Wojciech W. Charemza

Guesstimation

Department of Economics
University of Leicester
University Road
Leicester LE1 7RH, UK
Phone: +44 116 252 2899
e-mail: wch@le.ac.uk

February 1998 (revised September 1998)

Abstract

Macroeconomic model builders attempting to construct forecasting models frequently face constraints of data scarcity in terms of short time series of data, and also of parameter non-constancy and underspecification. Hence, a realistic alternative is often to guess rather than to estimate parameters of such models. This paper concentrates on repetitive guessing (drawing) parameters from iteratively changing distributions, with the straightforward objective function being that of minimisation of squares of ex-post prediction errors, weighted by penalty weights and subject to a learning process. The numerical Monte Carlo examples are those of a regression problem and a dynamic disequilibrium model.

Keywords: estimation, short data series, macromodels, computations, methodology

Wojciech W. Charemza*)

Guesstimation

1. Introduction

The noun 'guesstimation' has a bad press in econometric and forecasting literature. It ironically emphasises the fact that somebody, not being able to properly estimate an empirical model, is guessing its values, using own expertise and intuition. Such a procedure is generally dismissed as 'unscientific', sometimes even unethical and definitely not recommended to minors. It is, however, widely known but rarely acknowledged, that guessing parameters is a widespread procedure among macroeconomic model builders, non-econometric forecasters, policy analysts etc.. Suppose that a government policy advisor is asked about his/her projections concerning, say, future industrial output. The 'purist' way to do this is to collect long series of data on output, prices, labour and capital inputs, interest rates etc., built a model, check for integration, cointegration, heteroscedasticity, outliers, ARIMA properties and the like, carefully estimate the model, compute the prediction and forecasting intervals and then deliver the outcome to the government. It is a sad fact that, regrettably, such an ideal procedure rarely works. In practice econometricians are plagued with regime switching, unobservability of some important variables and, above all, short series of data. If two, or three quarters before the date of desired prediction a Statistical Office, which publishes data concerning industrial production, decided to redefine its index of industrial output, the only thing an honest econometrician can do is to wait for another twenty years for the time series to grow into a sufficient length (during this twenty years they will redefine it again, anyway). As far as modelling of newly established European economies is concerned, such as the Baltic States or the Balkan Republics, there is no

---

*) The financial support of A.C.E. Project Structural change and spillovers in the East European reform process is gratefully acknowledged. I am indebted to Maria Blangiewicz, Derek Deadman, Chao-Dong Huang, Krystyna Strzala and Anna Zalewska-Mitura for their helpful comments on an earlier draft of the paper. I am solely responsible for any remaining deficiences.

possibility for an econometrician wishing to built a traditional model describing the dynamics of these economies for quite a long time to come.

It is therefore no surprise that such econometrics, while confronted with the every day requirements of a policy analyst creates the desire to cut corners. 'If I cannot estimate marginal labour productivity, why should not I assume that it is equal to, say, 0.33? After all, I am a decent economist, with a lot of practice and I feel that this should be somewhere around 0.33'; this is something a lot of us are tempted to do and, perhaps, some even do. Especially that, very often, the price for being a purist is not to do the requested research at all. This seems to be confirmed indirectly, and perhaps inadvertently, by McCloskey and Zilak (1996), who gave a damming report of 'bad econometric practices' found in 182 empirical papers published in the American Economic Review. According to their findings, in most of the papers there have been substantial interpretational errors in empirical regression analyses, often performed with the use of large samples. It is difficult to believe that authors whose papers are admitted to such a prestigeous journal and their referees do not know basic econometrics. Presumably the authors squeezed what they could out of the empirical data; if they had followed 'best practice' to the letter, they would likely have finished with little or no conclusions. Then, their referees agreed, this was the best they could do under the limitations of the regression tool.

One may argue that a guess about parameters is indeed the prerequisite of any economic empirical research. Econometric estimation is just one of the methods which delivers, in certain situations and under certain conditions, an answer to the question: 'what is the best guess concerning the unknown parameters of the model?' Let an applied economist recall, and perhaps reconsider, the reason econometrics was invented: that, for the sake of conducting research we must have some knowledge about the parameters of the process we believe we are analysing. If econometrics cannot provide us with such parameters we have to do something more radical than estimation; guessing perhaps.

The problem this paper attempts to describe is the process of guessing the parameters of a complex, and generally large, empirically oriented model. It is assumed that data allow for simulation (solving) of such a model with guessed parameters, and that there is the possibility of checking the quality of the guess by computing some accuracy measure. Such an accuracy measure can be, for instance, the one-step ahead forecast error. Let us suppose that we have data necessary for making such a forecast. In such a case, it is possible to

perform guesstimation repeatedly and every time check the accuracy of forecasting results. If appropriate priors, representing the researcher's prior belief and expertise, can be applied, then the resulting objective function may have an economically sensible extremum, either global or local. It is argued that the behavioural algorithm presented in this paper represents, generally, the way the 'guesstimator' (no relation to 'estimator' here; a guesstimator is a person involved in guessing the parameters) acts in terms of formulating, applying and then correcting his/her priors.

The plan of the paper is as follows. Section 2 introduces the problem of guesstimation with the use of an example of a linear function with an infinite number of solutions. In section 3 a more general model and algorithm describing the guesstimator's behaviour in the case of repetitive guesstimation with a learning process is developed. Two artificial examples of repetitive stochastic guesstimation are given in section 4. The first is a simple one, of an ordinary least squares problem in a linear regression model and the other one is rather more complicated, describing the guesstimation of a canonical disequilibrium model with lagged unobservable dependent variables and time-varying parameters.

There are no theoretical pretences in this paper. I am quite convinced that its mathematics is well known, although I have been unable always to trace proper references. All this paper is trying to achieve is to show that where econometrics fails, the best alternative is not always to sit down, drink beer, and complain about data, the Statistical Office and the heterogeneity of the universe instead.

2. A simple example: adding two numbers together

To illustrate the problem, let us use probably the simplest and most well-rehearsed example of a 'model':

$$y_t = ax_t + bz_t + e_t \quad , \tag{1}$$

where a and b are unknown positive constants, and e_t is a loosely defined error term. Let us suppose that we have two observations for $t = 1, 2$, on x_t and z_t, each equal to one, and we have one observation on y_t, for $t = 2$ only, equal to one. Hence, the model essentially becomes:

$$y_t = \alpha + \beta + \varepsilon_t \quad .$$

Suppose further, that someone's objective is the evaluation of the parameters' values of this model according to some prior economic knowledge, acknowledging, at the same time, the fact that forecasting of y_t with the use of these parameters should be reasonably accurate. Obviously, such a 'model' cannot be estimated, due to the lack of identification. There is an infinite number of pairs of real numbers from the interval (0,1) which minimises square of prediction error of y_t to zero (e.g. 0.9 and 0.1; 0.8 and 0.2 ...). In other words, there is no unique minimum of the objective function, if the objective function is defined simply as the square of forecast error. This trivial example represents the essence of problems econometricians face with undersized (short) samples.

