EAGLE™ 1.0

A Capability Profile

Ike's Research Ltd

Technical Document
RD-11-2008
Edition 1.0

Principal Author

ISAAC SIWALE

Copyright © 1997-2008: Ike's Research Ltd 26 January 2008

CONTACTS

EAGLE™ is a trademark of Ike's Research Ltd.
Copyright © 1997-2008
All Rights Reserved Worldwide

Contact:

The Managing Director
Ike's Research Ltd
265, Southcroft Road
London
SW16 6QT

E-mail: Ike_Siwale@tiscali.co.uk

Phone: +44-(0)208-769-8336

PREFACE

NOTICE

This document contains proprietary material. No part of this material, which is protected by this copyright notice, may be reproduced or utilised in any form, or by any electronic or mechanical means, including photocopying, recording or any other informational storage and retrieval system, without written permission from the copyright owner.

Copyright © 1997-2008: Ike's Research Ltd.
All Rights Reserved.

This document outlines a proprietary program called EAGLE. EAGLE — an acronym for ***E***volutionary ***A***lgorithm for ***G***eneral ***L***ikelihood ***E***stimation — is a parameter estimation algorithm that employs the maximum likelihood principle and applies to linear and nonlinear models alike, using two modes of operation. The algorithm's uni-level mode is simply a direct numerical maximization method for the likelihood function: its bi-level mode allows for a decomposition of the estimation task into two interacting search processes: a 'Level 1' algorithm that drives the overall search and compensates for any regression "anomalies" that might arise during the solution process, and a 'Level 2' linear regression algorithm that computes OLS estimates, given data from the former.

The design of the algorithm is founded on three concepts: (a) model linearization and process decomposition via a partitioning of the parameter vector; (b) on-line regression diagnostic testing and control of regression residuals via multi-objective optimization — a process that is here called 'concurrent compensation'; (c) exploitation of the location and scale invariance properties of linear regression. The method is easy to implement and builds on the established methodology of linear regression about which there is extensive knowledge and experience.

Several types of numerical examples were estimated using the bi-level algorithm or its uni-level variant, and the results show the procedure to be very effective. The bi-level algorithm is statistically robust — it gives a good fit even where errors are non-normal.

Finally, it is important to note that, although EAGLE is explained in terms of the normal distribution, it is not restricted to this situation. The design philosophy allows any distribution for the response variable provided one can define the appropriate likelihood function and the statistical tests required for concurrent compensation. Hence, the term 'general' in the acronym 'EAGLE'.

Copyright © 1997-2008: Ike's Research Ltd

CONTENTS

Contacts i

Preface ii

Contents iii

1. Introduction 1

2. Some Numerical Examples 2

2.1. Preamble 2

2.2. Procedure and Results 2

- Example 3.1: Nonlinear Regression. Source: Spitzer (1984) 3
- Example 3.2: Nonlinear Regression. Source: Weisberg (1980) 4
- Example 3.3: Nonlinear Regression. Source: Draper and Smith (1981) 5
- Example 3.4: Nonlinear Regression. Source: Information Tech. Lab., NIST. 6
- Example 3.5: Nonlinear Regression. Source: Information Tech. Lab., NIST. 7
- Example 3.6: Nonlinear Regression. Source: Information Tech. Lab., NIST. 8
- Example 3.7: Nonlinear Regression. Source: Siwale (2002) 9
- Example 3.8: Nonlinear Regression. Source: Spitzer (1984); Siwale (2002) 10
- Example 3.9: Nonlinear Regression. Source: Spitzer (1984); Siwale (2002) 11
- Example 3.10: Nonlinear Regression. Source: Spitzer (1984); Siwale (2002) 12

3. Discussion 15

4. Summary and Conclusions 19

References 20

Copyright © 1997-2008: Ike's Research Ltd

1 Introduction

The purpose of statistical modelling is to derive, via regression techniques, a mathematical model that adequately explains the main variations of a random process, and thus enable one to make meaningful inferences about the said process. The extent to which such a derived model fulfils its purpose depends on the statistical efficiency of the parameter estimators used. Parameter estimation essentially involves solving an optimization problem, commonly formulated either on the 'Least Squares' or the 'Maximum Likelihood' principle,[1] and based on an initial statistical model of the process under study. An implicit assumption of either approach is that a solution that is the globally optimal can be found, for failure to find such a solution would result in estimators that are statistically inefficient, and even inconsistent.

The mechanism and final output of a regression study depend on the probability distribution of the disturbance terms in the initial model. Two fundamental assumptions concerning the nature of the disturbance terms are common: (a) that their effects are additive to those of the predictor variables; and (b) that they satisfy the 'classical' or 'full ideal' conditions — by which is meant that they are mutually independent; they have a mean value of zero; and they have the same constant variance. In general, these assumptions rarely hold, especially when one is dealing with observational data. Furthermore, other "anomalies" such as overly influential data points, inadequate specification of the functional form of the regression model, and multi-collinearity and errors in the independent variables, can adversely affect the regression process and hence the derived model.

Regression diagnostic tests whose purpose is to ascertain the quality and validity of the final model are therefore an important part of every regression study. Unfortunately, current methodologies do not allow for such testing concurrently with the parameter estimation process itself: diagnostic tests are normally only possible "after the fact", i.e. one must have derived a model in order to perform regression diagnostic tests. The lack of a "live feedback channel" between estimation and testing means that, if the diagnostic test results indicate some weakness in the derived model, one is required to repeat the entire modelling exercise. Currently therefore, statistical modelling is, to some extent, still a trial-and-error process.

The need for efficient estimates in statistical modelling cannot be over-emphasised, and finding estimators that are statistically efficient is usually an exercise in *global* numerical optimization (Tvrdík and Křivý, 2004). But many existing computational techniques are not always successful at this task: those that involve the inversion of the information matrix can break down if the said matrix is ill-conditioned; and those that are gradient-based can encounter convergence problems because of 'flat regions' in the landscapes of criterion functions typically associated with under-identified models (Harvey: 1990, p.115). A different approach is therefore called for.

The purpose of this document is to outline the capabilities of EAGLE.[2] EAGLE — an acronym for ***E***volutionary ***A***lgorithm for ***G***eneral ***L***ikelihood ***E***stimation — is a parameter estimator that employs the maximum likelihood principle and applies to linear and nonlinear models alike, using two modes of operation: it may be used in a stand-alone, uni-level mode as a numerical maximum likelihood estimator; or in a bi-level mode as a maximum likelihood estimator "sitting on top of" a linear regression package. The algorithm's bi-level mode concurrently compensates for commonly encountered "anomalies" of linear regression by optimising an appropriately defined vector-valued function. By this device, the bi-level mode alleviates the trial-and-error aspect of statistical modelling. The statistical efficiency requirement is addressed by recourse to evolutionary techniques.

[1] However, for linear models in which errors terms are normally distributed, estimators derived by the two methods are identical.

