}(a) S(a)^{-1} e(a) \\ & + \frac{\partial^2 e}{\partial a_j \partial a_i}(a) S(a)^{-1} e(a) - \frac{1}{T} e(a)' \frac{\partial R_i}{\partial a_j}(a) S(a)^{-1} e(a). \end{aligned} \tag{8}$$

For the approximation of the exact Hessian as given in (8) only first order information will be used.

---

[1] For more details see Weihs (1985). For a summary of results on trust-region algorithms see More (1983).

Amemiya (1977, 962) proposed the following Hessian approximation:

$$H_A(a) := G(a)' S(a)^{-1} \frac{\partial e}{\partial a}(a). \tag{9}$$

This matrix asymptotically coincides with the exact Hessian for linear models (see Amemiya (1977), 963). Dagenais (1978) experimented with a slightly different approximation, also mentioned by Amemiya (1977, 963):

$$H_D(a) := G(a)' S(a)^{-1} G(a). \tag{10}$$

Looking at (8), another variation of (9) seems to be natural (see Weihs (1985)):

$$H_W(a) := \frac{\partial e}{\partial a}(a)' S(a)^{-1} \frac{\partial e}{\partial a}(a). \tag{11}$$

Note that the last two approximations have the same form as the corresponding matrix used in Aitken-Zellner-estimation of linear models (cp. Dagenais (1978), 1354).

4.2 Exact Hessians for Linear Models

For models linear in the parameters, (8) simplifies to

$$\begin{aligned}(H_H(a))_{ij} := & \frac{\partial e}{\partial a_i}(a)' S(a)^{-1} \frac{\partial e}{\partial a_j}(a) \\ & - \frac{\partial e}{\partial a_i}(a)' S(a)^{-1} \frac{\partial S}{\partial a_j}(a) S(a)^{-1} e(a) \\ & + \frac{1}{T} e(a)' (C_{ij}(a) \otimes I) S(a)^{-1} e(a),\end{aligned} \tag{12}$$

$$\text{where } C_{ij}(a) := \sum_{t=1}^{T} J_t(a)^{-1} \frac{\partial J_t}{\partial a_j}(a) J_t(a)^{-1} \frac{\partial J_t}{\partial a_i}(a).$$

For nonlinear models one may hope that $H_H(a)$ is a better approximation to the Hessian than $H_W(a)$ (cp. (8), (11)). Note that no second order derivatives are used for the computation of H_H, but only first order derivatives and mixed derivatives concerning parameters and variables. In this sense, only first order information is used.

4.3 BFGS-Updating Formula

Up to this point, only approximations were described, which exploit the special structure of the likelihood function K. The following well-known BFGS-quasi-Newton-updating formula was established for the approximation of the Hessian in the general case (see e.g. Dennis, Moré (1977), 74):

Let $H_{B,k}$ be a positive definite matrix, then

$$H_{B,k+1} := H_{B,k} + \frac{v_k v_k'}{v_k' s_k} - \frac{H_{B,k} s_k s_k' H_{B,k}}{s_k' H_{B,k} s_k} \text{ is the BFGS-update, where}$$
$$v_k := \frac{\partial K}{\partial a}(a_{k+1}) - \frac{\partial K}{\partial a}(a_k),\ s_k := a_{k+1} - a_k \text{ (cp. (5))}. \tag{13}$$

4.4 Pure and Mixed Approximations

To combine the advantages of different Hessian approximations, the above approximation techniques are also mixed in some profitable way, as motivated (e.g.) in Weihs (1985, 12). Altogether, 5 distinct versions of the trust-region-algoritm 3. are investigated, differing only in the method of Hessian approximation:

(a) $H_k := H_W(a_k)$, $k = 0, 1, 2, \ldots$, (see (11) in 4.1). This approximation technique will be referred to as *GLS-solo*. Using H_A (see (9)) or H_D (see (10)) instead of H_W resulted in very similar convergence properties.

(b) $H_k := H_H(a_k)$, $k = 0, 1, 2, \ldots$, (see (12) in 4.2). This approximation technique will be referred to as *Hessian-solo*.

(c) $H_0 := H_W(a_0)$, $H_k := H_{B,k}$, $k = 1, 2, \ldots$, (see (13) in 4.3). This approximation technique will be referred to as *BFGS-solo*.

(d) $H_k := H_W(a_k)$, $k = 0, 1, 2, \ldots$, till $\| s_k \| / (1 + \| a_k \|) \leq 0.01$, then $H_k := H_{B,k}$. Obviously this is a mixture of (a), (c) and will be referred to as *BFGS-mix* (for theoretical properties cp. Weihs (1985)).