A guesstimator seems to encounter a similar problem. Whatever guess he/she makes concerning one of the parameters, there is always another one which can be set in such a way that the squared forecast error will be zero. This is only the case if the guesstimator is completely ignorant concerning the investigated economy. In practice, he/she nearly always has some prior knowledge about the 'true', empirical or theoretical, model. For instance, he/she can strongly reject the combinations $\alpha = 1$ and $\beta = 0$ (or vice versa) as economically nonsensical. He/she might be also inclined to dismiss the combination $\alpha = 0.95$ and $\beta = 0.05$ albeit, perhaps, less strongly. It is also possible that the guesstimator might prefer, on economic grounds, some combinations which lie outside the constraint given by non-identification. To illustrate this, let us suppose for a while that (1) is a production function, (x_t being a logarithm of labour input and z_t a logarithm of capital input), $y_2 = 1.1$ rather than $y_2 = 1$ and the guesstimator believes that there should be constant returns to scale, that is, that $\alpha + \beta = 1$. On economic grounds, he/she would accept that, for instance, $\alpha = 0.5$ and $\beta = 0.5$ rather than $\alpha = 0.55$ and $\beta = 0.55$, even if the latter combination forecasts better. Finally, we might assume that the restriction is not known to the guesstimator. This looks strange in the simple example given above, but it can be more plausible if we consider a complex, dynamic, multiequation model with a more complicated objective function instead.

Returning to the original example, where $y_2 = 1$, let us suppose that his/her prior beliefs concerning a is that it is equal to 0.5 and concerning b, that it is equal to 0.7. Every other value is also admissible, but subject to a penalty weight. Let us denote these values as $\hat{\alpha}$ and

$\hat{\beta}$ respectively and call them the guesses. It seems to be reasonable to assume that the penalties deviate from the initially believed values of a and b, according to, say, mutually independent standard normal distributions. Since the parameters might vary in their magnitude, it would be of advantage to scale them by their means. The lowest penalty (zero) is associated with the situation where the guesses are equal to the prior beliefs ($\hat{\alpha} = 0.5$ and $\hat{\beta} = 0.7$). Every other guess carries a non-zero penalty distributed as standard normal with the arguments $[(\hat{\alpha}/0.5) - 1]$ and $[(\hat{\beta}/0.7) - 1]$. Hence, the weight $\omega(\hat{\alpha}, \hat{\beta})$, is defined as the arithmetic averages of those two variates, is given by:

$$\omega(\hat{\alpha}, \hat{\beta}) = \frac{n\left(\frac{\hat{\alpha} - 0.5}{0.5}\right) + n\left(\frac{\hat{\beta} - 0.7}{0.7}\right)}{2 \cdot n(0)} \quad ,$$

where $n(\bullet)$ denotes the value of a standard normal probability density function. The guesstimator is interested in minimizing the weighted criterion function:

$$\varphi = [y_2 - \omega(\hat{\alpha}, \hat{\beta}) \cdot (\hat{\alpha} + \hat{\beta})]^2 \quad ,$$

and, at the same time, in minimizing the unweighted criterion function:

$$\tilde{\varphi} = [y_2 - (\hat{\alpha} + \hat{\beta})]^2 \quad .$$

After all, the guesstimator is still interested in obtaining such parameters which would give him/her the best possible forecast accuracy. It is easy to check that minimisation of $\tilde{\varphi}$ leads to the the result $\tilde{\varphi} = 0$ for any combination of $\hat{\alpha}$ and $\hat{\beta}$ which satisfies the restriciton $\hat{\alpha} + \hat{\beta} = 1$. For $\tilde{\varphi} = 0$ a unique minimum of φ, equal to 0.00018, can be obtained for $\hat{\alpha} = 0.433$ and $\hat{\beta} = 0.567$.

But what has this in common with the guesstimation procedure? This procedure illustrates a simple guessing rule, where the guesstimator is willing to modify his/her initial guess by accepting another one, but only if this would lead to a decrease in the criterion functions. Or, in other words, if one is guessing the parameters values at random and then computes the squared prediction error weighted by the penalty weight, it is likely that the final result will be close to 0.433 for a and to 0.567 for b. In fact a straightforward

simulation experiment of drawing 10,000 uniform random numbers from the interval (0,1), treating them as $\hat{\alpha}$'s, with $\hat{\beta}$'s computed from the restriction, and choosing such a pair of these numbers which minimises the criterion function gave values of 0.4329 and 0.5671 respectively. If the guesstimator is completely indifferent concerning the prior value of the parameters he/she attempt to guess and the penalty weight is uniformly equal to unity, then there is an infinite number of solutions minimising the criterion function. But, in a such case, this person should not perhaps be taken seriously as an expert in prediction, since he/she does not have any valuable knowledge on the subject.

3. Guesstimators' model and algorithm

Evidently, in practice, an experienced guesstimator applies, more or less consciously, a more complicated process of selection than that described above. It is likely that there are different degrees of uncertainty concerning particular parameters, Parameters might be regarded as being more or less difficult to guess. For instance, the possible interval in which the capital depreciation ratio is placed might be thought of as being narrower than, say, the short-run price elasticity of imports. Also, in some cases the guesstimator might revise the prior beliefs; if he/she realises that guesses widely apart from the expected value of the prior are giving sensible results, in terms of minimising the criterion function, it may happen that the person making guesses gets wiser, learns and modifies the priors. These priors can be modified in two ways. There might be something like 'learning eagerness', or 'learning aversion' where the researcher is increasingly eager (or reluctant) to modify the weights used for evaluating guesses against prior beliefs. It seems to be reasonable to assume that, with the increase in number of corrections of the priors, the researcher will increase his/her confidence in guesses from the priors. This would result in an increase in penalty weights with the increase in number of modifications of the priors. At the same time, it is reasonable to assume that, with the increase in the number of such revisions, the guesstimator would also express his/her increasing confidence by narrowing the interval from which the parameters are to be guessed, accordingly. It is equally possible to imagine a learning-averted guesstimator, who would decrease the weights put on the guesses and enlarging, rather than narrowing, the interval from which the guesses are made.

It is possible that the mean of the interval from which the parameters are drawn may also change in the process. Initially the guesstimator may draw a set of parameters from the interval he/she believes, at that stage, is the most likely to include the 'best' values of parameters and check the criterion function. Then, if an improvement has been made, he/she might revise the priors (that is, move the mean of the prior distribution to the point for which the improvement took place) and draw again. If there was no improvement, the drawing continues using the initial values of parameters (or values for which the criterion function previously reached its desired extremum) until the large number of unsuccessful drawings confirms that there is no room for improvement.