[2] A trial-version of the program may be obtained by contacting: info@Aptech.com or sales@tomopt.com

Copyright © 1997-2008: Ike's Research Ltd

2 Some Numerical Examples

2.1 Preamble

This section presents a sample of examples that sufficiently illustrate the capabilities of EAGLE. The presentation follows a set pattern as follows. First, a formal statement of the estimation problem is made; this is followed by a listing of EAGLE's output below which are appended the estimated parameters and equation, and some summary statistics of the regression; there then follows a list of diagnostic statistics at the final solution; and finally some pertinent remarks are made. The examples reported here are as follows.

- Example 1 is from Spitzer (1984); it serves to compare EAGLE with the algorithm reported therein.
- Example 2 is from Weisberg (1980, p.139); it serves to compare EAGLE with the two-phased Box-Cox / Box-Tidwell algorithm that Weisberg recommends.
- Example 3 is from Draper and Smith (1981, p.475), and it is reported in two parts. Example 3a serves to demonstrate that EAGLE is equally effective when used in uni-level mode. Example 3b is a repeat of (3a) but this time solved using the bi-level mode—its serves to illustrate how EAGLE may be used on nonlinear regression equations other than the Extended Box-Cox model.
- Examples 4, 5 and 6 are standard test problems from the Information Technology Laboratory of the National Institute on Standards and Technology.[3] They feature in a recent comparative study by Tvrdík and Křivý (2004), and therefore serve to compare EAGLE against the various algorithms tested therein. Example 4 is in two parts: the results of (4a) were obtained using the uni-level mode, and those of (4b) using the bi-level mode.
- Example 7 is typical of the artificial problems that were designed to test EAGLE's efficacy as a parameter estimator during its development.
- Examples 8, 9 and 10 test the effectiveness of EAGLE as a modelling tool in the absence of NIID errors. The error terms were intentionally set to violate the normal assumptions made regarding the error distribution. For Example 8, the error terms were drawn from a uniform distribution on the interval [-0.5, 0.5]; for Example 9, the error terms were assumed to follow a first-order auto-regressive process, namely:

$$\varepsilon_{it} = 0.6\varepsilon_{it-1} + w_t;\ w_t \sim \mathrm{NIID}(0, 0.2)\ . \tag{2.1a}$$

For Example 10, the error terms were a sum of the two previous types, i.e.:

$$\varepsilon_{it} = 0.6\varepsilon_{it-1} + v_t + w_t;\ v_t \sim \mathrm{U}(-0.5, 0.5);\ w_t \sim \mathrm{NIID}(0, 0.2)\ . \tag{2.1b}$$

2.2 Procedure and Results

The experimental procedure was simply to run EAGLE for a fixed number of generations, several times for each example, adjusting some parameters of the program between each execution. For each run, the maximum value attained by the likelihood component of the objective function was recorded; the results reported are those pertaining to the run with the best likelihood score. They include (1) the progress of the best member of the (genetic) population; (2) the estimated parameters at the end of the program run; and (3) the estimated equation together with its associated statistics—all figures reported to three decimal places. The equation statistics include (a) the adjusted coefficient of multiple determination, (b) the estimated variance, (c) the equation F-Statistic.

[3] http://www.itl.nist.gov/div898/strd/general/bkground.html

Copyright © 1997-2008: Ike's Research Ltd

Example 1

Regression Model: $y_t(\lambda_4) = \alpha_0 + \alpha_1 f_1(\lambda_1, x_{1t}) + \alpha_2 f_2(\lambda_2, x_{2t}) + \alpha_3 f_3(\lambda_3, x_{3t}) + \varepsilon_t; \quad f_i(\cdot, x_{it}) = x_{it}$

Data Source:[4] $y_t = -1.5 + x_{1t} + x_{2t} + x_{3t} + \omega_t; \quad \omega_t \sim \text{NIID}(0, 0.107474)$

I. EAGLE **Output**

Generation	*Objective 1*	*Objective 2*
0	-19.504083	0.336457
20	73.967673	0.000000
40	73.985527	0.000000
60	73.986482	0.000000
70	73.986486	0.000000
80	73.986486	0.000000
90	73.986487	0.000000
100	73.986487	0.000000

Nonlinear Parameter Vector: $\boldsymbol{\lambda}^* = (0.988348, 0.985113, 1.025304, 0.989334)^T$

Linear Parameter Vector: $\boldsymbol{\alpha}^* = (-1.445691, 1.001354, 1.003264, 0.939360)^T$

Estimated Equation: $y_t^{0.989} = \underset{(15.84)}{-1.446} + \underset{(91.08)}{1.001} x_{1t}^{0.988} + \underset{(89.54)}{1.003} x_{2t}^{0.985} + \underset{(85.89)}{0.939} x_{3t}^{1.025}$

Equation Statistics: $R^2_{adj} = 0.999$ $\hat{\sigma}^2_\varepsilon = 8.1706\text{E-}2$ $F = 7.230\text{E}3$

II. **Regression Diagnostics at Solution**

```
a. Residual Variance:                    0.006821

b. Condition Number:                     4.895601

c. Bamset Statistic:                     2.933684

d. Skewness Statistic:                   0.150015

e. Durbin-Watson Statistic:              1.824442

f. Shapiro-Francia Statistic:            0.973536
```

III. **General Remarks**

The parameter space $\boldsymbol{\Omega}$ is circumscribed via the input data that is supplied to EAGLE, and for this example, the search space was set as $\boldsymbol{\Omega} = [-4, 4]^4 \times \mathbf{R}^4 \times \mathbf{R}_+$. Note that one could choose to "switch off" λ_1 - λ_3 from the evolutionary search by fixing the values at '1' in the input data file since we know, from the data generator, that this is the correct value for these particular parameters (see General Remarks on Example 3b for illustration). However, for the purposes of this paper, it was felt that the stiffer test of estimating three extra parameters would be more informative of the capabilities of EAGLE. Observe that all diagnostic statistics are within their respective pre-specified bounds.

[4] **Source**: Spitzer (1984)

Copyright © 1997-2008: Ike's Research Ltd

Example 2

Regression Model: $y_t(\lambda_2) = \alpha_0 + \alpha_1 f_1(\lambda_1, x_{1t}) + \varepsilon_t; \quad f_i(\cdot, x_{it}) = x_{it}$

Data Source: Weisberg (1980, p.139)

I. EAGLE **Output**

Generation	*Objective 1*	*Objective 2*
0	0.748053	0.000000
20	3.084515	0.000000
40	3.085529	0.000000
60	3.085678	0.000000
80	3.085689	0.000000
100	3.085753	0.000000
120	3.085753	0.000000
140	3.085830	0.000000
160	3.085833	0.000000
180	3.085836	0.000000
200	3.085840	0.000000

Nonlinear Parameter Vector: $\boldsymbol{\lambda}^* = (0.098790, 0.599230)^T$

Linear Parameter Vector: $\boldsymbol{\alpha}^* = (-8.583067, 8.267309)^T$

Estimated Equation: $y_t^{0.599} = \underset{(-5.501)}{-8.583} + \underset{(11.300)}{8.267}\, x_{1t}^{0.099}$

Equation Statistics: $R^2_{adj} = 0.851$ $\quad \hat{\sigma}^2_\varepsilon = 9.391\text{E-}2$ $\quad F = 1.429\text{E}2$

II. **Regression Diagnostics at Solution**

```
a. Residual Variance:                    0.093913
b. Condition Number:                     9.650907
c. Bamset Statistic:                     4.441987
d. Skewness Statistic:                  -0.324212
e. Durbin-Watson Statistic:              1.854155
f. Shapiro-Francia Statistic:            0.970265
```

III. **General Remarks**

For this example, the search space was set as $\boldsymbol{\Omega} = [-4, 4]^2 \times \mathbf{R}^2 \times \mathbf{R}_+$. Although the quality of fit, as measured by R^2, is only marginally better than that reported by Weisberg (1980, p.139), namely $R^2 = 0.844$, EAGLE achieves an 88.3% reduction in the residual variance — $\sigma^2 = 0.093913$ as compared to $\sigma^2 = 0.802$. Furthermore, all diagnostic statistics are within their respective pre-specified bounds.