(e) $H_k := H_W(a_k)$, $k = 0, 1, 2, \ldots$, until $\| s_k \| / (1 + \| a_k \|) \leq 0.01$, then $H_k := H_H(a_k)$. Obviously this is the analogon of (d) for the Hessian approximation H_H which is exact for linear models. This technique will be referred to as *Hessian-mix*.

5. The Monte-Carlo (MC)-Experiment

To be able to compare the mean performance of the different versions of the algorithm using different Hessian approximations (see 3, 4), FIML-estimates for various real-world econometric models were computed for hundreds of artificially generated sets of model variables. The basic MC-experiment is the following[2]:

(i) The exogenous model variables in the reference period and the endogenous starting values (i.e. the realisations of lagged endogenous variables outside the reference period) are fixed to be equal to their "real-world-observations". For data generation, the parameter vector and the covariance matrix of the model residuals are fixed to be equal to the FIML-estimates $\hat{a}$ and $\hat{\Sigma}$ respectively, computed by using the observed endogenous variables also. Let $j := 1$.

---

[2] The MC-experiments were performed using the IAS-System Bonn, a software system developed to support econometric model building (cf. Weihs (1984)).

(ii) Determine realisations of the model residuals which are (pseudo-) $N(0, \hat{\Sigma})$ distributed, independently for all $t = 1, \ldots, T$.

(iii) Using the data from (i), (ii) determine realisation for the endogenous variables by solving the model for these variables for all $t = 1, \ldots, T$.

(iv) Using the data from (i) for the exogenous model variables and the endogenous starting values, and the data from (iii) for the endogenous variables in the reference period, compute OLS-estimates for the (functional) parameters a^* which serve as starting values $a_0^{(j)}$ for the computation of the FIML-estimates $\hat{a}^{(j)} = a_{Lj}^{(j)}$. During the computation of $\hat{a}^{(j)}$ record all the intermediate values $K_k^{(j)}$ of the objective function, $k = 0, \ldots, Lj$.

(v) Repeat steps (ii) – (iv) for $j = 2, 3, \ldots, N$.

For the comparison of the relative performance of different versions of the minimization algorithm two distinct aspects of the convergence process have been considered.

The *overall performance* (or *global convergence*) of the algorithm is represented by the evolution of the mean of the fraction of the distance between the values of the objective function in the starting point $a_0^{(j)}$ and the optimum $a_{Lj}^{(j)}$, which was covered after k iterations. To this end the optimal value is computed very exactly (with a tolerance of $5 \cdot 10^{-14}$).

To characterize the *convergence near the optimum* (*local convergence*) two different statistics are used. In both cases only the very last part of the course of iterations is considered. On the one hand the distribution of the number of iterations is reported which were required to reach the optimum, starting from that iteration where first more than 99.9% of the distance to the optimal value of the objective function was covered. On the other hand we report the mean number of digits of the optimum value of the objective function which was gained k iterations before the algorithm stopped.

The *analysis of the outcomes of the MC-experiment* can be formalized as follows:

(vi) Compute the error in the objective function after k iterations:

$$E_k^{(j)} := K_k^{(j)} - K_{Lj}^{(j)}, \; k = 1, \ldots, Lj, \; j = 1, \ldots, N.$$

(vii) Compute the mean of the relative reduction of the distance to the optimal value of the objective function, gained after k iterations:

$$\bar{R}_k := \frac{1}{N} \sum_{j=1}^{N} E_k^{(j)} / E_0^{(j)}, \; k = 1, \ldots, \max_j Lj,$$

where $E_k^{(j)} := 0$ for $k > Lj$. Then $EG_k := -\log_{10}(\bar{R}_k)$, $k = 1, \ldots, \max_j Lj$, is the number of digits up to which in iteration k the distance $E_0^{(j)} - E_k^{(j)}$ from the starting point coincides with the total distance $E_0^{(j)}$ on the average. The course of EG_k is utilized to characterize the mean global convergence of the algorithm.

(viii) Determine the number of iterations Nj where first more than 99.9% of the distance to the optimal value of the objective function was covered:

$E_{Nj}^{(j)}/E_0^{(j)} < 0.001$ and $E_{Nj-1}^{(j)}/E_0^{(j)} \geq 0.001$. The empirical distribution of the remaining iterations $Mj := Lj - Nj + 1$ characterizes the local convergence of the algorithm.