With these points in mind, the following guesstimators' model is proposed:

$$y_t = f(y_t, x_t, \varepsilon_t; \theta) \quad , \qquad (2)$$

where y_t , t = 1, 2,..., n, is the vector of current, observed, endogenous variables, x_t contains all other relevant and observable variables (at least weakly exogenous) and lagged endogenous variables and θ is the vector of K parameters which are to be guessed. Unlike as in a traditional econometric model, there is no identification restrictions and, in particular, the number of observations can be smaller than the number of unknown parameters. If (2) is a static model where all x_t variables are strongly exogenous, the minimal number of observations is one. If there are endogenous variables lagged by one included in x_t, then two observations are needed, etc.. Generally, the parameters θ are allowed to vary in time but, if the criterion is the minimisation of one (or more) step ahead forecast errors, they are supposed to be invariant intertemporarily (this creates a time consistency problem, not discussed in this paper). Finally, e_t is the random and unpredictable (in mean) process.

The model is completed by the guesstimator's prior (initial) beliefs concerning the parameters (the priors). The prior beliefs (also called the priors herein) are defined as a vector of K intervals, Θ^0, which are proportional to the intervals the guesstimator initially assumes the parameters are included in. These intervals are in turn defined by their mean values, $\theta^{(0)}$, and length, $\ddot{\Theta}^0$ The mean is essentially the guesstimator's prior best guess and the length of each interval corresponds to the uncertainty the guesstimator attributes to his/her knowledge concerning each parameter. Also, the model may include a number of maximal 'drift changes', that is the maximum number of times the guesstimator is prepared

to revise the mean of his/her priors. This is denoted below as N^d. Before the first step of the algorithm, it is necessary to derive the initial values of the criterion function. This can be done by using the initial values of the parameters, $\theta^{(0)}$, equal to the mean of initial intervals for solving the model (2) for y_t given $\varepsilon_t = 0$. Since this solution depends on $\theta^{(0)}$, let us denote it as: $y_t^{(0)} = \bar{f}^{-1}(x_t ; \theta^{(0)})$. This solution is needed in order to make an h-step ahead forecast for y_t, that is, finding $\hat{y}_{t+h}^{(0)} = \bar{f}^{-1}(\hat{x}_{t+h}^{(0)} ; \theta^{(0)})$, $h = 1, 2, \cdots$, where $\hat{x}_{t+h}^{(0)}$ is a forecast for x_{t+h} (the vector x_t may contain lagged y's and weakly exogenous variables; in the simplest case, where x_t is a vector of strongly exogenous variables, their future values must be known and $\hat{x}_{t+h}^{(0)} = x_{t+h}$). Compare the predictions $\hat{y}_{t+h}^{(0)}$ with the observed realisations of y_{t+h} by computing an initial value of the unweighted criterion function (UCF):

$$\tilde{\varphi}^{(0)} = \tilde{\varphi}^{(0)}(y_h, \hat{y}_h^{(0)}) \quad ,$$

where $y_h = \{y_{t+1}, y_{t+2}, \cdots\}$, $\hat{y}_h^{(0)} = \{\hat{y}_{t+1}^{(0)}, \hat{y}_{t+2}^{(0)}, \cdots\}$. A simple example of such an UCF is the sum of squares of one-step ahead prediction errors:

$$\tilde{\varphi}^{(0)} = \sum_{t=1}^{T-1} \sum_{y_{t+1}} (y_{t+1} - \hat{y}_{t+1}^{(0)})^2 \quad ,$$

where the symbol $\sum_{y_{t+1}} (\bullet)$ means the summation of all elements of vector y_{t+1} (that is, for all endogenous variables of the model). With these initial values, the algorithm of repetitive guessings (called herein the Repetitive Stochastic Guesstimation, the RSG) is the following:

1) In every iteration j (where 'iteration' relates to achieving an improvement in the criterion function) the previously obtained (or initial) set of admissible parameters intervals is modified through an application of the learning function $l_\Theta(j)$:

$$\Theta^{(j)} = \theta^{(j-1)} \pm \tfrac{1}{2} \ddot{\Theta}^{(j-1)} \lambda_\Theta(j) \quad .$$

2) From the set $\Theta^{(j)}$ draw (that is, guess or randomly generate) a sample of K parameters, $\theta_i^{(j)}$ and for these parameters compute model solutions:

$$y_{i,t}^{(j)} = \bar{f}_t^{-1}(x_t ; \theta_i^{(j)}) \quad ,$$

forecasts $\hat{y}_{i,h}^{(j)}$ (analogously to $\hat{y}_h^{(0)}$) and unweighted and weighted criterion functions, defined respectively as:

$$\tilde{\varphi}_i^{(j)} = \tilde{\varphi}_i^{(j)}(y_h, \hat{y}_{i,h}^{(j)}) \quad ,$$

and:

$$\varphi_i^{(j)} = \varphi_i^{(j)}[y_h, \hat{y}_{i,h}^{(j)}, \omega(\theta_i^{(j)}, \lambda_\varphi(j))] \quad ,$$

where $\hat{y}_{i,h}^{(j)} = \{\hat{y}_{i,t+1}^{(j)}, \hat{y}_{i,t+2}^{(j)}, \cdots\}$, and $\lambda_\varphi(j)$ is the learning function analogous to $l_\Theta(j)$, and is an argument of the penalty weight function $\omega(\bullet)$. By analogy to the initial UCF, the exemplary weighted criterion function, WCF, can be defined as one-step ahead prediction error:

$$\varphi_i^{(j)} = \sum_{t=1}^{T-1} \sum_{y_{t+1}} \left(y_{t+1} - \omega[\theta_i^{(j)} \lambda_\varphi(j)] \hat{y}_{i,t+1}^{(j)} \right)^2 \quad , \tag{3}$$

and $\tilde{\varphi}_i^{(j)}$ is analogous to $\tilde{\varphi}_i^{(0)}$, that is:

$$\tilde{\varphi}_i^{(j)} = \sum_{t=1}^{T-1} \sum_{y_{t+1}} (y_{t+1} - y_{i,t+1}^{(j)})^2 \quad . \tag{4}$$

For linear models with negative degrees of freedom (that is, not identified), the limit value for such defined $\tilde{\varphi}_i^{(j)}$ is obviously zero. Furtheron the random drawing of parameters within an iteration, identified by subscript i, is reffered to as replication.