Copyright © 1997-2008: Ike's Research Ltd

Example 3a

Regression Model: $y_t = \alpha_0 + (0.49 - \alpha_0)\exp(-\beta_1(x_{1t} - 8)) + \varepsilon_t$

Data Source: Draper and Smith (1981, p.475)

I. EAGLE **Output**

Generation	*Objective 1*	*Objective 2*
0	24.582465	0.327134
10	190.758330	0.000172
20	197.732063	0.000125
30	199.277558	0.000116
40	199.300319	0.000116
50	199.301601	0.000116
60	199.301671	0.000116
70	199.301995	0.000116
80	199.301995	0.000116
90	199.301996	0.000116
100	199.301996	0.000116

Nonlinear Parameter Vector: $\boldsymbol{\beta}^* = (0.390139, 0.101631)^T$

Estimated Equation: $y_t = 0.390 + 0.099\exp(-0.102(x_{1t} - 8))$

II. **Regression Diagnostics at Solution**

```
a. Residual Variance:                        0.000116

b. Residual Standard Deviation:              0.010785

c. Residual Sum of Squares:                  0.005002
```

III. **General Remarks**

This example is not particularly difficult to solve. In fact, Draper and Smith (pp.476-481) demonstrate that one can solve it "by hand" using an analytic-cum-graphical method, amongst other techniques. The point of including it here is merely to show that EAGLE's uni-level mode is effective, and to provide a simple example that illustrates how one applies the bi-level method to nonlinear models via variable transformations (see Example 3b below).

For this example, the search space was set as $\boldsymbol{\Omega}$ = [-10, 10] x [-10, 10], and the problem was solved by EAGLE in the uni-level mode. The results by EAGLE are practically similar to those reported by Draper and Smith; they however do not report an estimate for the equation variance.

Copyright © 1997-2008: Ike's Research Ltd

Example 3b

Regression Model: $\bar{y}_t(\lambda_2) = \alpha_0 + \alpha_1(1 - f_1(\beta_1, \lambda_1, x_{1t})) + \varepsilon_t; \quad f_1 = \exp(-\beta_1(x_{1t} - 8)); \quad \bar{y}_t = y_t - 0.49 f_1$

Data Source: Draper and Smith (1981, p.475)

I. EAGLE **Output**

Generation	*Objective 1*	*Objective 2*
0	190.440197	0.278732
20	198.804557	0.000000
40	198.805651	0.000000
80	198.805689	0.000000
100	198.805689	0.000000
120	198.805689	0.000000
140	198.805689	0.000000
160	198.805689	0.000000
180	198.805689	0.000000
200	198.805689	0.000000

Nonlinear Parameter Vector: $(\beta^*, \boldsymbol{\lambda}^*) = (0.099155, 1.000000, 1.000000)^T$

Linear Parameter Vector: $\boldsymbol{\alpha}^* = (-0.001122, 0.390750)^T$

Estimated Equation: $y_t = 0.390 + 0.099 \exp(-0.099(x_{1t} - 8))$

Equation Statistics: $R^2_{adj} = 0.991$ $\hat{\sigma}^2_\varepsilon = 1.190E\text{-}4$ $F = 4.495E3$

II. **Regression Diagnostics at Solution**

```
a. Residual Variance:                    0.000119
b. Condition Number:                     3.159270
c. Bamset Statistic:                     0.340370
d. Skewness Statistic:                   0.300266
e. Durbin-Watson Statistic:              1.980297
f. Shapiro-Francia Statistic:            0.937487
```

III. **General Remarks**

This example is a repeat of Example 3a but this time solved using the bi-level algorithm. An explanation of the regression equation is provided in § 4 below. The search space was set as $\boldsymbol{\Omega} = [0, 1] \times [1, 1]^2 \times \mathbf{R}^2 \times \mathbf{R}_+$. That is to say: $\beta \in [0, 1]$, $\lambda_1 \in \{1\}$, $\lambda_2 \in \{1\}$, $\boldsymbol{\alpha} \in \mathbf{R}^2$ and $\sigma^2 \in \mathbf{R}_+$. This input data specification effectively "switches off" the Box-Cox vector $\boldsymbol{\lambda}$ from the evolutionary search. Apart from the equation variance, the estimated parameters are essentially the same as those in Example 3a. But, of course, the bi-level method yields more information concerning the final solution in the form of diagnostic statistics. And, as can be seen above, all these are within their respective pre-specified bounds.

Copyright © 1997-2008: Ike's Research Ltd

Example 4a

Regression Model: $y_t = \beta_1 \left(1 - \exp(-\beta_2 x_{1t})\right) + \varepsilon_t$

Data Source:[5] National Institute of Standards and Technology (NIST)
Statistical Reference Datasets
Reference code: BoxBOD

I. EAGLE **Output**

Generation	Objective 1	Objective 2
0	-23.954259	2935.852000
20	-16.361263	233.633674
40	-16.360885	233.604216
60	-16.360854	233.601779
80	-16.360854	233.601775
100	-16.360854	233.601775
120	-16.360854	233.601775
140	-16.360854	233.601775
160	-16.360854	233.601775
180	-16.360854	233.601775
200	-16.360854	233.601775
220	-16.360854	233.601775
240	-16.360854	233.601775
260	-16.360854	233.601775
280	-16.360854	233.601775
300	-16.360854	233.601775

Nonlinear Parameter Vector: $\boldsymbol{\beta}^* = (213.809409, 0.547237)^T$

Estimated Equation: $y_t = 213.809\left(1 - \exp(-0.547 x_{1t})\right)$

II. **Regression Diagnostics at Solution**

```
a. Residual Variance:                    233.601775

b. Residual Standard Deviation:           15.284037

c. Residual Sum of Squares:             1168.008877
```

III. **General Remarks**

For this example, the search space was set as $\boldsymbol{\Omega}$ = [0, 1000] x [0, 1000], and the problem was solved by EAGLE in the uni-level mode. All estimates coincide exactly with the certified values reported by NIST.

