Investigating Nonlinearity: A Note on the Implementation of Hamilton's Methodology

by

D. Bond, M. Harrison, and E.J. O'Brien*

University of Ulster and Trinity College Dublin

25th October 2003

Abstract

(J.E.L. Classification C13, C51, C61)

In this paper we give an account of the new approach to nonlinear econometric modelling proposed by Hamilton (2001) and briefly describe some of the methods of nonlinear optimization that may be used in the *Gauss* computer program provided by Hamilton for the implementation of his methodology. The performance of this program is investigated using data relating to Hamilton's example concerning the US Phillips curve, two versions of the *Gauss* software and a range of alternative numerical optimization options and values for the important *Gauss* parameter **_oprteps**. Finally, the effects of changes in the sample data on the results produced by Hamilton's procedure are explored. The results presented suggest some clear conclusions, which will be of value to those contemplating working with Hamilton's new method.

Keywords: nonlinearity, numerical optimisation, US Phillips curve.

* E.J. O'Brien is a Government of Ireland Scholar and wishes to thank the Irish Council for the Humanities and Social Sciences for its generous funding.

Investigating Nonlinearity: A Note on the Implementation of Hamilton's Methodology

by

D. Bond, M. Harrison and E. O'Brien

Introduction

In an important recent paper, Hamilton (2001) proposed a new approach to nonlinear modelling of economic relationships that provides a single flexible parametric framework for testing for nonlinearity, drawing inference about the form of nonlinearity, and assessing the adequacy of the description of nonlinearity provided by specific models. Following Wecker and Ansley (1983), the approach treats functional form as the outcome of a latent stochastic process that is part of the data-generating process, i.e. the conditional expectation function associated with a regression model is thought of as being generated randomly prior to the generation of the data. This latent process is modelled using a new Gaussian random field concept that generalizes Brownian motion to k dimensions, and the parameters of the process are estimated by maximum likelihood. The method is a good deal more than an exploratory data or data-smoothing device. From the practicing economist's viewpoint, its importance lies in the valuable insights it can provide for model construction and the resulting enhancement of the forecasting ability of economic models.

However, the new methodology has been little used to date and its full potential remains to be established. As Hamilton (2001, p. 552) points out, its usefulness for particular sample sizes and nonlinearities is a matter for empirical investigation. Yet, citing his own three examples and the Monte Carlo studies by Dahl (1998), he suggests that the method holds much promise. We very much agree.

The aims of the present paper are modest. The main purpose is to address a number of practical issues that arise when using the Hamilton approach. The first of these concerns computation: we report on our experience with Hamilton's software to implement the method.[1] It appears that the numerical optimization involved is not an entirely straightforward matter, either when using Hamilton's own dataset or our alternative samples. The second issue concerns the sensitivity of the method to changes in data. Our experiments suggest that minor data changes can have implications for computation and big effects on the results. Another aim, given the length and difficulty of the original paper, is to provide a concise and reasonably accessible account of Hamilton's methodology for non-specialist practitioners, though the nature of the subject is such that it is not possible to avoid technicalities.

The structure of the paper is as follows. Section 2 describes the new approach to modelling nonlinearity; sections 3 and 4 report on computational and data-sensitivity matters, respectively; and section 5 contains a brief summary and conclusion.

[1]The program is written in *Gauss* and can be freely downloaded from http://weber.ucsd.edu/~jhamilto.

2. Hamilton's method of parametric flexible nonlinear inference [2]

The model

Our interest is in the nonlinear regression denoted by

$$y_t = \mu(\mathbf{x}_t) + \varepsilon_t, \qquad t = 1, 2, ..., T \tag{1}$$

where y_t is scalar and $\mathbf{x}_t = [x_{it}]$ is a $k \times 1$ vector of observations on the explanatory variables at time t, ε_t is a stochastic disturbance with zero mean and constant variance, independent of lagged values of $\mathbf{x}_t$ and y_t, and $\mu(\mathbf{x})$ denotes the conditional expectation function $E(y \mid \mathbf{x})$. The nature of $\mu(\mathbf{x})$ is fundamental to Hamilton's approach and is considered to be determined as

$$\mu(\mathbf{x}) = \alpha_0 + \boldsymbol{\alpha}'\mathbf{x} + \lambda m(\mathbf{g} \odot \mathbf{x}) \tag{2}$$

where α_0 and λ are scalar parameters, $\boldsymbol{\alpha} = [\alpha_i]$ and $\mathbf{g} = [g_i]$ are $k \times 1$ vectors of parameters, $m(\cdot)$ is a realization of a stochastic process called a random field, and $\odot$ denotes element-by-element multiplication. The realization $m(\cdot)$, hence $\mu(\mathbf{x})$, is assumed to be generated by nature prior to and independently of all of the observations. Given this fixed $\mu(\mathbf{x})$, the values for ε_t and $\mathbf{x}_t$ are then generated and y_t, is determined according to the regression (1).

The interpretation of the parameters in (2) is vitally important for the application of Hamilton's method. In particular, the scalars λ and g_i, $i = 1, 2, ..., k$, characterize the relationship between $m(\cdot)$ and the conditional expectation function $\mu(x_1, x_2, ..., x_k)$.

[2] To facilitate cross-reference to the original paper, the notation used in this section is similar to that used by Hamilton (2001).

Specifically, λ is a measure of the overall 'weight' of the process $m(\cdot)$ in the conditional expectation, while the magnitudes of the g_i indicate the degree of nonlinearity associated with their respective x_i. Thus $\lambda = 0$ indicates that $m(\cdot)$ makes no contribution and the conditional expectation is linear, in which case (1) is the familiar general linear model. Similarly, $g_i = 0$ implies that the conditional expectation is linear in x_i, while $g_i \neq 0$ signifies that it is nonlinear in x_i. If all of the $g_i \to 0$, the contribution of $m(\cdot)$ to the conditional expectation, hence to y_t, becomes indistinguishable from that of α_0; if all of the $g_i \to \infty$, the contribution to y_t is indistinguishable from that of ε_t. The interpretation associated with α_0 and the α_i is the standard one.

The key component in (2), on which the interpretation of the g_i depends, is the random realization $m(\cdot)$, whose nature and role require explanation before the practical matters of estimation and testing are considered. First, consider a uniform orthogonal grid in $\Re^k$, bounded in the direction of each of the k standard basis vectors or Cartesian co-ordinates by some lower value a_j and some upper value b_j, $j = 1, 2, ..., k$.[3] Let the set of all nodes in the grid be A_N, where $N-1$ is the number of grid intervals in each direction and N^k is therefore the number of distinct points in A_N. For each point $\mathbf{x} \in A_N$, let $e(\mathbf{x}) \sim \mathbb{N}(0,1)$ and be independent of $e(\mathbf{z})$ for all $\mathbf{x} \neq \mathbf{z}$; let $B_N(\mathbf{x}) = \{\mathbf{z} \in A_N : (\mathbf{x}-\mathbf{z})'(\mathbf{x}-\mathbf{z}) \leq 1\}$, i.e. the set of all points in A_N whose distance from $\mathbf{x}$ is less than or equal to unity; and let $n_N(\mathbf{x})$ denote the number of points in $B_N(\mathbf{x})$. Hamilton (2001, p.540) then defines the scalar

[3] By *uniform* we mean that the intervals defined by the grid are of equal length in the direction of each of the k co-ordinates, and the number of intervals in each direction is the same. Note that this does not imply that the intervals in different directions have to be the same length unless the a_j are equal and the b_j are equal $\forall\, j$.

process $m_N(\mathbf{x})$ as[4]

$$m_N(\mathbf{x}) = [n_N(\mathbf{x})]^{-\frac{1}{2}} \sum_{\mathbf{z} \in B_N(\mathbf{x})} e(\mathbf{z}). \tag{3}$$

Taking the limit of (3) as the grid partition becomes finer, i.e. the interval length in each direction of the grid tends to zero, we have the notion of a continuous scalar-valued k-dimensional random field. The stochastic nature of this is such that for any $\mathbf{x} \in A_N$, $m(\mathbf{x}) \sim \mathbb{N}(0,1)$. The similarity to standard Brownian motion is apparent.