3) In each replication the value of the function $\varphi_i^{(j)}$ is compared with that obtained in the previous iteration $\varphi^{(j-1)}$ and the value of the function $\tilde{\varphi}_i^{(j)}$ is compared with $\varphi^{(j-1)}$. It is often convenient to use $\varphi^{(0)} = \tilde{\varphi}^{(0)}$ as the initial value. Suppose that we are interested in minimisation of the criterion functions. If $\varphi_i^{(j)} < \varphi^{(j-1)}$ and $\tilde{\varphi}_i^{(j)} \leq \tilde{\varphi}^{(j-1)}$, then the algorithm moves to next iteration ($j = j + 1$) and steps 1) –3) are repeated starting from $i = 1$. While repeating step 1) the priors can be modified in two ways: i) by setting $\lambda_\Theta(j) \neq \lambda_\Theta(j-1)$ and, additionally ii) by imposing $q^{(j)} = q_i^{(j)}$. If only modification i) is imposed and $\theta^{(j)} = \theta^{(0)}$ for all j, the algorithm is called the constant mean RSG; otherwise we are dealing with the non-constant mean RSG. Also, the weights in the weighted criterion may change, if $\lambda_\varphi(j) \neq \lambda_\varphi(j-1)$. If $\varphi_i^{(j)} > \varphi^{(j-1)}$, or if $\varphi_i^{(j)} < \varphi^{(j-1)}$

but $\tilde{\varphi}_i^{(j)} > \tilde{\varphi}^{(j-1)}$, then the admissible intervals do not change and steps 2) – 3) are repeated for unchanged j and i = i + 1; the algorithm moves to the next replication within the same iteration. A new set of potential parameters is drawn from the same intervals as before and this is repeated until there is an improvement on the objective function, or the stopping rule is fulfilled.

An important question is how to define the learning functions. A simple proposition for the learning function is:

$$l_\theta(j) = \bar{l}_\theta \sqrt{1 + d_\theta \frac{j-1}{N^1}} \quad ,$$

where $\bar{l}_\theta$ is a constant which reflects the impact of the learning process on the penalty weights, N^1 is a constant which is inversely proportional to the learning speed and d_θ is a constant positive or negative value, depending whether the guesstimator expresses 'learning aversion' or 'learning eagerness'. The other learning function, $\lambda_\varphi(j)$, is defined analogously.

Given $\lambda_\varphi(j)$, the penalty weights for the criterion function are said to be normally distributed according to the difference between the actually guessed and the previous best guess (that is, the middle of the admissible interval):

$$\omega(\theta_i^{(j)} \lambda_\varphi(j)) = \frac{\sum_{k=1}^{K} n\left(\lambda_\varphi(j) \cdot \psi_k \cdot \frac{\theta_{k,i}^{(j)} - \theta_k^{(j)}}{\theta_k^{(j)}}\right)}{K \cdot n(0)} \quad ,$$

where $\theta_{k,i}^{(j)}$ denotes the k-th parameter drawn in the j-th iteration and, within it, in the i-th drawing (replication) and $q_k^{(j)}$ is such a value of the k-th parameter which ended the j-1 iteration (in another words, $q_k^{(j)}$ is equal to the k-th element of $q^{(j-1)}$) and ψ_k is the scaling factor for the k-th parameter. If the guesstimator is putting equal 'faith' into a guess of any parameter, regardless of its scale than $\psi_k = 1/K$ for all k. Otherwise it varies and it might be reasonably assumed that values of ψ_k are inversely proportional to the size (absolute value) of a corresponding k-th parameter. There are obviously numerous other ways for the penalty weights and learning function to be formulated, but those given above are simple to compute and intuitively appealing. For instance, the penalty weight function has a maximum

equal to unity for $\theta_{k,i}^{(j)} = \theta_{k}^{(j)}$, that is where the guessed parameter is equal to the best one (so far).

Technically, there is nothing new in the above algorithm. The question of finding a solution in an undersized optimisation problem has been discussed for a long time in the literature of stochastic optimal control of an econometric model, (see e.g. Hughes Hallet and Rees (1983) for a thorough description of early economic applications, Arkin and Evstigneev (1987), Holly and Hughes Hallett (1989) for a more sophisticated approach and Cividini (1992) for a comparison with alternative methods). The algorithm is also similar to that of 'training' the weights in the simple two-layer stochastic neural network, with the squashing function given by the UCF and WCF (see e.g. Hornik et al. (1990); for a simple introduction see Fausett (1994, p. 329); for its analogy to the multivariate least squares and the generalised least squares method see e.g. Angus (1989); for a general overview see Barndorff-Nielsen, Jensen and Kendall (1993)). The principal difference with the theory of neural network training is in the fact that in the guesstimation algorithm, the outputs are not normalised within the interval (0,1) – see Wasserman (1989, pp. 45). Conceptually, the guesstimation seems to be close to 'calibration' of parameters of general equilibrium models as suggested by Kydland and Prescott (1982), (1991); see also Kydland (1992) and for details of the computational algorithm and description of software, Greenaway et al. (1993). The reader must be on the alert here, since the concept of 'calibration' used in the literature has various, sometimes quite confusing, meanings. The philosophical underlings of both the Kydland and Prescott 'calibrator' and the guesstimator, are indeed very close: both are inventing their parameters up to the best of their prior knowledge, both are verifying a criterion function and then attempting to revise the priors, if the result is not up to their liking. The principal difference is that the guesstimator's action does not depend so heavily on microeconomic assumptions and constraints. Instead, he/she repeats the process of selection of the parameters more often and is prepared to learn, that is to modify his/her prior (that is, also microeconomic) knowledge. If one does not like the notion of 'guesstimation', one might use that of 'repetitive stochastic calibration' instead. It seems, nevertheless, that despite the fact that prior beliefs concerning the parameters are widely used, the guesstimation is not a Bayesian analysis, at least not in the traditional sense. The posterior distribution is not computed, neither directly nor indirectly. In particular, Bayesian analysis requires a full-sized sample (the degrees of freedom constraint). This is not required

for guesstimation. In its extreme case, with only two pieces of information available, the guesstimation is (nearly) a fully subjective enterprise; one might note that in the example given in Section 2, the final result would always change with the revision of the priors.

An interesting attempt of combining the RSG with the genetic evaluation strategies has been made by Plata-Przechlewski (1997). He analysed a number of different genetic algorithms, with the results pointing out at the usefulness of the defining of the drawing process of $\theta_i^{(j)}$ as:

$$\theta_i^{(j)} = \theta^{(j-1)} + N(0, \sigma_\theta^{(j)}) \quad ,$$

where is $N(0, \sigma_\theta^{(j)})$ is a random variate generated from a normal distribution with a zero mean and standard deviation $\sigma_\theta^{(j)}$, evaluated according to the 1/5 success rule (see Rachtenberg (1973)). It has been shown that, in some instances, drawing of parameters according to the rules of the genetic evaluation strategies gives reults superiour to that of the original RSG. Detailed analysis of the genetic algorithms is, however, beyond the scope of this paper.