[5] http://www.itl.nist.gov/div898/strd/nls/data/boxbod.shtml

Copyright © 1997-2008: Ike's Research Ltd

Example 4b

Regression Model: $y_t(\lambda_2) = \alpha_0 + \alpha_1 f_1(\beta_1, \lambda_1, x_{1t}) + \varepsilon_t; \quad f_1(\beta_1, \cdot, x_{1t}) = -\exp(-\beta_1 x_{1t})$

Data Source:[6] National Institute of Standards and Technology (NIST)
Statistical Reference Datasets
Reference code: BoxBOD

I. EAGLE **Output**

Generation	*Objective 1*	*Objective 2*
0	-12.526729	0.734905
20	-13.835578	0.223179
40	-13.857588	0.214621
60	-13.875464	0.207697
80	-13.875508	0.207680
100	-13.875508	0.207680
180	-13.875508	0.207680
260	-13.875508	0.207680
300	-13.875508	0.207680

Nonlinear Parameter Vector: $(\beta^*, \boldsymbol{\lambda}^*) = (0.118188, 1.000000, 1.000000)^T$

Linear Parameter Vector: $\boldsymbol{\alpha}^* = (292.776993, 196.251984)^T$

Estimated Equation: $y_t = 292.777 - 196.252\exp(-0.118x_{1t})$

Equation Statistics: $R^2_{adj} = 0.958$ $\hat{\sigma}^2_\varepsilon = 102.020$ $F = 91.780$

II. **Regression Diagnostics at Solution**

```
a. Residual Variance:                    102.020043
b. Condition Number:                      16.567751
c. Bamset Statistic:                       4.968858
d. Skewness Statistic:                     0.055823
e. Durbin-Watson Statistic:                2.707680
f. Shapiro-Francia Statistic:              0.880162
```

III. **General Remarks**

A quasi-linearization technique similar to that employed in Example 3b was applied to the original model $y_t = b_1(1 - \exp[-b_2 x_{1t}])$. The search space was set as $\boldsymbol{\Omega} = [0, 1] \times [1, 1]^2 \times \mathbf{R}^2 \times \mathbf{R}_+$, thus effectively "switching off" the Box-Cox vector $\boldsymbol{\lambda}$ from the evolutionary search. The residual variance is much reduced, although, obviously the estimated equation is different in structure from the original model.

[6] http://www.itl.nist.gov/div898/strd/nls/data/boxbod.shtml

Copyright © 1997-2008: Ike's Research Ltd

Example 5

Regression Model: $\bar{y}_t(\lambda_2) = \alpha_0 + \alpha_1 f_1(\beta_1, \lambda_1, x_{1t}) + \varepsilon_t; \quad f_1(\beta_1, \cdot, x_{1t}) = (x_{1t} + \beta_1)^{-1}; \quad \bar{y}_t = \mathrm{Ln}(y_t)$

Data Source:[7] National Institute of Standards and Technology (NIST)
Statistical Reference Datasets
Reference code: MGH10

I. EAGLE **Output**

Generation	Objective 1	Objective 2
0	55.380769	1.357790
40	139.010110	0.000000
80	139.011494	0.000000
120	139.011494	0.000000
160	139.011494	0.000000
200	139.011494	0.000000
240	139.011494	0.000000
280	139.011494	0.000000
300	139.011495	0.000000
400	139.011495	0.000000
500	139.011495	0.000000

Nonlinear Parameter Vector: $(\beta^*, \boldsymbol{\lambda}^*) = (344.424058, 1.000000, 1.000000)^T$

Linear Parameter Vector: $\boldsymbol{\alpha}^* = (-5.155381, 6157.861348)^T$

Estimated Equation: $\bar{y}_t = -5.155 + 6157.861(x_{1t} + 344.424)^{-1}$

II. **Regression Diagnostics at Solution**

```
a. Residual Variance:                        0.000000
b. Condition Number:                     16049.902801
c. Bamset Statistic:                         2.571227
d. Skewness Statistic:                       0.500043
e. Durbin-Watson Statistic:                  1.585406
f. Shapiro-Francia Statistic:                0.907896
```

III. **General Remarks**

A quasi-linearization approach similar to that used in Example 3b may, in principle, be applied directly to the original model $y_t = b_1 \exp[b_2(x_{1t} + b_3)^{-1}]$, but better results accrue by first transforming the model into the additive structure, $\mathrm{Ln}(y_t) = \mathrm{Ln}(b_1) + b_2(x_{1t} + b_3)^{-1}$, and then applying the said method. Although the condition number seems large, EAGLE's estimates of the parameters b_1, b_2 and b_3 almost coincide with the NIST values; the much reduced residual variance suggests that the former may be closer to the true parameter values.

[7] http://www.itl.nist.gov/div898/strd/nls/data/mgh10.shtml

Copyright © 1997-2008: Ike's Research Ltd

Example 6

Regression Model: $y_t = (\beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3)\,(1 + \beta_5 x + \beta_6 x^2 + \beta_7 x^3)^{-1} + \varepsilon_t$

Data Source:[8] National Institute of Standards and Technology (NIST)
Statistical Reference Datasets
Reference code: Thurber

I. EAGLE **Output**

Generation	Objective 1	Objective 2
0	216.0429101	17947.531325
40	-126.335747	924.228482
80	-98.015677	199.965475
120	-96.312494	182.377868
160	-96.286273	182.119560
200	-96.283877	182.095975
240	-96.282545	182.082866
280	-96.280425	182.062002
320	-96.279969	182.057518
360	-96.278832	182.046323
400	-96.277112	182.029398
440	-96.275290	182.011473
480	-96.273570	181.994554
500	-96.273389	181.992774

Nonlinear Parameter Vector: $\boldsymbol{\beta}^* = (1286.786376, 1506.499750, 595.180611, 77.685168, 0.981125, 0.404239, 0.050574)^T$

Estimated Eqn: $y_t = (1286.79 + 1506.50x + 595.181x^2 + 77.685x^3)\,(1 + 0.981x + 0.404x^2 + 0.051x^3)^{-1}$

II. **Regression Diagnostics at Solution**

```
a. Residual Variance:                        181.992774

b. Residuals Standard Deviation:              13.490470

c. Residuals Sum of Squares:                5641.775992
```

III. **General Remarks**

As with Example 5 above, EAGLE's estimates of the parameters b_1, b_2, b_3 b_4, b_5, b_6 and b_7 almost coincide with the NIST values; however, the residual variance is smaller and this suggests that the EAGLE parameters may be closer to the true values.