For arbitrary points in $\Re^k$, say, $\mathbf{x}$ and $\mathbf{z}$, the correlation between $m(\mathbf{x})$ and $m(\mathbf{z})$ is zero if the distance between $\mathbf{x}$ and $\mathbf{z}$ is greater than 2. If this distance is not greater than 2, it can be shown, though the proofs are difficult,[5] that

$$H_k(h) = Cov_k\left(m(\mathbf{x}), m(\mathbf{z})\right) = \frac{G_{k-1}(h,1)}{G_{k-1}(0,1)} \tag{4}$$

where $G_{k-1}(h,1) = -\frac{h}{k}(1-h)^{\frac{k-1}{2}} + \frac{k-1}{k}G_{k-3}(h,1)$, h is one-half the distance between $\mathbf{x}$ and $\mathbf{z}$, $k = 2, 3, ...$ and the initial values are $G_0(h,1) = 1-h$ and $G_1(h,1) = \frac{\pi}{4} - \frac{1}{2}h(1-h^2)^{\frac{1}{2}} - \frac{1}{2}\sin(h)$. Thus (4) can be calculated recursively, but fortunately its values for $k=1$ to 5 inclusive are provided in Table I of Hamilton (2001, p. 542). It is this covariance that provides the means by which the g_i govern the curvature of $\mu(\mathbf{x})$ in (2); see the illustrative case of $k=1$ in Hamilton (2001, p. 540).

Estimation

Assuming normality of the ε_t, it follows from (1), (2) and (4) that

$$\mathbf{y} \sim \mathbb{N}(\mathbf{X}\boldsymbol{\beta}, \mathbf{C} + \sigma^2 \mathbf{I}_T) \tag{5}$$

[4]This processs is illustrated for $k=2$, $a_1 = a_2 = 0$, $b_1 = 5$, $b_2 = 3$, and equal interval lengths in Hamilton (2001, p.541), so that the number of intervals in each direction is not the same, as required by the definition of A_N.

[5]See Lemma 2.1 and Theorem 2.2 in Hamilton (2001, p. 541). Note also that the details relating to Equation (4) are expressed slightly differently than in Hamilton's lemma and theorem.

where $\mathbf{y}$ is the $T\times 1$ vector of observations on the dependent variable in (1), $\mathbf{X}$ is the $T\times(k+1)$ matrix of observations on the k explanatory variables and a column of ones associated with the intercept, $\boldsymbol{\beta} = [\begin{array}{cc} \alpha_0 & \boldsymbol{\alpha}' \end{array}]'$ is the $(k+1)\times 1$ vector of parameters of the linear component of the conditional expectation, $\mathbf{C} = [\lambda^2 H_k(h_{ts})]$ is a $T \times T$ variance-covariance matrix whose typical element is $\lambda^2 Cov_k\left(m(\mathbf{g}\odot\mathbf{x}_t), m(\mathbf{g}\odot\mathbf{x}_s)\right)$, and h_{ts} is one-half the distance between $\mathbf{g}\odot\mathbf{x}_t$ and $\mathbf{g}\odot\mathbf{x}_s$. The likelihood function follows straightforwardly from (5) as

$$\ln f(\mathbf{y};\boldsymbol{\beta},\mathbf{g},\lambda,\sigma^2) = -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{C}+\sigma^2\mathbf{I}_T| - \frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'(\mathbf{C}+\sigma^2\mathbf{I}_T)^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}). \quad (6)$$

Maximum likelihood provides the basis for inference concerning the parameters $\boldsymbol{\beta}$, $\mathbf{g}$, λ and σ^2; and as Hamilton shows, the procedure is valid for regressors that are deterministic or lagged values of the dependent variable. However, in the interests of simplifying the calculations, (6) is re-written. Defining $\zeta = \frac{\lambda}{\sigma}$, letting $\boldsymbol{\psi} = [\begin{array}{cc} \boldsymbol{\beta}' & \sigma^2 \end{array}]'$ be the $(k+2)\times 1$ vector of parameters for the linear part of the model and $\boldsymbol{\theta} = [\begin{array}{cc} \mathbf{g}' & \zeta \end{array}]'$ be the $(k+1)\times 1$ vector of parameters of the nonlinear component, and setting $\mathbf{W}(\mathbf{X};\boldsymbol{\theta}) = \zeta^2\mathbf{C}^* + \mathbf{I}_T$, where $\mathbf{C}^* = \lambda^{-2}\mathbf{C}$, the right-hand side of (6) can be written as

$$-\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\sigma^2 - \frac{1}{2}\ln|\mathbf{W}(\mathbf{X};\boldsymbol{\theta})| - \frac{1}{2\sigma^2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})'\mathbf{W}(\mathbf{X};\boldsymbol{\theta})^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}). \quad (7)$$

The values of the elements of $\boldsymbol{\psi}$ that maximize (7) for given $\boldsymbol{\theta}$ can then be calculated analytically as

$$\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta}) = [\mathbf{X}'\mathbf{W}(\mathbf{X};\boldsymbol{\theta})^{-1}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{W}(\mathbf{X};\boldsymbol{\theta})^{-1}\mathbf{y}, \quad (8)$$

$$\widetilde{\sigma}^2(\boldsymbol{\theta}) = \frac{1}{T}[\mathbf{y}-\mathbf{X}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta})]'\mathbf{W}(\mathbf{X};\boldsymbol{\theta})^{-1}[\mathbf{y}-\mathbf{X}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\theta})]. \quad (9)$$

Thus (7) may be concentrated as

$$\phi(\boldsymbol{\theta};\mathbf{y},\mathbf{X}) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln\widetilde{\sigma}^2(\boldsymbol{\theta}) - \frac{1}{2}\ln|\mathbf{W}(\mathbf{X};\boldsymbol{\theta})| - \frac{T}{2}. \tag{10}$$

Hence numerical maximization of (10) gives the maximum likelihood estimate of $\boldsymbol{\theta}$, which via (8) and (9) yields the estimate of $\boldsymbol{\psi}$.