Particular variations of the general RSG algorithm can be illustrated by surface plots representing the guesstimates of the parameters α and β of model (1) together with the corresponding values of the WCF. Strictly speaking, the surface plots correspond to about 10,000 three-dimensional points: $\alpha_i^{(j)}$, $\beta_i^{(j)}$ and $\varphi_i^{(j)}$. Figures 1-4 show such surfaces for the constant and non-constant RSG's, with and without weighted scaling. The surfaces for the constant mean RSG (Figures 1 and 2) are bimodal, with the minimum of the criterion function being in the 'valley' between the hills. These hills are clearly visible for the case with uniform scaling of priors; where weighted scaling is used, the lower (forward) hill can hardly be noticed. The plots suggest that the non-weighted algorithm 'wasted' a lot of replications, at the early stages of computations, searching for the minimum well away from its actual point. The algorithm with the weighted scaling of priors, with a steeper slide towards the minimum, can be regarded, in the analysed model, as computationally more efficient. Even more efficient seem to be the algorithms with non-constant means (Figures 3 and 4). It has to be stressed, however, that the entire reason for using non-constant rather than constant

mean algorithms is of subjective nature; whether or not the guesstimator is prepared to revise his/her beliefs regarding the mean of the priors has to be decided prior to guesstimation, on the grounds of some external information. A mistake in this respect can have rather dire consequences since, as in the example (1) above, every revision of the mean of the prior leads to a different minimum of the WCF.

Surface plots of drawings in Model 1:

Fig. 1: constant-mean RSG, uniform scaling of priors

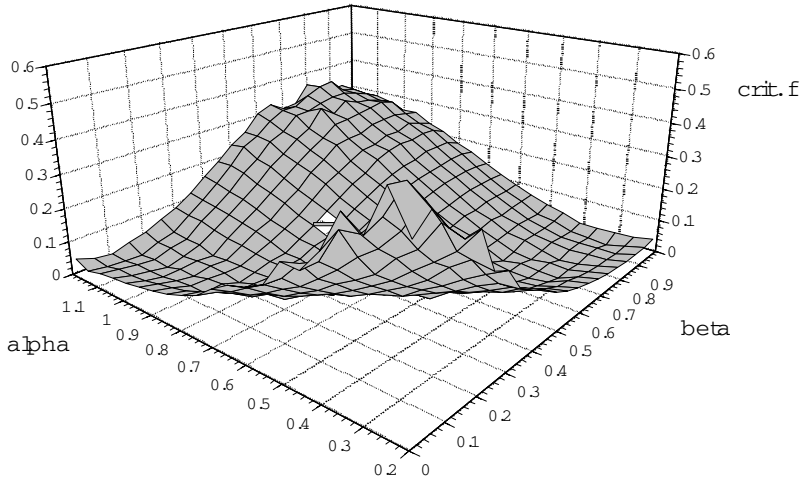


Fig. 2: Constant-mean RSG, weighted scaling of priors

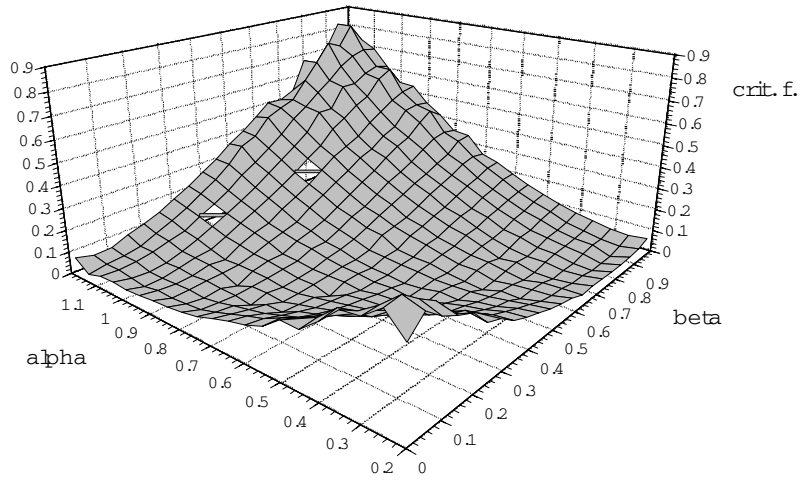


Fig.3: Non-constant-mean RSG, no scaling of priors

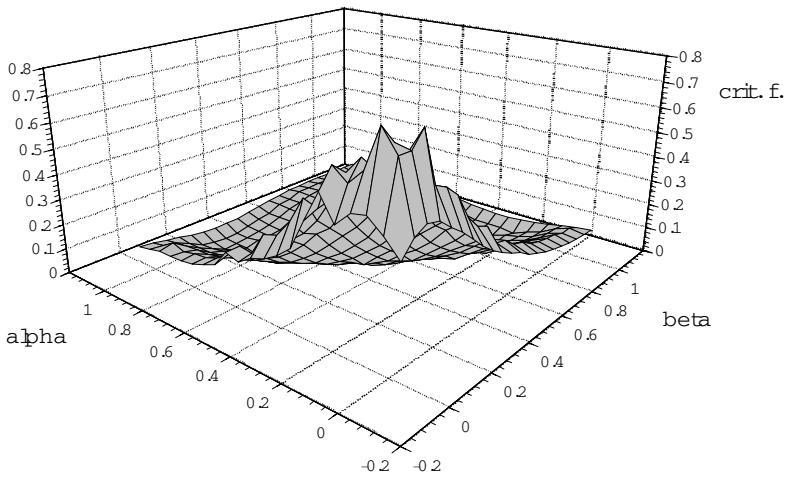
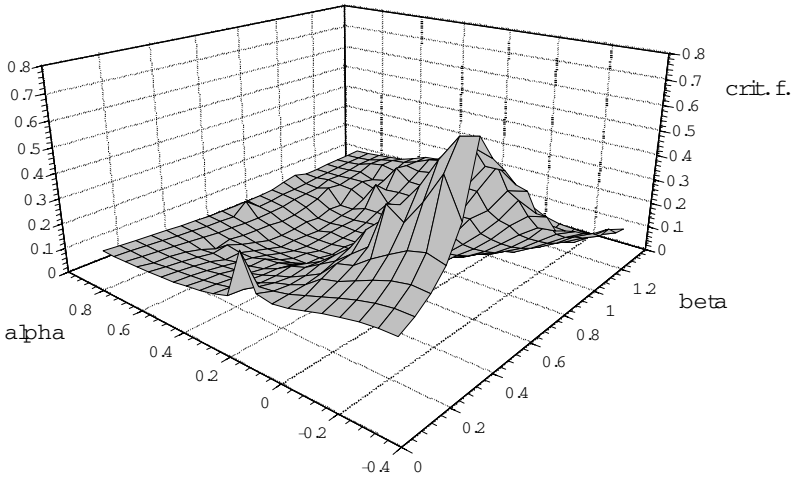


Fig.4: Non-constant-mean RSG, weighted scaling of priors



4. Some Monte Carlo examples

4.1. Linear single equation model

An evident question asked by an empirical analyst is: is there any empirical, or pseudo-empirical parameters' evaluation procedure; estimation, guesstimation, calibration, etc. which may allow for any improvement over the parameters values the investigator initially believes in? If answer to this question is positive, then it can be argued that the procedure is, in some sense, efficient, since it leads to better (more accurate) evaluation of parameters, than simple guesses. Obviously, if more than one method is compared, then this one can be

regarded as better (more efficient), which either produces such an improvement more frequently, and/or approximates the true value of the parameter with better accuracy. The problem simplifies in the case where the number of degrees of freedom is negative that is, where the number of observations is smaller than the number of parameters. In such a case, 'proper' econometric methods cannot be applied, and the investigator is left with two options: to believe in his/her initial guess, or to apply the RSG. Therefore, it can be asserted that the RSG, makes sense (is, in some sense, superior to the initial guess) if, on average, it does give an improvement in the accuracy of approximation of the true parameter, relatively to the initial guess, more often than does not. If, for instance, the true value of the parameter is one, the initial guess is 0.5, and the RSG produces the number of 1.2 then, in this particular case, the RSG scores a point against the initial guess, since 1.2 is closer to one than 0.5. If such a situation happens more often than the opposite, then one might regard the RSG as being efficient relatively to the initial guess.