[8] http://www.itl.nist.gov/div898/strd/nls/data/thurber.shtml

Copyright © 1997-2008: Ike's Research Ltd

Example 7

Regression Model: $y_t(\lambda_4) = \alpha_0 + \alpha_1 f_1(\lambda_1, x_{1t}) + \alpha_2 f_2(\lambda_2, x_{2t}) + \alpha_3 f_3(\lambda_3, x_{3t}) + \varepsilon_t; \quad f_i(\cdot, x_{it}) = x_{it}$

Data Source:[9] $y_t^{0.55} = 2x_{1t}^{0.22} + 3x_{2t}^{0.33} + 4x_{3t}^{0.44} + \omega_t; \quad \omega_t \sim \text{NIID}(0, 0.001); \quad x_{it} \sim 10\text{U}(0,1)$

I. EAGLE Output

Generation	*Objective 1*	*Objective 2*
0	-95.443155	0.000000
20	0.457075	0.000000
40	65.053261	0.000000
60	108.181866	0.000000
80	125.908667	0.000000
100	210.723094	0.000000
120	210.773047	0.000000
140	213.506937	0.000000
160	213.506970	0.000000
180	213.516967	0.000000
200	213.516967	0.000000

Nonlinear Parameter Vector: $\boldsymbol{\lambda}^* = (0.220080, 0.329999, 0.440010, 0.550100)^T$

Linear Parameter Vector: $\boldsymbol{\alpha}^* = (-0.001952, 2.000582, 3.002066, 4.002574)^T$

Estimated Equation: $y_t^{0.550} = \underset{(>1000)}{-0.002} + \underset{(>1000)}{2.001}\, x_{1t}^{0.220} + \underset{(>1000)}{3.002}\, x_{2t}^{0.330} + \underset{(>1000)}{4.003}\, x_{3t}^{0.440}$

Equation Statistics: $R^2_{adj} = 1.000$ $\quad \hat{\sigma}^2_\varepsilon = 0.000$ $\quad F = 2.141\text{E}10$

II. Regression Diagnostics at Solution

```
a. Variance:                          0.000000
b. Condition Number:                  8.661957
c. Bamset Statistic:                  5.669785
d. Skewness Statistic:                0.493459
e. Durbin-Watson Statistic:           1.780419
f. Shapiro-Francia Statistic:         0.961181
```

III. General Remarks

For this example, the search space was set as $\Omega = [-4, 4]^4 \times \mathbf{R}^4 \times \mathbf{R}_+$. The estimation results above speak for themselves.

[9] Siwale (2002)

Copyright © 1997-2008: Ike's Research Ltd

Example 8

Regression Model: $y_t(\lambda_4) = \alpha_0 + \alpha_1 f_1(\lambda_1, x_{1t}) + \alpha_2 f_2(\lambda_2, x_{2t}) + \alpha_3 f_3(\lambda_3, x_{3t}) + \varepsilon_t; \quad f_i(\cdot, x_{it}) = x_{it}$

Data Source: $y_t = -1.5 + x_{1t} + x_{2t} + x_{3t} + \omega_t; \quad \omega_t \sim U(-0.5, 0.5)$

I. EAGLE **Output**

Generation	*Objective 1*	*Objective 2*
0	-20.086104	0.000000
10	35.072423	0.000000
20	36.547247	0.000000
30	36.889041	0.000000
40	36.902160	0.000000
50	36.904674	0.000000
60	36.904765	0.000000
70	36.904832	0.000000
80	36.904835	0.000000
90	36.904835	0.000000
100	36.904835	0.000000

Nonlinear Parameter Vector: $\boldsymbol{\lambda}^* = (1.014330, 0.819143, 1.056153, 0.981380)^T$

Linear Parameter Vector: $\boldsymbol{\alpha}^* = (1.328126, 0.828662, 1.396256, 0.847389)^T$

Estimated Equation: $y_t^{0.981} = \underset{(15.60)}{1.328} + \underset{(88.45)}{0.829}\, x_{1t}^{1.014} + \underset{(88.45)}{1.396}\, x_{2t}^{0.819} + \underset{(84.89)}{0.847}\, x_{3t}^{1.056}$

Equation Statistics: $R^2_{adj} = 0.988$ $\quad \hat{\sigma}^2_\varepsilon = 7.596\text{E-}2$ $\quad F = 7.418\text{E}2$

II. **Regression Diagnostics at Solution**

```
a. Variance:                          0.075962
b. Condition Number:                  5.023523
c. Bamset Statistic:                  3.090866
d. Skewness Statistic:                0.325685
e. Durbin-Watson Statistic:           1.673699
f. Shapiro-Francia Statistic:         0.949291
```

III. **General Remarks**

For this example, the search space was set as $\boldsymbol{\Omega} = [-4, 4]^4 \times \mathbf{R}^4 \times \mathbf{R}_+$. And, as can be seen from the diagnostic tests at the solution, the algorithm "normalises" the error distribution and yields a very good fit to the data.

Copyright © 1997-2008: Ike's Research Ltd

Example 9

Regression Model: $y_t(\lambda_4) = \alpha_0 + \alpha_1 f_1(\lambda_1, x_{1t}) + \alpha_2 f_2(\lambda_2, x_{2t}) + \alpha_3 f_3(\lambda_3, x_{3t}) + \varepsilon_t; \quad f_i(\cdot, x_{it}) = x_{it}$

Data Source: $y_t = -1.5 + x_{1t} + x_{2t} + x_{3t} + \omega_t; \quad \omega_t = 0.6\omega_{t-1} + \phi_t; \quad \phi_t \sim \text{NIID}(0, 0.2)$

I. EAGLE Output

Generation	Objective 1	Objective 2
0	-16.349574	0.000000
20	37.635034	0.000000
40	38.384242	0.000000
60	38.553593	0.000000
80	38.554459	0.000000
100	38.554463	0.000000
120	38.554463	0.000000
140	38.554463	0.000000
160	38.554463	0.000000
180	38.554463	0.000000
200	38.554463	0.000000

Nonlinear Parameter Vector: $\boldsymbol{\lambda}^* = (0.819380, 1.161990, 1.064060, 0.346684)^T$

Linear Parameter Vector: $\boldsymbol{\alpha}^* = (1.588204, 0.114435, 0.056432, 0.071302)^T$

Estimated Equation: $y_t^{0.347} = \underset{(15.60)}{1.588} + \underset{(88.45)}{0.114}\, x_{1t}^{0.819} + \underset{(88.45)}{0.056}\, x_{2t}^{1.162} + \underset{(84.89)}{0.071}\, x_{3t}^{1.064}$

Equation Statistics: $R^2_{adj} = 0.984$ $\quad \hat{\sigma}^2_\varepsilon = 5.28\text{E-}4$ $\quad F = 5.478\text{E}2$

II. Regression Diagnostics at Solution

```
a. Variance:                          0.000528
b. Condition Number:                  7.112158
c. Bamset Statistic:                  2.874424
d. Skewness Statistic:                0.103995
e. Durbin-Watson Statistic:           1.500000
f. Shapiro-Francia Statistic:         0.977277
```

III. General Remarks

For this example, the search space was set as $\boldsymbol{\Omega} = [-4, 4]^4 \times \mathbf{R}^4 \times \mathbf{R}_+$. As can be seen from the diagnostic tests at the solution, the algorithm "normalises" the error distribution and yields a very good fit to the data.