Testing for nonlinearity

The form of the model used in Hamilton's approach suggests that a simple method of testing for nonlinearity is to check if λ, or λ^2, is zero or not. Hamilton shows that if $\lambda^2 = 0$, and the nonlinear model is estimated, then for fixed $\mathbf{g}$, the maximum likelihood estimator $\widetilde{\lambda}^2$ is consistent for the true value of zero and asymptotically normal. Thus a test based on the use of standard normal tables is suggested. However, given the maximum likelihood approach to estimation and the linearity of (1) under the null hypothesis that $\lambda^2 = 0$, an obvious and perhaps more appealing way of testing is to use the Lagrange multiplier principle, which requires only a simple linear regression to be estimated. Under the assumption of normality, Hamilton derives the appropriate score vector of first derivatives and the associated information matrix and proposes a form of LM test for practical application. The procedure is as follows.

- Set $g_i = \frac{2}{\sqrt{ks_i^2}}$, where s_i^2 is the variance of explanatory variable x_i, excluding the constant term whose variance is zero.
- Calculate the $T \times T$ matrix, $\mathbf{H}$, whose typical element is $H_k\left(\frac{1}{2}\|\mathbf{g}\odot\mathbf{x}_t - \mathbf{g}\odot\mathbf{x}_s\|\right)$, i.e. the function $H_k(h_{ts})$ defined in (4) and (5).

- Use OLS to estimate the standard linear regression $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ and obtain the usual residuals, $\widehat{\boldsymbol{\epsilon}}$, and standard error of estimate, $\widehat{\sigma} = (T-k-1)^{-\frac{1}{2}}\sqrt{\widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}}}$.

- Finally, compute the statistic

$$\varkappa^2 = \frac{\left[\widehat{\boldsymbol{\epsilon}}'\mathbf{H}\widehat{\boldsymbol{\epsilon}} - \widehat{\sigma}^2 tr(\mathbf{MHM})\right]^2}{\widehat{\sigma}^4\left\{2tr\left(\left[\mathbf{MHM} - (T-k-1)^{-1}\mathbf{M}tr(\mathbf{MHM})\right]^2\right)\right\}}, \qquad (11)$$

where $\mathbf{M} = \mathbf{I}_T - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the familiar symmetric idempotent matrix.

As $\varkappa^2 \overset{A}{\sim} \chi_1^2$ under the null hypothesis, linearity ($\lambda^2 = 0$) would be rejected if $\varkappa^2$ exceeded the critical value, $\chi^2_{1,\alpha}$, for the chosen level of significance, α. Otherwise the null of linearity would not be rejected. For example, at the 5 per cent significance level, the null would be rejected if $\varkappa^2 > 3.84$. In this case the alternative nonlinear specification given by (1) and (2) would be preferred.[6]

3. Computational issues

The implementation of Hamilton's methodology is straightforward, in principle, given the availability of Hamilton's software (see Footnote 1 above). In practice, however, difficulties may await the unwary. These difficulties relate to the nonlinear optimization algorithms in the OPTMUM procedure in *Gauss,* which is at the heart of Hamilton's program. Indeed, in our original attempts to run the program using *Gauss 5*, the optimization procedure failed completely and no nonlinear estimates were obtained. It was this experience that motivated the research for the present paper.

[6] The identification of a specific form of nonlinearity is greatly aided by the estimate of the conditional expectation $\mu(\mathbf{x}_t)$ and, specifically, the $\widetilde{\zeta}$ and $\widetilde{g}_i$. The matter is explained in Hamilton (2001, Section 5) and illustrated in the three examples in his Section 7; note the role of his Equation (5.17). However, this is not pursued in the present paper, which concentrates on the nonlinear estimation *per se.*

The methods of nonlinear optimization are familiar to most econometricians, and the particular algorithms available in *Gauss* may be familiar to some *Gauss* users. However, most economists probably do not share this familiarity. Therefore, to assist with the understanding of the material in this and the following section, a brief description of numerical optimization and the relevant algorithms is provided in the following subsection.[7]

Nonlinear optimization

The OPTMUM procedure in *Gauss* maximizes the function (10), i.e. $\phi(\boldsymbol{\theta}; \mathbf{y}, \mathbf{X})$, by minimizing the negative of the function with respect to its vector of parameters, $\boldsymbol{\theta}$. Given the derivatives of this objective function with respect to $\boldsymbol{\theta}$, i.e. the gradient vector, which it computes numerically, and initial values for $\boldsymbol{\theta}$, OPTMUM proceeds iteratively, computing a *direction*, $\mathbf{d}$, and a *step length*, s, at each iteration. The quantity $s\mathbf{d}$ is a vector of values that is added to the current estimate of $\boldsymbol{\theta}$, and therefore has the same dimension as $\boldsymbol{\theta}$, and s is a scalar. Thus, given a value for $\mathbf{d}$, the current estimate, $\widetilde{\boldsymbol{\theta}}$, is updated as

$$\widetilde{\boldsymbol{\theta}}_{+} = \widetilde{\boldsymbol{\theta}} + s\mathbf{d}; \tag{12}$$

hence s may be interpreted as changing the rate of descent of the objective function in the given direction. We will describe in turn how $\mathbf{d}$ and s are computed, concentrating on the former.

Defining $\mathbb{G}$ to be the $(k+1) \times 1$ gradient vector and $\mathbb{H}$ to be a $(k+1) \times (k+1)$ symmetric matrix, a standard method of calculating $\mathbf{d}$ is as

$$\mathbf{d} = \mathbb{H}^{-1}\mathbb{G}. \tag{13}$$

[7] Further details on numerical optimization may be found in the texts by Brent (1973), Greene (2003), Murray (1972) and the *Gauss* reference manual *Optimization,* Aptech Systems, Inc., 2001, especially chapters 2 and 3.

However, as numerical matrix inversion may be a risky process, the OPTMUM procedure can avoid it by computing $\mathbf{d}$ as the solution of the equation

$$\mathbb{H}\mathbf{d} = \mathbb{G}, \tag{14}$$

which is thought to be numerically more reliable. While $\mathbb{G}$ is calculated in a standard manner, $\mathbb{H}$ may be calculated in different ways depending on which algorithm is selected. Several approaches are available in OPTMUM.

The *steepest descent* algorithm simply sets $\mathbb{H} = \mathbf{I}_{(k+1)}$. While this is computationally undemanding and therefore attractive, the descent may be slow and require many iterations before convergence.

The PRCG or Polak and Ribiere (1969) *conjugate gradient* method is a development of the steepest descent method that also uses only the gradient but updates the direction as

$$\mathbf{d}_+ = \mathbb{G}_+ + r\mathbf{d}, \text{ where } r = \frac{(\mathbb{G}_+ - \mathbb{G})'\mathbb{G}_+}{\mathbb{G}'\mathbb{G}}. \tag{15}$$

There are several more complex methods. The *Newton* algorithm equates $\mathbb{H}$ to the Hessian of the objective function, which may be computed numerically as the gradient of the gradient. Unfortunately, this computation is generally a formidable numerical problem and, as it is required at each iteration, makes the algorithm slow and possibly unreliable. However, when it works smoothly, the Newton algorithm may converge in fewer iterations than other methods.