In order to evaluate the efficiency of the RSG in relation to initial guesses and, in the case of positive degrees of freedom, in relation to some alternative methods, Monte Carlo experiments have been performed on data generated by the following data generating process (DGP):

$$y_t = \sum_{k=1}^{K} \alpha_k x_{k,t} + \varepsilon_t \quad ; \quad t = 1,2,\cdots 10 \quad ,$$

where ε_t is generated from standard normal distribution and values of $x_{k,t}$ are fixed in repetitive samples. In individual experiments the number of explanatory variables, K, is altered from 1 to 20. Since the sample size remains unchanged at the level of 10, the number of degrees of freedom changes, in individual experiments, from 9 ($K = 1$) to -10 ($K = 20$). The parameters values α_k are fixed (that is, drawn once from a uniform [1,10] distribution).

It is also assumed that, in each case, the investigator possesses prior knowledge regarding the parameters. This knowledge is, however, imperfect and he/she errs regularly by a given ratio. Hence, it is assumed that the initial values of the investigator differs, randomly, from the true parameters values by 0.25% 50% and 75% respectively. These all, for the RSG, gives the total number of experiments as equal to 60 (20 DGP's with 1 to 20 parameters times 3 different initial guess error ratios). The number of parameter evaluations (Monte Carlo replications) for each experiment is 500. In each evaluation, the constant-mean

RSG has been applied, with maximum number of RSG iterations (that is, changes in learning functions and priors intervals length) equal to 150 and the maximum number of replications within each iteration equal to 500.[1]

Where the number of degrees of freedom is positive (that is, for k = 1, 2, ..., 9) it is possible to compute, for the sake of a comparison, an econometric alternative to the RSG. Two such alternatives have been used: the ordinary least squares method (OLS) and a simple Bayesian estimator, with the prior distribution for parameters given by the multivariate normal distribution with diagonal covariance matrix (see e.g. Judge et al. (1988), pp. 284-287). While the OLS is an example of the method where no prior knowledge of the investigator is used (initial values of the parameters are ignored), the Bayesian estimates can be seen as a logical alternative to the RSG, where the investigator's prior knowledge is explicitly applied. Consistently with the assumptions used for the RSG, for Bayesian estimation it has been assumed that means of the prior distributions for the parameters are equal to those initially guessed by the investigator, that is they differ by, respectively, 25%, 50% and 75% form the true parameters' values. For computational simplicity is also assumed that the standard deviation of the error term is known and equal to unity.

Figures 5-7 show the proportions of average improvements, across K evaluated parameters, over the initial values given by the particular methods applied. For the case of positive number of the degrees of freedom, where both the OLS and Bayesian methods are applied, the results suggest superiority of the RSG over these two methods in the case where the initial guesses are relatively close to the true values of parameters. With the increase of distance of the initial guesses from the true values (that is, where the guesses are becoming worse), the OLS, which does not require any priors at all, gains relatively to the RSG. The Bayesian method is losing dramatically with the decrease in the accuracy of the initial guess. In fact, the entire experiment can be regarded as being set unfairly against the Bayesian method, since it is claimed that the Bayesian estimation of a linear model is efficient if the true parameters is equal to the expected value of the prior distribution.

---

[1] All computer programs and data used in this paper are written in G A U S S and are available on request.

Where the number of the degrees of freedom is negative and the only alternative to the RSG is the initial guess, the RSG shows its efficiency by producing the proportion of cases where there has been an improvement over the initial guess consistently at a level exceeding 0.5. It should also be observed that this proportion rises with the increase of the initial guess inaccuracy. At the same time, as shown by Figure 8, the average (across the parameters) root mean square errors (RMSE's) of the parameters' evaluations obtained with the use of the RSG, have a tendency to decrease with the increase of the accuracy of the initial guesses. For larger initial guess errors, there is also visible a slight tendency of the average RMSE to decrease with the decrease in the degrees of freedom. This, apparently absurd, situation, can be intuitively explained by the fact that, where the number Monte Carlo analysis of the linear model:Proportion of average improvements over initial guesses of degrees of freedom is decreasing, the amount of prior information is increasing with the increase in number of parameters.