Copyright © 1997-2008: Ike's Research Ltd

Example 10

Regression Model: $y_t(\lambda_4) = \alpha_0 + \alpha_1 f_1(\lambda_1, x_{1t}) + \alpha_2 f_2(\lambda_2, x_{2t}) + \alpha_3 f_3(\lambda_3, x_{3t}) + \varepsilon_t; \quad f_i(\cdot, x_{it}) = x_{it}$

Data Source:

$$y_t = -1.5 + x_{1t} + x_{2t} + x_{3t} + \omega_t; \quad \omega_t = 0.6\omega_{t-1} + \phi_t + \varphi_t; \quad \phi_t \sim \text{NIID}(0, 0.2) \quad \varphi_t \sim \text{U}(-0.5, 0.5)$$

I. EAGLE **Output**

Generation	*Objective 1*	*Objective 2*
0	-17.754724	0.000000
20	34.702001	0.000000
40	35.472699	0.000000
60	35.519152	0.000000
80	35.522458	0.000000
100	35.523771	0.000000
120	35.523772	0.000000
140	35.523772	0.000000
160	35.523772	0.000000
180	35.523772	0.000000
200	35.523782	0.000000

Nonlinear Parameter Vector: $\boldsymbol{\lambda}^* = (0.790770, 0.932208, 1.216555, 1.252945)^T$

Linear Parameter Vector: $\boldsymbol{\alpha}^* = (-0.861939, 3.162547, 2.635666, 1.482148)^T$

Estimated Equation: $y_t^{1.253} = -\underset{(15.60)}{0.862} + \underset{(89.71)}{3.162} x_{1t}^{0.790} + \underset{(88.45)}{2.636} x_{2t}^{0.932} + \underset{(84.89)}{1.482} x_{3t}^{1.217}$

Equation Statistics: $R^2_{adj} = 0.990$ $\hat{\sigma}^2_\varepsilon = 8.077\text{E-}1$ $F = 8.30\text{E}2$

II. **Regression Diagnostics at Solution**

```
a. Variance:                          0.431597
b. Condition Number:                  4.408000
c. Bamset Statistic:                  2.392576
d. Skewness Statistic:                0.097269
e. Durbin-Watson Statistic:           1.500000
f. Shapiro-Francia Statistic:         0.983248
```

III. **General Remarks**

For this example, the search space was set as $\boldsymbol{\Omega} = [-4, 4]^4 \times \mathbf{R}^4 \times \mathbf{R}_+$. Once again, diagnostic statistics at the final solution show that EAGLE manages to "normalise" the error distribution and yields a very good fit to the data.

Copyright © 1997-2008: Ike's Research Ltd

3 Discussion

On Example 1

This example serves to compare the performance of EAGLE to the method advocated by Spitzer (1982; 1984). In his study, Spitzer generates 60 data points in which the regressor values are from a uniform distribution with a mean of 2.25 and variance of 1.021. He computes the response values using the model:

$$y_i = -1.5 + x_{i1} + x_{i2} + x_{i3} + \varepsilon_i; \quad \varepsilon_i \sim \text{NIID}(0, 0.107474) \tag{3.1}$$

He proceeds to fit the normal Box-Cox model to the data thus generated, using the maximum likelihood principle. To solve the likelihood maximisation problem, Spitzer first re-scales the Box-Cox model and then derives a supposedly equivalent non-linear least squares problem. His solution procedure is a modification of Newton's algorithm, which he informally describes as follows:

> "First minimise the [sum of squares function] $S(\lambda, \boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$, conditional on λ. The $\boldsymbol{\beta}$ that minimises the $S(\lambda, \boldsymbol{\beta})$ a function of λ, and therefore $S(\lambda, \boldsymbol{\beta})$ is a function of λ only. The (k+1)-dimensional problem has now been transformed into a one-dimensional problem. Second, take a single Newton iteration towards minimising $S(\lambda, \boldsymbol{\beta})$ with respect to λ. These two steps are repeated until a convergence criterion is met" (Paraphrased from Spitzer: 1982, p.761)

His estimated equation is:

$$y_i^{1.188} = \underset{(29.27)}{-3.491} + \underset{(11.78)}{1.486}\, x_{i1} + \underset{(19.8)}{1.523}\, x_{i2} + \underset{(11.75)}{1.479}\, x_{i3}\,.$$

Example 1 is essentially equivalent to the model studied by Spitzer. However, the generalised nature of the EBC specification presumably renders it more difficult to solve due to the larger number of parameters — eight as compared to five in Spitzer's model. Nevertheless, EAGLE's bi-level procedure yields an equation that is much closer to the true equation than Spitzer's algorithm, namely:

$$y_t^{0.989} = \underset{(15.84)}{-1.446} + \underset{(91.08)}{1.001}\, x_{1t}^{0.988} + \underset{(89.54)}{1.003}\, x_{2t}^{0.985} + \underset{(85.89)}{0.939}\, x_{3t}^{1.025}\,.$$

In addition, the theoretical foundation of EAGLE is firmer than Spitzer's method since it does not involve a re-scaling of the regression model—a technique that is rather problematic. The said re-scaling procedure has been common practice since Zarembka (1968). Of late however, others have questioned its validity. Dagenais and Dufour (1994) show that because the scaling factor applied in deriving the 'scaled model' is in fact a random variable, one is lead to use an incorrect concentrated log likelihood function for the scaled variables. They write:

> "Re-scaling the dependent variable with the sample geometric mean (without taking into account the random nature of the scaling factor) leads one to use the wrong likelihood function: the estimated model is not the same as the original specification. The sample geometric mean modifies the model in basic way and thus involves much more than a mere unit change [. . .] regression standard errors obtained from the re-scaled model are not generally valid, so that these standard errors cannot be used to build confidence sets for either the original or the re-scaled model" (Paraphrased from p.571)

Finally, note that, because both levels of the EAGLE algorithm employ the maximum likelihood principle, the computed estimates are asymptotically efficient.

Copyright © 1997-2008: Ike's Research Ltd

On Example 2

This example affords a comparison of EAGLE to the two-phase Box-Cox/Box-Tidwell estimation procedure suggested by Box and Cox (1964). Weisberg (1980, p.138-144) uses the Box-Cox/Box-Tidwell approach to fit the EBC model to some data pertaining to the growth rate of Oak trees. Weisberg initially estimates the optimum transform for the response variable as $\lambda_2 = 1.3$ by the graphical method of Box and Cox, but rationalises down to $\lambda_2 = 1$ as follows:

> "The 90% confidence interval for λ_2 is the set [0.9, 1.7]. Since this interval includes the value $\lambda_2 = 1$ corresponding to no transformation, it appears that little is to be gained by transforming the response variable. Box and Cox suggest using relatively simple choices for λ_2 since the difference between $\lambda_2 = 1.3$ and $\lambda_2 = 1$ is small in terms of the values of y^{λ} but the interpretation of the latter is generally simpler than that of the former." (Paraphrased from p.141)

Having adopted the value of $\lambda_2 = 1$, Weisberg then uses the Box-Tidwell method to estimate the optimum transform for the independent variable as $\lambda_1 = 0.0926$. He then asserts (p.142) that:

> "for most practical purposes, it is adequate to note that λ_1 is nearly zero, and a log transformation is suggested".