The large computational problems associated with the calculation of the Hessian in the Newton method are avoided by certain so-called *quasi-Newton* algorithms. These start with an initial estimate of the Hessian and employ updates that add information at each iteration without requiring the calculation of second derivatives. Although they generally need more iterations to achieve convergence than the Newton method, their numerical efficiency means that they are usually faster and, furthermore, tend to be more robust to the condition of the model and data. The OPTMUM procedure contains three such algorithms: the BFGS method due to Broyden (1967), Fletcher (1970), Goldfarb (1970) and Shanno (1970), the DFP method of Davidon (1968) and Fletcher and Powell (1963), and BFGS-SC, which is a modified BFGS algorithm in which the formula for the computation of the update of the Hessian estimate has been changed to make it scale free. In all three cases, the OPTMUM implementation of the algorithm uses the Cholesky factorization of the approximation to the Hessian in (14), i.e. $\mathbb{H} = \mathbb{C}'\mathbb{C}$, before solution for $\mathbf{d}$.

The BFGS algorithm is the default choice in OPTMUM, while the other five are available as options.

The OPTMUM procedure in *Gauss 5* also includes a number of methods for computing the step length, s. The default method is called STEPBT, which is described in Dennis and Schnabel (1983). It first attempts to fit a quadratic to the objective function and computes an estimate of s that minimizes the quadratic. If that fails, it tries a cubic function, which is rather more versatile in cases where the objective function is not well approximated by a quadratic.

If STEPBT fails, then BRENT is used, a technique due to Brent (1972) that evaluates the objective function at a sequence of test values for s, determined by extrapolation and interpolation using the inverse of the "golden ratio", namely, the constant $\frac{(\sqrt{5}-1)}{2} = 0.61803$. This method is generally more efficient than STEPBT but requires significantly more function evaluations.

If, in turn, BRENT fails, then a procedure called HALF is used. Denoting the objective function by $F(\widetilde{\boldsymbol{\theta}} + s\mathbf{d})$, this method first sets $s = 1$. If $F(\widetilde{\boldsymbol{\theta}} + s\mathbf{d}) < F(\widetilde{\boldsymbol{\theta}})$, then s is set to 1; if not, then $s = 0.5$ and $F(\widetilde{\boldsymbol{\theta}} + s\mathbf{d})$ is tried. The attempted step length is halved each time the objective function fails to decrease. When the function does decrease, s is set to its current value. This method usually requires the fewest function evaluations but is most likely to fail to find the step length that decreases the objective function.

Thus, if HALF fails, a final search for a random direction that decreases the objective function is implemented. The radius of the random search is fixed via an important global variable in OPTMUM called **_oprteps**, the default value of which is 0.01. It is, however, possible to specify any positive value for **_oprteps**.

Computations and results

The computations and results in this subsection and the following section relate to Hamilton's Example 3 concerning the post-war US Phillips curve.[8] Noting that an OLS regression of inflation (π_t) on unemployment (u_t), lagged inflation (π_{t-1}) and a time trend (t) reveals statistically insignificant evidence of an inflation-unemployment trade-off using annual data for the period $t = 1949$ to 1997, Hamilton (2001, Section 7) investigates whether a nonlinear relation like that defined in (1) and (2), of the specific form

$$\pi_t = \mu(u_t, \pi_{t-1}, t) + \varepsilon_t, \tag{16}$$

might be an improvement. His results appear on page 563 of his paper. However, as

[8] No results for the test statistic (11) are given as they derive from a simple ordinary least squares regression, which is unproblematical. However, they were checked for all of the cases considered in sections 3 and 4 and, without exception, the null of linearity was rejected.

previously mentioned, we were originally unable to reproduce Hamilton's results using *Gauss 5* and Hamilton's data: the numerical optimization associated with the maximization of (10) failed.

Examination of Hamilton's program revealed that it employs the BFGS algorithm and relies on the default value of **_oprteps.** It also appeared that it had been implemented by Hamilton using an earlier version of *Gauss.* Indeed, we were able to reproduce his results using *Gauss* 3, for example, although where he reports a value for $\widetilde{g}_2$ of 0.16, we find a value of -0.16, when rounded to two decimal places like all of his results.[9] However, when different algorithms were specified under *Gauss* 3 the results, when they were produced using Hamilton's data file, were not always similar to those reported by Hamilton, as shown in Table 1. In this table, algorithms 1, 2, 3, 4, 5 and 6 refer to the steepest descent, BFGS, BFGS-SC, DFP, Newton and PRCG methods, respectively; and the g_i and α_i refer to the parameters in the nonlinear and the linear components of the conditional expectation function, respectively. The values of $i = 1, 2, 3$ relate to u_t, π_{t-1} and t, respectively, while α_0 is the constant.

The results for BFGS (algorithm 2) in Table 1 are those corresponding to Hamilton's. Apart from the one difference in sign for $\widetilde{g}_2$, they are identical to his. However, we find that BFGS-SC (algorithm 3) and PRCG (algorithm 6) fail for Hamilton's dataset, steepest descent (algorithm 1) produces noticeably different numerical results from Hamilton's, Newton (algorithm 5) produces very similar results except for the sign on the nonlinear parameter estimate $\widetilde{g}_2$, and DFP (algorithm 4) replicates the results of BFGS, the Hamilton case. Despite the big numerical differences in the results produced using algorithm 1, the

[9]This difference in sign may be a typographical error in Hamilton's paper.

high statistical significance of $\widetilde{g}_3$ remains and the inference concerning nonlinearity would be basically the same as that drawn by Hamilton.

Following an amount of experimentation, we eventually reproduced Hamilton's results using *Gauss 5,* though not all values of **_oprteps** proved successful and led to results. Table 2 contains the results for what was perhaps the most successful value for this parameter, namely, **_oprteps** = 0.00001, while Table A1 in Appendix A also gives results for certain other **_oprteps** values, i.e. 0.001, 0.1 and 1.0. As can be see from Table 2, the results from *Gauss 5,* algorithm 2, are identical to those produced by the same algorithm in *Gauss 3,* and confirm the negative sign on $\widetilde{g}_2$. The results from *Gauss 5,* algorithms 1 and 5 are similar in absolute terms to those given by algorithm 2 but there are some sign changes on $\widetilde{g}_2$ and $\widetilde{g}_3$. In contrast to what was found using *Gauss 3,* there are surprisingly large numerical differences between the results from algorithms 4 and 6 and those from algorithm 2 when using *Gauss 5.* Despite these various changes across some algorithms and the two versions of *Gauss,* $\widetilde{g}_3$ remains the most statistically significant of the nonlinear parameter estimates, though $\widetilde{g}_2$ is marginally significant for most of the algorithms and **_oprteps** values. Algorithm 3 failed in all experiments due to a problem with the Cholesky decomposition,[10] and we conclude that this may be due to a program error in *Gauss,* which remains to be investigated. The numbers of iterations used by the alternative algorithms are, in relative terms, broadly in line with what was said about relative efficiencies in section 3.

[10]The *Gauss* diagnostic message produced was "Cholesky downdate failed".