(ix) Compute the number of digits of the optimal value of the objective function, gained k iterations before the algorithm stopped:

$$D_k^{(j)} := -\log_{10}(E_{Lj-k}^{(j)}/|K_{Lj}^{(j)}|),\ k = 1, \ldots, \max_j Lj,$$

where $E_{Lj-k}^{(j)} := 0$ for $k > Lj$.

Then the course of the mean of these digits for small k:

$$EL_k := \frac{1}{N} \sum_{j=1}^{N} D_k^{(j)},\ k = 1, \ldots, 10 \text{ (say)},$$

characterizes the local convergence of the algorithm.

Note that the statistic EG_k will be strongly dominated by replications with untypically slow convergence. Therefore EG_k tends to be too pessimistic judging the mean global performance of the algorithm.

Further note that the MC-experiment can be easily generalized so that the exogenous model variables are independently, identically normally distributed with means and covariance-matrix estimated from historical data (cp. (i)). The results, however, should not be particularly sensitive to such different choice of the exogenous variables (cp. Calzolari, Panattoni (1984), 19).

6. Empirical Econometric Models

The MC-experiment described in 5. was applied to some "real-world" econometric models in order to judge the performance of the FIML-algorithm in relevant situations. These models can briefly be characterized as follows:

1. The well-known Klein-1 model has 6 equations, 3 of which stochastic, and 12 unknown functional parameters. The historical reference period has length $T = 21$ (see e.g. Theil (1971), 432 – 434, 456).

2. A strongly nonlinear version of the Klein-1 model was obtained by using a Cobb-Douglas-type consumption function (with additive error term) involving the same arguments as in the linear case.

3. A strongly nonlinear model for the German economy (see Weihs (1987)) consists of 30 equations, 5 of which stochastic with 18 unknown functional parameters. To investigate the influence of sample sizes T on convergence, $T = 21$ (model 3a) and $T = 27$ (model 3b) were used. Historical data were only available for the first period. For the extension to $T = 27$, exogenous variables were fixed to follow some reasonable course. For both sample sizes the FIML-estimates $\hat{a}$, $\hat{\Sigma}$, computed on the basis of historical data only ($T = 21$), were used to generate endogenous simulation data in the MC-experiment (cp. 5 (i)).

The MC-experiments on models 1, 2, 3a, 3b were performed using $N=100$ replications each (cp. 5 (v)). In each replication all the 5 versions of the trust-region-algorithm 3., described in 4.4, performed on the same set of data. In this way the comparison of the results seems to be reasonable without further increasing the number of replications.

The results of the experiments may be summarized as follows:

(i) The version *Hessian-solo* of the algorithm never led to convergence in more than 50% of the replications. This type of approximation turned out to be useful only in the very close neighborhood of the optimum. This obviously corresponds to a well-known property of the classical Newton-method (cp. e.g. Dennis, Moré (1977), 49). Because of these unfavorable results this type of approximation *was excluded from comparison*. Moreover, the statistics characterizing convergence (cp. 5 (vii)–(ix)) have been computed, considering only those replications where the remaining 4 versions of the algorithm converged to (nearly) the same optimum. This generally resulted in quite a few rejected replications: 15% for model 1, 2% for model 2, 0% for models 3a, 3b.

(ii) Concerning *local convergence* there can be no doubt that there is a unique ranking of the remaining 4 versions of the algorithm. Indeed, the relative performance of the different versions turned out to be remarkably stable. Both local statistics resulted in the same *ranking of the versions* for all models:

I: Hessian-mix, II: BFGS-mix, III: BFGS-solo, IV: GLS-solo.

The typical plot of the empirical distribution function of the number of iterations Mj, needed to cover the last 0.1% of the distance to the optimum (cp. 5 (viii)), is reproduced in Fig. 1. Note that the scale of the horizontal axis naturally differs for different models.

From Fig. 1 the superiority of *Hessian-mix* (I) is quite obvious. This might indicate that none of the considered model is that much nonlinear to cause a

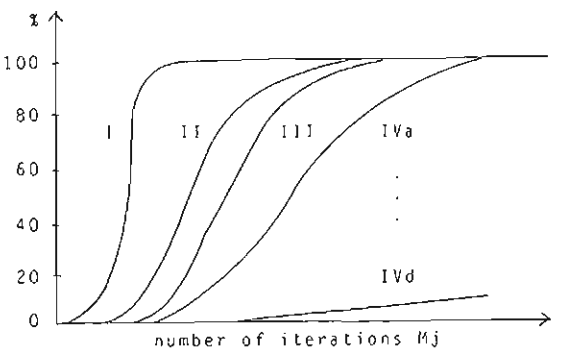

Fig. 1. Empirical distribution function of Mj

bad fit of H_H to the exact Hessian. Indeed, it is well-known that exact Newton steps (cp. (6) with H_k equal to the exact Hessian) are leading to quadratic convergence (cp. Dennis, Moré (1977), 62).