Thus, he reports his final estimated equation as:

$$y_t = \underset{(5.15)}{-3.928} + \underset{(11.64)}{2.980}\,\mathrm{Ln}(x_t)$$

$$R^2 = 0.844 \qquad \hat{\sigma}_{\varepsilon}^2 = 0.802$$

By contrast, the EAGLE only needs a single run to estimate all the parameters of the regression model. The estimated equation is:

$$y_t^{0.599} = \underset{(-5.501)}{-8.583} + \underset{(11.300)}{8.267}\, x_{1t}^{0.099}$$

$$R_{adj}^2 = 0.851 \qquad \hat{\sigma}_{\varepsilon}^2 = 0.094$$

Note that, although EAGLE reports only a slightly higher quality of fit, the equation variance is significantly better than that reported by Weisberg.

On Example 3

The first part of this example is a direct numerical maximization of the log likelihood function; it needs no further comment except to point out that EAGLE's solution agrees with that reported by Draper and Smith (pp.476-481). The purpose of the second part is to provide a simple illustration of how one would implement the bi-level mode in this case.

The original regression model is:

$$y_t = \alpha + (0.49 - \alpha)e^{-\beta(x_t - 8)} + \varepsilon_t \qquad (3.2a)$$

Copyright © 1997-2008: Ike's Research Ltd

To estimate equation (3.2a) using EAGLE's bi-level mode, one assumes the quasi-linear model:

$$y_t = \delta_0 + \delta_1 z_t + \varepsilon_t \tag{3.2b}$$

where: $\delta_0 \equiv \alpha$; $\delta_1 \equiv (0.49 - \alpha)$; $z_t = \exp(-\hat{\beta}^k (x_t - 8))$

At the *k*-th iteration, the Level 1 algorithm transforms the original regressor data according to

$$z_t = \exp(-\hat{\beta}^k (x_t - 8)) \tag{3.2c}$$

Given the transformed data, the Level 2 program evaluates the OLS estimators for $\hat{\delta}_0^k$, $\hat{\delta}_1^k$, $\hat{\sigma}^k$. Note that because the linear parameters in (3.2b) are not independent, the parameter α can only be retrieved from δ_0 or δ_1 if the following constraint is satisfied exactly:

$$\delta_0 + \delta_1 - 0.49 = 0 \tag{3.2d}$$

Although EAGLE accommodates such constraints on OLS parameters, there is no guarantee that the constraint will be satisfied at the final solution. An alternative approach that estimates the original parameters directly and does not require a constraint follows a trivial manipulation of (3.2b) to yield the regression equation:

$$y_t - 0.49 z_t = \alpha(1 - z_t) + \varepsilon_t \tag{3.2e}$$

The results reported for Example 3b were obtained using this regression model. Apart from the equation variance, the estimated parameters are essentially the same as those in Example 3a. But, of course, the bi-level method yields more information concerning the final solution in the form of diagnostic statistics.

On Example 4, 5 and 6

These examples are part of the NIST test suite for nonlinear regression algorithms. They feature in a recent comparative study by Tvrdík and Křivý (2004) in which several statistical packages for least squares regression were tested, namely: NCSS 2001, where the Levenberg-Marquardt algorithm is used; S-PLUS 4.5 which employs the Gauss-Newton algorithm; SPSS 10.0 which uses the modified Levenberg-Marquardt algorithm; and SYSTAT 8.0 where both a modified Gauss–Newton algorithm and the simplex method are implemented. These three examples were selected because they were the ones that most of the statistical packages in the Tvrdík- Křivý study found difficult to solve. Table 4.1 below is an extract from there to which has been added the result by EAGLE for comparison. And, as can be seen, most of the packages were not completely successful in all cases despite the fact that the recommended starting values are actually very near to the certified estimates.

For Example 4b, a quasi-linearization technique similar to that employed in Example 3b was applied to the original model $y_t = b_1(1 - \exp[-b_2 x_{1t}])$. Thus, one assumes the quasi-linear model:

$$y_t = \delta_0 + \delta_1 z_t + \varepsilon_t \tag{3.3}$$

where: $\delta_0 = -\delta_1 \equiv b_1$; $z_t = \exp(-\hat{\beta}^k x_t)$

The parameter constraint corresponding to (3.2d) is simply: $\delta_0 + \delta_1 = 0$.

Copyright © 1997-2008: Ike's Research Ltd

For Example 5 and Example 6, EAGLE's estimates of the model parameters almost coincide with the NIST values; however, the residual variance is smaller in both cases, and this suggests that the EAGLE parameters may be closer to the true values.

Table 3.1: The Reliability of Statistical Packages (**Source**: Tvrdík and Křivý, 2004)

	EAGLE	NCSS 2001	S-PLUS 4.5	SPSS 10.0	SYSTAT 8.0 [Simplex]	SYSTAT 8.0 [Gauss-Newton]
Example 4a	OK	F	OK	F / OK	F	OK
Example 4b	OK	-	-	-	-	-
Example 5	OK	F	F / OK	F	F / OK	OK
Example 6	OK	F	F	F	OK	OK

Legend: OK — success; F — Failure; F / OK — Failure from first starting point; success from second starting point

On Example 7

Example 7 is an artificial problem whose regressor data is random sample from a uniform distribution on the interval [0.5, 4]; a linear model, together with a normally distributed random variable, generated the response data. The sample size was set at 30. The variance of the random errors was deliberately set low in order to discount the possibility of estimating models other than those used to generate the data. The estimated equation for this example is practically the same as the true model and needs no further comment.

On Example 8, 9 and 10

The results for Example 8 through 10 show that the bi-level technique estimates the best equation for any given set of data. Even where the errors are significantly non-NIID, the method manages to "normalise" the error distribution and provide a good fit, as may be seen from the equation statistics for Example 8, 9 and 10. Other studies with comparable error distributions such as Spitzer (1978; 1984) and Showalter (1994) did not achieve a similar quality of fit, despite using considerably larger sample sizes.

In General

Maximum likelihood estimation is usually performed by numerical optimization methods, and finding ML estimators that are statistically efficient is an exercise in *global* optimization (Tvrdík and Křivý, 2004). But, as pointed out earlier (infra, p.2), many existing techniques are not always successful at this task. However, it is increasingly being recognised that evolutionary methods can be efficient estimators because a well-designed Evolutionary Algorithm does not depend on the convexity or smoothness of the likelihood function, and it converges to the global solution, albeit in probability only (Rudolph, 1994). EAGLE has, therefore, a distinct advantage over traditional methods. Because it is an evolutionary approach, it is numerically more stable, and its estimates are relatively more efficient in many difficult cases. The examples presented here support this assertion. Finally, it is important to note that, although EAGLE has been described in terms of the normal distribution, it is not restricted to this situation. The design philosophy allows any distribution for the response variable provided one can define the appropriate likelihood function and the statistical tests required for concurrent compensation. Hence, the term 'general' in the acronym 'EAGLE'.

Copyright © 1997-2008: Ike's Research Ltd

4 Summary and Conclusions

This document has outlined EAGLE. EAGLE is a parameter estimation algorithm that employs the maximum likelihood principle and applies to linear and nonlinear models alike, using two modes of operation. The algorithm's uni-level mode is simply a direct numerical maximization method for the likelihood function: its bi-level mode allows for a decomposition of the estimation task into two interacting search processes: a 'Level 1' algorithm that drives the overall search and compensates for any regression "anomalies" that might arise during the solution process, and a 'Level 2' linear regression algorithm that computes OLS estimates, given data from the former.