4. Sensitivity to data

This section reports on the performance of the program and *Gauss* algorithms, and the results produced, when various small changes are made to the dataset used in Hamilton's Example 3. Three kinds of change were considered. The first deleted observations at the start of the dataset and the second deleted observations at the end to give successively smaller samples. The third added new observations to create successively larger, more up-to-date samples. The additional US unemployment data were obtained from http://www.bls.gov/cps/cpsaat1.pdf and the new values for the US consumer price index from ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt. Checks confirmed that the observations for the period 1949 to 1997, also available from these websites, were identical to those in Hamilton's dataset.

In all, ten alternative samples were used. Hamilton's original dataset is called dataset 1. Deleting the first observation from Hamilton's data gives dataset 2; deleting the first and second observations gives dataset 3; deleting the first, second and third observations gives dataset 4. Similarly, deleting the last observation gives dataset 5; deleting the last two observations gives dataset 6; deleting the last three observations gives dataset 7. Finally, adding the observation for 1998 gives dataset 8; adding the two observations for 1998 and 1999 gives dataset 9; adding the three observations for 1998, 1999 and 2000 gives dataset 10; and adding the four observations for 1998 to 2001, inclusive, gives dataset 11.

For each of the 10 alternative samples, Hamilton's program was implemented using *Gauss 5* and the values for **_oprteps** of 0.00001, 0.001, 0.1 and 1.0, which were referred to in the previous section and appear in Table A1. The *Gauss 3* implementation was also used with datasets 2, 5, 7, 8 and 11. A large volume of results was therefore produced and the relevant details are tabulated in Appendix B. Table 3 summarizes the nonlinear estimates given by the *Gauss 3* implementation of the program using dataset 2. These results are typical in regard to the incidence of failure of nonlinear optimization algorithms and of the differences in the results produced by different algorithms that did not fail. The results from *Gauss 3* for data sets 5, 7, 8 and 11 are contained in tables B11, B12, B13 and B14, respectively, and the level of program failure in each case is at least as great as that observed using dataset 2.

As can be seen from Table 3, algorithm 2, which is Hamilton's default method, as well as algorithms 3 and 4, fail in *Gauss 3.* The reason is that, after one or several iterations, the algorithm encountered a non-positive definite matrix.[11] Of the methods that worked, algorithm 1 and 6 give similar results but algorithm 5 gives very different results from these, including different signs for all of the $\widetilde{g}_i$ coefficients. These differences are noteworthy, as is the fact that algorithm 2 fails for all of the modified datasets examined.

The details concerning the *Gauss 5* implementation using datasets 2 to 11, inclusive, are contained in tables B1 to B10, respectively. For convenience, the information from these tables on success and failure of the algorithms is summarized in Table 4, along with similar information for Hamilton's data (dataset 1). Of the 264 program runs, 102 or 39 per cent failed to produce nonlinear estimates. At the extremes, algorithm 3 (BFGS-SC) failed in all cases, while algorithm 1 (steepest descent) was successful in all cases, albeit requiring the maximum number of iterations permitted.[12] Algorithm 5 (Newton) was the

[11] One or other of two *Gauss* diagnostic messages were obtained in this event. The first was "Negative of Hessian is not positive definite"; the second was "Matrix not positive definite".

[12] Increasing the maximum number of iterations to 250, for those algorithms that reached the original

most efficient in terms of number of iterations but it failed in 10 out of 44 runs, i.e. in 23 per cent of cases. Algorithm 2 (BFGS), the default in Hamilton's program, failed in 28 out of 44 runs or 64 per cent of cases.

As noted in the case of *Gauss 3,* there are also many differences in the nonlinear estimates obtained from a given data set when different algorithms work under *Gauss 5*, including some sign changes. There are also some big changes in numerical estimates of parameters, including sign changes, when marginal changes in the dataset, such as the addition or deletion of just one observation, are made. The reader's attention is also drawn to the interesting results obtained using dataset 11. However, the relatively high statistical significance of $\widetilde{g}_3$ is generally maintained across the range of experiments we have conducted.

5. Conclusion

In this paper we have given an account of the new approach to nonlinear econometric modelling proposed by Hamilton (2001) and briefly described some of the methods of nonlinear optimization that may be used in the *Gauss* computer program provided by Hamilton for the implementation of his methodology. The performance of this program has been investigated using data relating to Hamilton's example concerning the US Phillips curve, two versions of the *Gauss* software and a range of alternative numerical optimization options and values for the important *Gauss* parameter **_oprteps**. Finally, the effects of changes in the sample data on the results produced by Hamilton's procedure have been explored.

The results we have presented suggest some clear conclusions, which we hope will

maximum of 150, did not alter the results obtained to three places of decimals.

be of value to those contemplating working with Hamilton's new method. First, different algorithms used for the numerical optimization have different chances of success. Hamilton's choice of the BFGS algorithm fails in over 60 per cent of the cases examined in our study, while the less computationally efficient steepest descent method succeeds in all cases. Secondly, when different algorithms work, they may produce significantly different numerical results and different signs for parameter estimates. Thirdly, minor changes in data can have significant effects, both in terms of whether an algorithm operates or not and, in the case of it operating, the numerical results its produces. For example, it is interesting to note that if Hamilton's data had just one less observation at either end of the sample, or one more observation at the end, his version of the program would have failed to produce nonlinear estimates, not only with *Gauss 3* but also with *Gauss 5* and all of the values of **_oprteps** used in this study. Moreover, if his dataset had contained the four additional observations for 1998 to 2001 (dataset 11), while the program would have produced results, all three nonlinear parameter estimates would have been significant, in contrast to just $\widetilde{g}_3$ as found in his original study (dataset 1). Thus his inferences concerning the form of nonlinearity would also have been different. However, finally, despite the sensitivity of results to algorithm and data changes, the statistical significance of the nonlinear parameter estimates, hence the inference about the form of nonlinearity, generally seems to be little affected according to the findings that we have reported.

The present paper is only a beginning of the work advocated by Hamilton as required to establish the usefulness of his new methodology, and our empirical investigations are continuing. Indeed, the approach used above has already been extended to the implementation of Hamilton's program for his Example 1 and Example 2 [Hamilton (2001, Section 7)], though the results are not discussed here. For the interested reader, some of the results are presented in the tables in Appendix C.

References

Brent, Richard P. (1972). "A New Algorithm for Minimizing a Function of Several Variables without Calculating Derivatives"; in Anderssen, R.S., L.S. Jennings and D.M. Ryan. *Optimization*, Brisbane: University of Queensland Press.

Brent, Richard P. (1973). *Algorithms for Minimization without Derivatives,* Englewood Cliffs, N.J.: Prentice-Hall.

Broyden, C.G. (1967). "Quasi-Newton Methods and their Application to Function Minimization", *Mathematics of Computation*, 21, 368-381.

Dahl, Christian M. (1998). "An Investigation of Tests for Linearity and the Accuracy of Flexible Nonlinear Inference", University of Aarhus, Working Paper.

Davidon, W.C. (1968). "Variance Algorithms for Minimization", *The Computer Journal*, 10, 406-410.

Dennis, Jr., J.E. and S.B. Schnabel (1983). *Numerical Methods for Unconstrained Optimization and Nonlinear Equations,* Englewood Cliffs, NJ: Prentice-Hall.

Fletcher, R. (1970). "A New Approach to Variable Metric Algorithms", *The Computer Journal,* 13, 317-322.