Fig.5: Initial guess error ratio: 25%

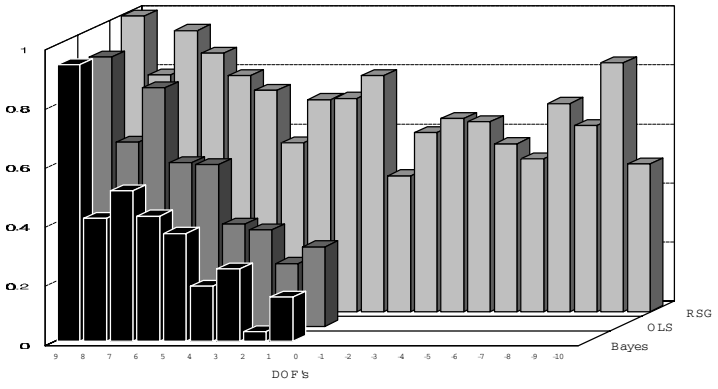


Fig.6: Initial guess error ratio: 50%

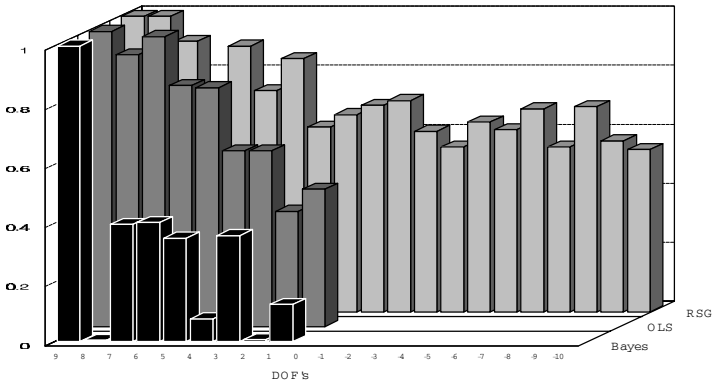


Fig.7: Initial guess error ratio: 75%

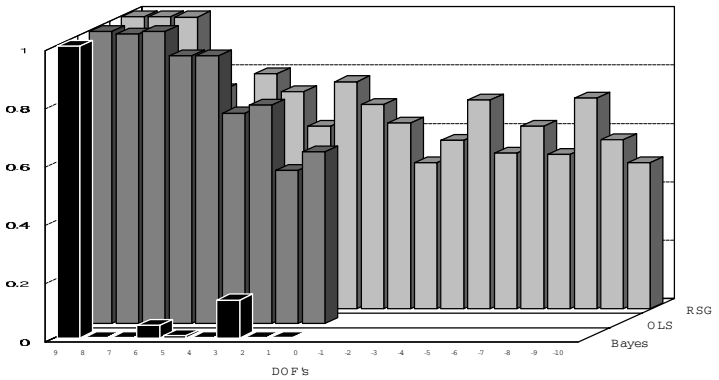
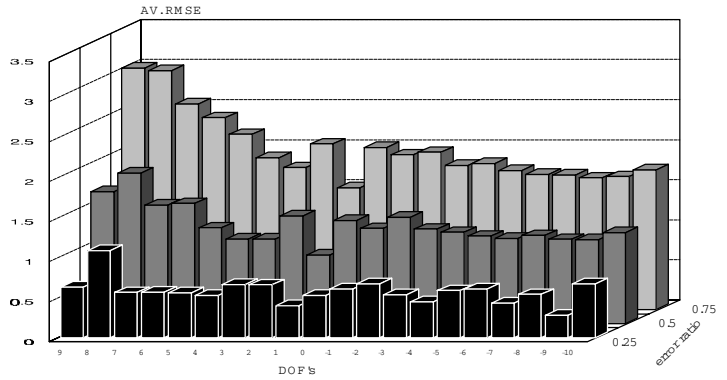


Fig.8: Average RMSE for the RSG in the linear model



4.2 A non-linear model: dynamic canonical disequilibrium model

The next model to be considered is also an artificial one, although of a more complicated, nonlinear structure. Suppose that there is a market in disequilibrium described by the following dynamic model:

$$D_t = \theta_1 D_{t-1} + \theta_2 x_{1,t} + \varepsilon_{1t} \quad ,$$

$$S_t = \theta_3 x_{2,t} + \varepsilon_{2t} \quad ,$$

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

Here D_t denotes demand, S_t is supply and Q_t is the quantity transacted. It is assumed that demand and supply are not directly observable, in that they are not equal to the quantity transacted. Such a model is called a dynamic canonical model, and is regarded as being particularly difficult to estimate, due to the presence of the lagged unobservable variable D_{t-1} (see e.g. Quandt (1988), pp. 132-140). The method recently proposed for estimation of such a model is the simulated pseudo-maximum likelihood (SPML) method by Laroque and

Salanié (1995); for further development see Lee (1997a,b). The method, which allows for a very general specification of the model, consists in simulating first and second order moments of the endogenous variables in h independent drawings and then averaging the results.

The canonical disequilibrium model can easily be evaluated with the use of the RSG. If the criterion functions are given by (3) and (4), that is set to minimise the one-step ahead forecast errors, the one-step ahead prediction can be computed as:

$$Q_{i,t+1}^{(j)} = \min(\theta_{1,i}^{(j)} D_{i,t}^{j} + \theta_{2,i}^{(j)} x_{1,t+1}, \theta_{3,i}^{(j)} x_{2,t+1})$$

with $D_{i,t}^{(j)}$ computed recursively in each replication and $\theta_{1,i}^{(j)}$, $\theta_{2,i}^{(j)}$, $\theta_{3,i}^{(j)}$ are drawings of parameters obtained in particular iteration and replication.

In order to compare the performance of the SPML and RSG, a series of Monte Carlo experiments were performed. The DGP is essentially that used by Laroque and Salanié (1995), that is where $\theta_1 = 0.5$, $\theta_2 = \theta_3 = 1$, standard errors of ε_{1t} and ε_{2t} are equal to unity, $x_{2,t}$ is a unitary variable with all its values equal to 5, $x_{1,t}$ is defined as $x_{1,t} = (1-\theta_1)\cdot[5\cdot n_t(0,1)]$, where $n_t(0,1)$ stands for a simulated pseudo-random standard normal variate, and the initial value for D_{t-1} is given as:[2]

$$D_0 = 5\cdot\theta_2 + \frac{n_0(0,1)\cdot\sqrt{1+5^2\theta_2^2(1-\theta_2^2)}}{\sqrt{1-\theta_1^2}}$$

As before, it is assumed that the investigator is making errors in his/her initial guesses regarding the parameters, respectively by 25%, 50% and 75% of their true values. These initial guesses are used as means of initial drawing intervals in the RSG and as the starting values in the optimisation routine of the SPML. The initial standard deviations for parameters (in case of the RSG, for the initial intervals for priors) have been assumed, for both methods, as being equal to their true values. For each method, one hundred replications were made for sample sizes of N = 3 (RSG only), 10 and 100, with the first observation discarded for lags. The number of degrees of freedom is equal to -1, 6 and 96 respectively. The RSG applied was the constant-mean algorithm, with the maximum number of learning

---

[2]) I wish to express my thanks to Guy Laroque and Bernard Salani. for allowing me to use their G A U S S S P M L program

function changes (iterations) equal to 150 and the maximum number of replications within each iteration equal to 3,000. For comparison, for sample size equal to 100, the results of the non-constant mean algorithm are also shown (for sample size equal to 10 the results obtained for the non-constant mean RSG are clearly inferior and, for sample size of 2, nonsensical). The SPML method has a limit of 150 optimisation iterations and the number of SPML drawings h is equal to 20. The idea of these settings was to make the computing time used by both methods in one Monte Carlo replication as being of a similar magnitude. In practice, however, the SPML method turned out to be about four times more expensive (in terms of computing time) than the RSG, for the same sample size. There has been also some cheating imposed in computations in favour of the SPML. It was found out that, for a sample size of 10 (and, in one case for a sample size of 100), the SMPL diverge, in iterations, from the true parameters, leading the estimated parameters astray. Hence, where such a diversion was noticed, the particular Monte Carlo replication was repeated.

Table 1 summarises the results. For a sample size of 100 the SPML is clearly superior to the constant-mean RSG, both in terms of average RMSE and average frequencies of improvements over initial guesses. The RSG, however, increases its efficiency faster than the SPML with an increase in the initial guess accuracy. However, the non-constant mean RSG results seem to be nearly as accurate (and, for the initial error ratio of 0.25, more accurate) as the SPML computations. For a sample size of 10, the RSG performs better for 'close' initial guesses and similar to the SPML for 'average' and 'bad' initial guesses. It is worth noting, however, that in about 10% cases the SPML gave totally unreliable (divergent) results, discarded for calculation of the RMSE and average frequencies of improvements and that it is also about 4 times computationally less expensive. For a sample size of 3, the Monte Carlo characteristics worsened only slightly in comparison with those obtained for N = 10.