The design of the algorithm is founded on three concepts: (a) model linearization and problem decomposition via a partitioning of the parameter vector; (b) on-line regression diagnostic testing and control of regression residuals via multi-objective optimization — a process that is here called 'concurrent compensation'; (c) exploitation of the location and scale invariance properties of linear regression. The idea of linearizing a nonlinear model by a transformation of the variables is certainly not original to EAGLE. In fact, Draper and Smith (1981, p.459) point out that most authors use the words 'intrinsically linear' in this sense. However, as of this writing, I am unaware of any other regression technique that employs regression diagnostic tests in the manner described here. The method is easy to implement and builds on the established methodology of linear regression about which there is extensive knowledge and experience.

The algorithm's uni-level mode is a bi-objective evolutionary design that directly maximises the likelihood function and minimises the residual variance. But unlike some existing maximum likelihood algorithms, about which Harvey (1990) writes:

> "Most non-linear routines are subject to problems of execution, while the results are liable to misinterpretation. A good deal of care is therefore required" (p.122)

EAGLE is a numerically stable global optimizer: none of the numerical problems that are typically associated with traditional optimization techniques arise in the case of EAGLE because it is an evolutionary approach that does not require the information matrix, or any its approximations.

Both modes of operation allow for functional constraints of the form $g(\mathbf{\Phi}_1) \leq 0$ on the nonlinear parameters. In addition, the criterion vector $\mathbf{J}$ of the bi-level scheme may be augmented by a component that restricts the OLS parameters to a particular region of the search space.

Several types of numerical examples were estimated using the bi-level algorithm or its uni-level variant, and the results show the procedure to be very effective. The bi-level algorithm is statistically robust — it gives a good fit even where errors are non-normal.

Copyright © 1997-2008: Ike's Research Ltd

REFERENCES

BAŞAR, T. and G. J. Olsder (1999). *Dynamic Non-cooperative Game Theory*. SIAM Classics in Applied Mathematics, **23**, SIAM, Philadelphia

BOX, G. E. P. and D. R. Cox (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society*, Series B, **26**, 211-243.

BOX, G. E. P. and P. W. Tidwell (1962). Transformation of the independent variables. *Technometrics*, **4**, 531-550.

COELLO COELLO, C. A. (1996). *An Empirical Study of Evolutionary Techniques for Multi-objective Optimization in Engineering Design*. Unpublished PhD Dissertation, Tulane University, New Orleans, Louisiana, U.S.A.

COELLO COELLO, C. A. (2006). 20 Years of Evolutionary Multi-Objective Optimization: What Has Been Done and What Remains to be Done. In Gary Y. Yen and David B. Fogel (eds.), *Computational Intelligence: Principles and Practice*, Chapter 4, pp. 73--88, IEEE Computational Intelligence Society.

COURNOT, A. A. (1838): Recherches sur les Principes Mathématiques de la Théorie des Richesse, M. Riviere, Paris. Translated in Kelly, A. M., *Researches into the Mathematical Principles of Wealth*, New York, 1960.

DAGENAIS, M. G. and J. Dufour (1994). Pitfalls of Re-scaling Regression Models with Box-Cox Transformations. *The Review of Economics and Statistics*, **76**, 570-575.

DRAPER N. R. and H. Smith (1981). *Applied Regression Analysis*. New York: John Wiley.

EDGEWORTH, F. Y. (1881). *Mathematical Psychics*. P. Keagan, London, England.

HARVEY, A. (1990). *The Econometric Analysis of Time Series*. London: Phillip Allen.

LARDARO, L. (1993). *Applied Econometrics*. New York: Harper Collins Publishers.

NASH, J. (1951). Non co-operative Games. *Annals of Mathematics*, **54**, 287-295.

PARETO, V. (1896). *Cours D'Economie Politique*, Volume I and II. F. Rouge, Lausanne

RAMSEY, J. B. (1969). Tests for Specification Errors in Classical Linear Least-squares Regression Analysis. *Journal of the Royal Statistical Society*, Series B, **31**, 350-371.

RAWLINGS, J. O. (1988). *Applied Regression Analysis: A Research Tool*. Pacific Cole: Wadsworth and Brooks.

RUDOLPH, G. (1994). "Convergence Properties of Canonical Genetic Algorithms", *IEEE Trans. on Neural Networks*, **5**, pp. 96-101.

SHAPIRO, S. S. and M. B. Wilk (1965). An Analysis of Variance Test for Normality (Complete Samples), *Biometrika*, **52**, 591-611.

SHAPLEY, L. S. (1959). Equilibrium Points in Games with Vector Payoffs, *Naval Research Logistics Quarterly*, **1**, pp.57-61

SHOWALTER, M. H. (1994). A Monte Carlo Investigation of the Box-Cox Model and a Non-linear Least Squares Alternative. *The Review of Economics and Statistics*, 561-570.

SIWALE, I. (2002). EAGLE: An Evolutionary Algorithm for General Likelihood Estimation. *Technical Report No. RD-2-2002*, Ike's Research Ltd, London.

SPITZER, J. J. (1978). A Fast and Efficient Algorithm for the Estimation of Parameters in Models with the Box-Cox Transformation. *Journal of the American Statistical Association*, Vol. 77, 760-765.

SPITZER, J. J. (1982). Variance Estimates in Models with the Box-Cox Transformation: Implications for Estimation and Hypothesis Testing. *The Review of Economics and Statistics*, **66**, 645-652.

SPITZER, J. J. (1984). A Primer on Box-Cox Estimation. *The Review of Economics and Statistics*, **64**, 645-652.

Copyright © 1997-2008: Ike's Research Ltd

STATISTICAL REFERENCE DATASETS. Information Technology Laboratory, National Institute of Standards and Technology. [online]. Available from: http://www.itl.nist.gov/div898/strd/

THEIL, H. (1968). A Simplification of the BLUS Procedure for Analysing Regression Disturbances. *Journal of the American Statistical Association*, **63**, 242-251.

THISTED, R. A. (1980). Comment on paper by Smith and Campbell in *Journal of the American Statistical Association*, **75**, 74-81.

TVRDÍK, J. and L. Křivý (2004). Comparison of Algorithms for Nonlinear Regression Estimates. In Antoch, J (ed.), COMPSTAT 2004. *Physica-Verlag*, 1917-1924.

WEISBERG, S. (1974). An Empirical Comparison of the Percentage Points of *W* and *W'* . *Biometrika*, **61**, 644-646.

WEISBERG, S. (1980). *Applied Regression Analysis*. New York: John Wiley.

WILLAN A. R. and D. G. Watts (1978). Meaningful Multi-collinearity Measures. *Technometrics*, **20**, 407-412.

ZAREMBKA, P. (1968). Functional Form in the Demand for Money. *Journal of the American Statistical Association*, **63**, 502-511.

ZAREMBKA, P. (1974). Transformation of variables in Econometrics. In Zarembka, P. (ed.), *Frontiers in Econometrics*. New York: Academic Press, 81-104.

Copyright © 1997-2008: Ike's Research Ltd