Fletcher, R. and M.J.D. Powell (1963). "A Rapidly Convergent Descent Method for Minimization", *The Computer Journal*, 6, 163-168.

Goldfarb, D. (1970). "A Family of Variable-Metric Methods Derived by Variational Means", *Mathematics of Computation*, 24, 23-26.

Greene, William H. (2003). *Econometric Analysis (5^{th}* edn.), Upper Saddle River, N.J.: Pearson Education, Inc.

HAMILTON, JAMES D. (2001). "A Parametric Approach to Flexible Nonlinear Inference", *Econometrica,* 69, 537-573.

MURRAY, W. (ED.) (1972). *Numerical Methods for Unconstrained Optimization,* London: Academic Press.

POLAK, E. AND G. RIBIERE (1969). "Note Sur la Convergence de Method des Direction Conjugees", University of California, Berkeley, Department of Electrical Engineering and Computer Sciences, Working Paper.

SHANNO, D.F. (1970). "Conditioning of Quasi-Newton Methods for Function Minimization", *Mathematics of Computation*, 24, 647-656.

WECKER, WILLIAM E. AND CRAIG F. ANSLEY (1983). "The Signal Extraction Approach to Nonlinear Regression and Spline Smoothing", *Journal of the American Statistical Association,* 78, 81-89.

Table 1: Dataset 1 (Hamilton's US Phillips curve data), *Gauss* 3 results.

Algorithm	Iterations		$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
1	150	Coefficient	0.089	0.127	0.072	-51.903	-0.873	0.406	0.030	1.827	1.431
		Standard Error	0.140	0.071	0.016	153.654	0.431	0.253	0.078	1.019	0.355
2	28	Coefficient	0.142	-0.155	0.136	-88.482	-0.922	0.436	0.049	2.047	1.237
		Standard Error	0.167	0.078	0.032	127.436	0.455	0.232	0.065	1.294	0.441
3	1	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
4	150	Coefficient	0.143	-0.155	0.136	-88.728	-0.919	0.437	0.049	2.036	1.240
		Standard Error	0.168	0.077	0.031	126.952	0.457	0.231	0.064	1.283	0.440
5	17	Coefficient	0.142	0.155	0.136	-88.482	-0.922	0.436	0.049	2.047	1.237
		Standard Error	0.167	0.078	0.032	127..214	0.455	0.232	0.065	1.294	0.441
6	11	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-

Key:
algorithm 1 - Steepest descent
algorithm 2 - BFGS
algorithm 3 - BFGS-SC
algorithm 4 - DFP
algorithm 5 - Newton
algorithm 6 - PRCG

A dash (-) denotes no estimate due to algorithm failure

Table 2: Dataset 1 (Hamilton's US Phillips curve data), *Gauss* 5 results, _oprteps=0.00001.

Algorithm	Iterations		$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
1	150	Coefficient	0.144	0.157	-0.135	-89.083	-0.912	0.440	0.049	1.988	1.256
		Standard Error	0.170	0.078	0.033	126.027	0.456	0.230	0.064	1.228	0.432
2	28	Coefficient	0.142	-0.155	0.136	-88.483	-0.922	0.436	0.049	2.047	1.237
		Standard Error	0.167	0.078	0.032	127.113	0.455	0.232	0.065	1.294	0.441
3	1	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
4	150	Coefficient	-0.051	0.131	0.541	-123.141	-0.654	0.418	0.066	7.010	0.365
		Standard Error	0.382	0.127	0.202	108.257	0.335	0.336	0.056	10.033	0.502
5	15	Coefficient	0.142	-0.155	-0.136	-88.481	-0.922	0.436	0.049	2.047	1.237
		Standard Error	0.167	0.078	0.032	127.058	0.455	0.232	0.065	1.294	0.441
6	150	Coefficient	-2.260	0.930	-0.078	-71.093	-0.429	0.668	0.038	4.562	0.491
		Standard Error	1.198	0.473	0.029	63.154	0.288	0.129	0.032	4.782	0.490

Key: algorithm 1 - Steepest descent
algorithm 2 - BFGS
algorithm 3 - BFGS-SC
algorithm 4 - DFP
algorithm 5 - Newton
algorithm 6 - PRCG

A dash (-) denotes no estimate due to algorithm failure

Table 3: Dataset 2, *Gauss* 3 results.

Algorithm	Iterations		$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
1	150	Coefficient	-0.082	1.071e-9	0.220	-25.236	-0.833	0.239	0.017	3.497	0.803
		Standard Error	0.131	0.040	0.030	132.164	0.331	0.176	0.0367	1.905	0.332
2	54	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
3	1	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
4	150	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
5	7	Coefficient	0.500	-1.283	-0.082	-38.335	-0.335	0.660	0.021	1.154	1.468
		Standard Error	0.365	0.805	0.037	0.678	0.280	0.135	0.035	0.678	0.451
6	150	Coefficient	0.047	0.0004	0.221	-28.182	-0.868	0.207	0.019	3.744	0.771
		Standard Error	0.162	0.033	0.045	140.213	0.311	0.172	0.071	1.914	0.311

Key: algorithm 1 - Steepest descent
algorithm 2 - BFGS
algorithm 3 - BFGS-SC
algorithm 4 - DFP
algorithm 5 - Newton
algorithm 6 - PRCG

A dash (-) denotes no estimate due to algorithm failure

Table 4: Summary of *Gauss* 5 results.

Dataset		1		2		3		4		5		6		7		8		9		10		11	
Algorithm	_oprteps																						
1: Steepest descent	0.00001	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S
	0.001	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S
	0.1	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S
	1.0	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S	150	S
2: BFGS	0.00001	28	S	45	F	31	S	26	S	118	F	60	F	52	F	57	F	65	F	29	F	27	S
	0.001	29	S	45	F	31	S	26	S	118	F	60	F	50	F	57	F	65	F	29	F	27	S
	0.1	47	F	45	F	31	S	26	S	118	F	60	F	46	F	57	F	65	F	29	F	27	S
	1.0	20	S	45	F	31	S	26	S	118	F	60	F	32	S	57	F	65	F	29	F	27	S
4: DFP	0.00001	150	S	150	S	74	S	150	F	150	S	150	F	150	F	150	S	150	S	150	S	150	S
	0.001	150	S	150	S	74	S	150	F	150	S	150	F	150	S	150	S	150	S	150	S	150	S
	0.1	150	S	150	S	74	S	150	F	150	S	150	F	150	S	150	S	150	S	150	S	150	S
	1.0	150	S	150	S	74	S	150	F	150	S	150	F	139	S	150	S	150	S	150	S	150	S
5: Newton	0.00001	15	S	23	F	8	S	18	S	36	F	11	S	11	S	150	F	22	S	150	F	10	S
	0.001	15	S	23	F	8	S	18	S	36	F	11	S	9	S	134	S	22	S	39	S	10	S
	0.1	15	S	23	F	8	S	18	S	36	F	11	S	9	S	16	S	22	S	15	S	10	S
	1.0	15	S	23	F	8	S	18	S	36	F	11	S	9	S	13	S	22	S	15	S	10	S
6: PRCG	0.00001	150	S	150	S	150	S	88	F	150	S	150	F	150	F	150	S	150	F	150	S	150	S
	0.001	150	S	150	S	150	S	150	S	150	S	150	F	150	F	150	S	150	F	150	S	150	S
	0.1	150	S	150	S	150	S	150	S	150	S	150	S	150	F	150	S	150	F	150	S	150	S
	1.0	150	S	150	S	150	S	150	F	150	S	150	S	150	F	150	S	150	S	150	S	150	S

Key: S - success; F - failure.
Numbers within table are numbers of iterations.