Table 1: Monte Carlo comparison of the SPML and RSG methods

		Initial error ratio: 0.25		Initial error ratio: 0.50		Initial error ratio: 0.75	
	N	SPML	RSG	SPML	RSG	SPML	RSG
Av. RMSE	3	N/A	0.216	N/A	0.337	N/A	0.509
	10	0.257	0.198	0267	0.274	0.397	0.489
	100	0.068	0.202 0.071*	0.067	0.268 0.076*	0.101	0.489 0.071*
Av. Impr.	2	N/A	0.620	N/A	0.640	N/A	0.643
	10	0.627	0.710	0.877	0.677	0.663	0.689
	100	0.977	0.757 0.980*	1.000	0.717 0.997*	0.997	0.700 1.000*
No. of diversions	10	6	0	12	0	10	0
	100	0	0	1	0	0	0

Results marked by * are for the non-constant mean RSG algorithm. All other RSG results are for the constant-mean algorithm.

6. Summary and conclusions

It seems that guesstimation, in the sense discussed above may, in some situations, be not a totally idiotic idea. If a guesstimator has either the patience or a lot of time, or an access to a decent computer so that he/she can repeat the process and learn on the way, the result might be of some practical importance. In particular, he/she might improve on the initial guess, might discover parameter values which have some interesting forecasting properties, and may not bother with identification, short samples and time-varying parameters. Often and sometimes unknowingly, he/she might solve an optimal control problem on the way, do a 'calibration' or train a neural network. What is equally important, models on which the guesstimation is performed might be highly nonlinear, truncated and even formulated in a 'fuzzy' fashion. In other words, they might be closer to the underlying economic theory than typical econometric models which, for estimation purposes, are often nearly linear or

linearised. Even more relevant is the fact that equations in the guesstimated models might be incomplete, and subject to misspecification errors.

In fact the RSG algorithm has also been used for estimation a real-life nonlinear forecasting models, giving decent forecasting properties. In particular, the parameters of a series of quarterly models of East European economies (where some of the economies, like the Baltic States and the Czech and Slovak Republics are only a few years old and the data series cannot be long) have been systemmaticaly 'guesstimated' and used for short and medium term forecasting (for the description of the models see Charemza (1994), the RSG algorithm see Blangiewicz and Charemza (1994) and for the independent comparison of various forecasts including those made with the use of 'guesstimated' models see Maciejewski (1997).

The described algorithm of guessing the parameters is far from being ideal. It depends heavily on the choice of initial values and on other assumptions concerning the process of learning, impact of weights on the criterion function and indeed on the choice of the distribution which is supposed to represent the drawing process. One might say, notwithout justification, that so many assumptions creates a conducive environment to 'torturing the data until nature confesses', that is, a researcher might change the assumptions as until a desired result is obtained. This is undoubtedly true, but is also true for traditional econometric models. If the proposed procedure represents a 'back to basics' empirical methodology rather than a joke, then it is likely that further steps in its development will concern about the proper (optimal) choice of penalty weights, constants in the learning formulae and an evaluation of the number of times the guesstimator is willing to correct the priors. If the whole idea of the repetitive stochastic guesstimation is ridiculed, I do hope that somebody will point out to better alternative, or will explain why doing nothing and complaining that 'bad' data do not fit 'good' econometrics is superior to doing something.

References

Angus, J.E. (1989), 'On the connection between neural network learning and multivariate nonlinear least squares estimation', Neural Networks 1, pp. 42-47.

Arkin, V.I. and I.V. Evstigneev (1987), Stochastic models of control and economic dynamics, Academic Press, London.

Barndorff-Nielsen, J.L. Jensen and W.S. Kendall, eds. (1993) Networks and chaos: statistical and probabilistic aspects Chapman and Hall, London.

Blangiewicz, M. and W. Charemza (1994), 'Evaluation of parameters of LAM models', paper presented at the seminar: LAM models: principles, constrution and first results, University of Gdask.

Charemza, W. (1994), 'LAM models for East European economies: general description', paper presented at the seminar: LAM models: principles, constrution and first results, University of Gdask.

Cividini, A. (1992), 'Implementing stochastic optimal control of nonlinear models: a comparison with alternative solution methods', Temi di discussione del Servizio Studi No. 179, Banca d'Italia, Roma.

Fausett, L. (1994), Fundamentals of neural networks, Prentice Hall, Englewood.

Greenaway, D., S.J. Leybourne, G.V. Reed and J. Whalley (1993), Applied general equilibrium modelling: applications, limitations and future development, HMSO, London.

Holly, S. and A. Hughes Hallett (1989), Optimal control, expectations and uncertainty Cambridge U.P., Cambridge.

Hornik, K.M., M. Stinchcombe and H. White (1990), 'Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks', Neural Networks 3, pp. 359-366.

Hughes Hallett, A. and H. Rees (1983), Quantitative economic policies and interactive planning, Cambridge U.P., Cambridge.

Kydland, F.E. and E.C. Prescott (1982), 'Time to built and aggregate fluctuations', Econometrica 50, pp. 1345-1370.

Kydland, F.E. and E.C. Prescott (1991), 'The econometrics of the general equilibrium approach to business cycles' Scandinavian Journal of Economics 93, pp. 161–178.

Kydland, F.E. (1992), 'On the econometrics of world business cycles', European Economic Review 36, pp. 476–482.

Laroque, G. and B. Salanié (1985), 'Macroeconometric disequilibrium models', in H. Pesaran and M. Wickens, eds. Handbook of applied econometrics: Vol. 1: Macroeconomics, Basil Blackwell, Oxford.

Lee. L-F (1997a), 'Simulation estimation of dynamic switching regression and dynamic disequilibirum models: some Monte Carlo results', Journal of Econometrics 78, pp. 179–204.

Lee. L-F (1997b), 'A smooth likelihood simulator for dynamic disequilibrium models', Journal of Econometrics 78, pp. 257–294.

Maciejewski, W. (1997), 'Macroeconomic forecasts for the Polish economy 1995–1996: a comparison', paper presented at the US-Polish Economic Roundtable, Polanica.

McCloskey, D.N. and S.T. Zilak (1996), 'The standard error of regressions', Journal of Economic Literature 34, pp. 97–114.

Plata-Przechlewski, T. (1997), 'Zastosowanie algorytmów genetycznych w modelowaniu nieliniowych zale¨no.ciekonomicznych', Ph.D. dissertation, University of Gda¤sk.

Quandt, R.E. (1988), The econometrics of disequilibrium, Basil Blackwell, Oxford.

Rachtenberg, I. (1973), Evolutionstrategie: Optimierung technisher Systems nach Prinzipien der biologishen Evolution, Fromman-Holzboog, Stuttgart.

Wasserman, P.D. (1989), Neural computing: theory and practice, Van Nostrand Reinhold, New York.