On the other hand, the *GLS-solo* version performed reasonably well in comparison with the other versions only for model 3. Really competitive results GLS-solo only produced for model 3 and 27 observations (cp. IV a in Fig. 1). Indeed, increasing the number of observations results in less iterations for all versions and in a relatively better performance of the GLS-solo version. The latter may be motivated by the consistency of GLS-type approximations for linear models (cp. 4.1). Concerning the *BFGS-versions* (II, III), the results indicate that using BFGS-updates only near the optimum profitably influence local convergence. Thus, improving the starting matrix for the updating process may well be paid off by local speed. But switching too late may offset this effect. Therefore the goodness of the switching criterion has to be discussed mainly in the context of global convergence (cp. (iii)).

These results are strongly supported by the other local statistic EL_k, the mean number of digits of the optimal value of the objective function, gained k iterations before the algorithm stopped (cp. 5(ix)). The typical plot of EL_k is reproduced in Fig. 2. In particular, Fig. 2 once more demonstrates that the performance of GLS-solo strongly depends on the model experimented with. Note that even for linear models GLS-solo may perform horribly bad (IV d represents the performance of the linear model 1).

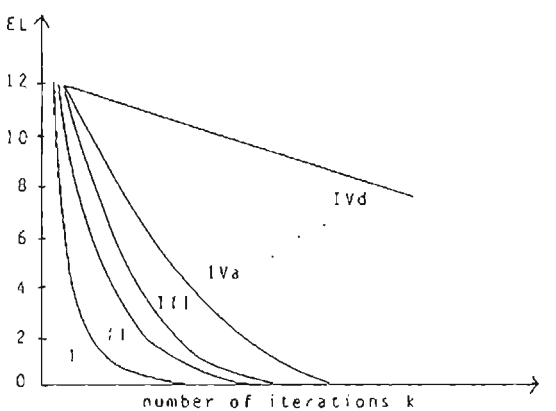

Fig 2. Mean number of correct digits, k iterations before the algorithm stopped

(iii) Stable and fast *global convergence* is crucial for econometric model builders who are most of the time interested in getting only few correct digits of the optimum in minimum time. Fortunately the ranking of the versions concerning global convergence turned out to be essentially the same as for local convergence. There are only minor differences. Note in particular that the superiority of *Hessian-mix* to *BFGS-mix* does not seem to be very dramatic in

general. Only for model 3 and 21 observations BFGS-mix performed substantially worse. This might be a result of the very few "degrees of freedom" in this case (21 observations, 18 parameters).

Concerning global convergence, both "pure" versions *BFGS-solo* and *GLS-solo* generally performed substantially slower than the mixed versions, though due to very different reasons. Whereas BFGS-solo generally needed many iterations to reach a neighborhood of the optimum, GLS-solo had strong difficulties to improve near to the optimum. The latter obviously coincides with the results on local convergence (cp. (ii)). Fig. 3 shows the typical plot of EG_k, the number of correct digits of the distance to the optimum, gained after k iterations on the average. Note in particular that for model 3 and 27 observations (cp. IVa) GLS-solo performs faster than BFGS-solo, even if one is interested in 7 correct digits.

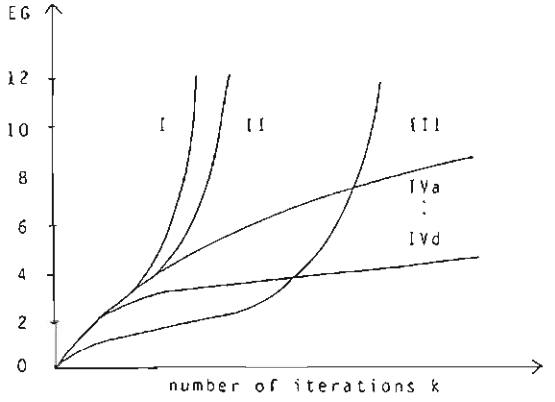

Fig. 3. Number of correct digits of the distance to the optimum, on average gained after k iterations

Moreover, Fig. 3 indicates that the criterion for switching to the locally preferable approximations (cp. 4.4 d), e)) has been posed reasonably well (cp. the increase of curves I, II, IVa, IVd).