Note: Algorithm 3 (BFGS-SC), omitted from the table, failed in all cases.

Appendix A

Summary of *Gauss* 5 Results.

Table A1: Dataset 1, *Gauss* 5.

Coefficient Estimates				$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	_oprteps	Iterations										
1: Steepest Descent*	0.00001	150	Coefficient	0.144	0.157	-0.135	-89.083	-0.912	0.440	0.049	1.988	1.256
			Standard Error	0.170	0.078	0.033	126.027	0.456	0.230	0.064	1.228	0.432
2: BFGS	0.00001	28	Coefficient	0.142	-0.155	0.136	-88.483	-0.922	0.436	0.049	2.047	1.237
			Standard Error	0.167	0.078	0.032	127.113	0.455	0.232	0.065	1.294	0.441
	0.001	29	Coefficient	0.142	-0.155	0.136	-88.482	-0.922	0.436	0.049	2.047	0.441
			Standard Error	0.167	0.078	0.032	127.193	0.455	0.232	0.065	1.294	0.441
	0.1	47	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
	1.0	20	Coefficient	0.142	-0.155	0.136	-88.482	-0.922	0.436	0.049	2.047	1.237
			Standard Error	0.167	0.078	0.032	127.353	0.455	0.232	0.065	1.294	0.441
3: BFGS-SC*	0.00001	1	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
4: DFP	0.00001	150	Coefficient	-0.051	0.131	0.541	-123.141	-0.654	0.418	0.066	7.010	0.365
			Standard Error	0.382	0.127	0.202	108.257	0.335	0.336	0.056	10.033	0.502
	0.001	150	Coefficient	-0.009	0.175	0.540	-111.722	-0.671	0.489	0.060	8.410	0.302
			Standard Error	0.360	0.081	0.134	87.365	0.321	0.199	0.045	15.513	0.545
	0.1	150	Coefficient	-0.005	0.150	0.559	-114.967	-0.650	0.462	0.0613	5.559	0.450
			Standard Error	0.359	0.086	0.178	90.850	0.334	0.227	0.047	6.615	0.508
	1.0	150	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
5: Newton*	0.00001	15	Coefficient	0.142	-0.155	-0.136	-88.481	-0.922	0.436	0.049	2.047	1.237
			Standard Error	0.167	0.078	0.032	127.058	0.455	0.232	0.065	1.294	0.441
6: PRCG*	0.00001	150	Coefficient	-2.260	0.930	-0.078	-71.093	-0.429	0.668	0.038	4.562	0.491
			Standard Error	1.198	0.473	0.029	63.154	0.288	0.129	0.032	4.782	0.490

* indicates same results for all values of _oprteps.

Appendix B

Detailed results, *Gauss* 3 & 5, for all datasets.

Table B1: Dataset 2, *Gauss* 5 results.

Coefficient Estimates				$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	_oprteps	Iterations										
1: Steepest Descent*	0.00001	150	Coefficient	0.099	1.965e-10	0.221	-26.144	-0.806	0.265	0.017	3.011	0.892
			Standard Error	0.154	0.044	0.032	130.494	0.357	0.194	0.066	1.654	0.356
2: BFGS*	0.00001	45	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
3: BFGS-SC*	0.00001	1	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
4: DFP*	0.00001	150	Coefficient	0.111	0.001	0.447	-67.915	-0.603	0.257	0.038	113.968	0.022
			Standard Error	0.244	0.060	0.093	97.525	0.372	0.180	0.050	2662.126	0.520
5: Newton*	0.00001	23	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
6: PRCG	0.00001	150	Coefficient	-0.043	5.354e-5	0.222	-24.321	-0.883	0.174	0.017	6.401	0.491
			Standard Error	0.107	0.031	0.027	149.158	0.289	0.143	0.076	4.692	0.325
	0.001	150	Coefficient	-0.042	7.674e-6	0.222	-24.257	-0.884	0.173	0.017	6.539	0.481
			Standard Error	0.106	0.031	0.027	146.673	0.288	0.142	0.074	4.881	0.326
	0.1	150	Coefficient	-0.041	5.990e-5	0.221	-24.113	-0.885	0.172	0.017	6.700	0.471
			Standard Error	0.109	0.030	0.044	150.050	0.289	0.145	0.076	5.106	0.327
	1.0	150	Coefficient	-0.032	2.017e-5	0.221	-23.576	-0.893	0.153	0.016	11.967	0.274
			Standard Error	0.111	0.028	0.036	151.022	0.281	0.137	0.077	15.682	0.347

* indicates same results for all values of _oprteps.

Table B2: Dataset 3, *Gauss* 5 results.

Coefficient Estimates				$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	_oprteps	Iterations										
1: Steepest Descent*	0.00001	150	Coefficient	0.171	-1.113e-9	0.143	-119.44	-0.857	0.524	0.064	1.632	1.214
			Standard Error	0.187	0.039	0.032	114.294	0.452	0.142	0.058	0.778	0.255
2: BFGS*	0.00001	31	Coefficient	-0.298	-2.698e-6	0.471	-100.439	-0.642	0.543	0.054	2.708	0.752
			Standard Error	0.242	0.104	0.079	80.059	0.331	0.231	0.041	3.676	0.791
3: BFGS-SC*	0.00001	1		-	-	-	-	-	-	-	-	-
				-	-	-	-	-	-	-	-	-
4: DFP*	0.00001	74	Coefficient	0.298	7.595e-6	0.471	-100.441	-0.642	0.543	0.054	2.710	0.751
			Standard Error	0.242	0.111	0.079	80.090	0.331	0.231	0.041	3.686	0.792
5: Newton*	0.00001	8	Coefficient	0.299	4.393e-6	0.471	-100.438	-0.642	0.543	0.054	2.705	0.752
			Standard Error	0.242	0.108	0.079	80.034	0.331	0.230	0.041	3.664	0.789
6: PRCG	0.00001	150	Coefficient	0.119	0.0004	0.136	-109.930	-0.971	0.470	0.059	2.272	1.070
			Standard Error	0.146	0.034	0.0347	140.275	0.451	0.177	0.071	1.756	0.385
	0.001	150	Coefficient	0.107	-0.0003	-0.137	-106.029	-1.000	0.455	0.058	2.481	1.029
			Standard Error	0.154	0.033	0.035	149.642	0.471	0.215	0.076	2.607	0.520
	0.1	150	Coefficient	0.107	-0.0006	0.137	-106.168	-0.100	0.455	0.058	2.473	1.031
			Standard Error	0.153	0.033	0.035	148.540	0.468	0.213	0.075	2.560	0.512
	1.0	150	Coefficient	0.115	-0.0002	0.137	-109.364	-0.979	0.467	0.059	2.302	1.065
			Standard Error	0.143	0.034	0.035	141.438	0.446	0.180	0.072	1.825	0.396

* indicates same results for all values of _oprteps.