8. Conclusion

Reconsidering all the results of the MC-experiments, the superiority of the mixed approximations (Hessian-mix and BFGS-mix) should be obvious. Using such mixtures instead of the pure approximations (Hessian-solo, BFGS-solo or GLS-solo) will surely be paid off to the model builder by obtaining estimations much faster. Indeed, the pure approximations all had more or less serious defects concerning either global or local convergence.

Furthermore, it seems that one does not have to be too unlucky not being able to compute exact Hessians, because BFGS-mix never performed that bad that it was not able to offset the greater computational effort per iteration needed by Hessian-mix.

The relevance of these results is strongly supported by analogous results for a line-search-based algorithm in Calzolari, Panattoni and Weihs (1985). Thus it may be interesting to compare the relative performance of BFGS-mix (say) in the trust-region algorithm and the line-search-based algorithm. But up to now no attempt was made in this direction, mainly because the algorithms ran under incompatible regimes.

References

Amemiya, T.: The maximum likelihood and the nonlinear three-stage least squares estimator in the general nonlinear simultaneous equation model. Econometrica *45*, 955–968 (1977)

Belsley, D. A.: On the computational competitiveness of full-information maximum-likelihood and three-stage least-squares in the estimation of nonlinear simultaneous-equation models. Journal of Econometrics *9*, 315–342 (1979)

Belsley, D. A.: On the efficient computation of the nonlinear full-information maximum-likelihood estimator. Journal of Econometrics *14*, 203–225 (1980).

Berndt, E. K., Hall, B. H., Hall, R. E., Hausman, J. A.: Estimation and inference in nonlinear structural models. Annals of Economic and Social Measurement *3*, 653–665 (1974).

Calzolari, G., Panattoni, L.: A simulation study on FIML covariance matrix. Pisa: Centro Scientifico IBM, paper presented at the European Meeting of the Econometric Society, Madrid, September 3–7 (1984).

Calzolari, G., Panattoni, L.: Gradient methods in FIML estimation of econometric models. Pisa: Centro Scientifico IBM, paper presented at the International Conference on Economic Policies and Control Theory, Venezia, January 29–February 1 (1985).

Calzolari, G., Panattoni, L., Weihs, C.: Computational efficiency of FIML estimation. Pisa: Centro Scientifico IBM, mimeo (1985)

Dagenais, M. G.: The computation of FIML-estimates as iterative generalized least-squares estimates in linear and nonlinear simultaneous equation models. Econometrica *46*, 1351–1362 (1978).

Dennis, J. E. Jr., Moré, J. J.: Quasi-Newton methods, motivation and theory. SIAM Review *19*, 46–89 (1977).

Moré, J. J.: Recent developments in algorithms and software for trust-region methods. In: Mathematical Programming, The State of the Art, Bonn 1982 (Bachem, A., Groetschel, M., Korte, B., eds.), pp. 258–287. Springer-Verlag 1983.

Parke, W. R.: An algorithm for FIML and 3SLS estimation of large nonlinear models. Econometrica *50*, 81–95 (1982).

Powell, M. J. D.: Convergence properties of a class of minimization algorithms. In: Nonlinear Programming 2 (Mangasarian, O. L., Meyer, R. R., Robinson, S. M., eds.), pp. 1–27. Academic Press 1975.

Theil, H.: Principles of Econometrics. J. Wiley 1971.

Weihs, C.: IAS-System Bonn, User's Manual. University of Bonn, Wirtschaftstheoretische Abt. I, discussion paper No. 144a (1984).

Weihs, C.: Convergence of an algorithm for FIML-estimation in (non-)linear econometric models. In: Methods of Operations Research 49 (Brucker, P., Pauly, R., eds.), pp. 101–126. Verlag A. Hain 1985.

Weihs, C.: Auswirkungen von Fehlern in den Daten auf Parameterschätzungen und Prognosen (Effects of data errors on parameter estimates and forecasts). Dissertation, University of Trier (FRG) (to appear) (1987)

C. Weihs
SFB 303, University of Bonn (FRG)
Current address:
CIBA-GEIGY AG
Mathematische Applikationen, R-1036.P.06
CH-4002 Basel
Switzerland

G. Calzolari and L. Panattoni
Centro Scientifico IBM
via S. Maria 67
I-56100 Pisa
Italy