Table B3: Dataset 4, *Gauss* 5 results.

Coefficient Estimates				$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\widetilde{\sigma}}$
Algorithm	_oprteps	Iterations										
1: Steepest Descent*	0.00001	150	Coefficient	0.106	2.111e-8	-0.132	-191.748	-1.003	0.408	0.101	2.997	0.870
			Standard Error	0.120	0.127	0.021	153.395	0.424	0.145	0.078	1.684	0.289
2: BFGS*	0.00001	26	Coefficient	0.281	-3.111e-6	0.458	-142.422	-0.614	0.445	0.075	92.029	0.024
			Standard Error	0.211	0.103	0.064	89.201	0.349	0.152	0.046	3360.837	0.894
3: BFGS-SC*	0.00001	1	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
4: DFP*	0.00001	150	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
5: Newton*	0.00001	18	Coefficient	0.137	0.039	0.128	-201.126	-0.991	0.391	0.106	5.063	0.575
			Standard Error	0.122	0.044	0.017	166.279	0.471	0.163	0.085	6.120	0.576
6: PRCG	0.00001	88	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
	0.001	150	Coefficient	1.544	1.173	-0.087	-102.800	-0.474	0.751	0.054	4.273	0.450
			Standard Error	0.622	0.518	0.057	62.306	0.258	0.118	0.032	3.649	0.363
	0.1	150	Coefficient	1.465	1.190	0.084	-103.752	-0.473	0.751	0.054	3.725	0.513
			Standard Error	0.598	0.539	0.061	63.237	0.259	0.118	0.032	2.901	0.371
	1.0	150	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-

* indicates same results for all values of _oprteps.

Table B4: Dataset 5, *Gauss* 5 results.

Coefficient Estimates				$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	_oprteps	Iterations										
1: Steepest Descent*	0.00001	150	Coefficient	0.157	-0.158	-0.140	-136.518	-0.965	0.446	0.073	1.852	1.280
			Standard Error	0.180	0.080	0.033	123.255	0.440	0.217	0.063	1.123	0.423
2: BFGS*		118	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
3: BFGS-SC*	0.00001	1	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
4: DFP*	0.00001	150	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
5: Newton*	0.00001	36	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
6: PRCG	0.00001	150	Coefficient	-0.146	0.151	-0.142	-135.371	-1.035	0.419	0.073	2.376	1.115
			Standard Error	0.153	0.076	0.032	135.720	0.446	0.235	0.069	1.890	0.540
	0.001	150	Coefficient	-0.147	0.152	-0.141	-135.525	-1.023	0.424	0.073	2.271	1.145
			Standard Error	0.157	0.076	0.034	133.185	0.443	0.232	0.068	1.684	0.508
	0.1	150	Coefficient	-0.143	0.149	-0.142	-135.172	-1.052	0.412	0.073	2.546	1.070
			Standard Error	0.146	0.075	0.026	139.223	0.451	0.242	0.071	2.303	0.605
	1.0	150	Coefficient	-0.152	0.155	-0.140	-136.044	-0.993	0.436	0.073	2.039	1.216
			Standard Error	0.168	0.077	0.032	127.990	0.439	0.223	0.065	1.332	0.453

* indicates same results for all values of _oprteps.

Table B5: Dataset 6, *Gauss* 5 results.

Coefficient Estimates				$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	_oprteps	Iterations										
1: Steepest Descent*	0.00001	150	Coefficient	0.151	-0.155	0.141	-143.357	-0.987	0.438	0.077	1.874	1.286
			Standard Error	0.179	0.081	0.034	129.535	0.441	0.222	0.066	1.203	0.449
2: BFGS*	0.00001	60	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
3: BFGS-SC*	0.00001	1	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
4: DFP*	0.00001	150	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
5: Newton*	0.00001	11	Coefficient	0.151	0.154	0.141	-143.253	-0.990	0.437	0.077	1.895	1.279
			Standard Error	0.177	0.081	0.034	130.389	0.441	0.223	0.066	1.228	0.453
6: PRCG	0.00001	150	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
	0.001	150	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
	0.1	150	Coefficient	-0.151	0.154	-0.141	-143.245	-0.990	0.437	0.077	1.897	1.278
			Standard Error	0.177	0.081	0.034	130.254	0.441	0.223	0.066	1.229	0.453
		150	Coefficient	-0.151	0.154	-0.141	-143.251	-0.990	0.437	0.077	1.895	1.279
	1.0		Standard Error	0.177	0.081	0.034	130.267	0.441	0.222	0.066	1.228	0.453

* indicates same results for all values of _oprteps.

Table B6: Dataset 7, *Gauss* 5 results.

Coefficient Estimates				$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	_oprteps	Iterations										
1: Steepest Descent*	0.00001	150	Coefficient	0.155	0.150	0.143	-174.889	-1.032	0.418	0.093	1.889	1.283
			Standard Error	0.189	0.083	0.034	136.196	0.450	0.225	0.069	1.258	0.465
2:BFGS	0.00001	52	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
	0.001	50	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
	0.1	46	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
	1.0	32	Coefficient	0.159	-0.152	0.143	-174.902	-1.013	0.425	0.093	1.775	1.324
			Standard Error	0.198	0.084	0.035	132.935	0.450	0.221	0.069	1.123	0.444
3: BFGS-SC*	0.00001	1	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
4: DFP	0.00001	150	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
	0.001	150	Coefficient	-0.193	-0.094	0.682	-187.177	-0.681	0.413	0.098	8.305	0.291
			Standard Error	0.371	0.109	0.238	104.643	0.380	0.233	0.054	15.097	0.518
	0.1	150	Coefficient	-0.116	-0.137	0.565	-192.728	-0.780	0.419	0.101	7.017	0.356
			Standard Error	0.366	0.098	0.280	125.823	0.372	0.275	0.065	10.953	0.537
	1.0	139	Coefficient	0.159	-0.152	0.143	-174.902	-1.013	0.425	0.093	1.775	1.324
			Standard Error	0.198	0.084	0.035	132.867	0.450	0.221	0.068	1.123	0.443
5: Newton	0.00001	11	Coefficient	0.159	0.152	0.143	-174.902	-1.013	0.425	0.093	1.775	1.324
			Standard Error	0.198	0.084	0.035	133.131	0.450	0.221	0.068	1.122	0.443
	0.001	9	Coefficient	0.159	0.152	0.143	-174.902	-1.013	0.425	0.093	1.775	1.324
			Standard Error	0.198	0.084	0.035	133.131	0.450	0.221	0.068	1.122	0.443
6: PRCG*	0.00001	150	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-

* indicates same results for all values of _oprteps.

Table B7: Dataset 8, *Gauss* 5 results.

Coefficient Estimates				$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	_oprteps	Iterations										
1: Steepest Descent*	0.00001	150	Coefficient	0.101	-0.120	0.072	-27.747	-0.814	0.431	0.0177	1.743	1.452
			Standard Error	0.147	0.069	0.015	146.613	0.457	0.243	0.074	0.948	0.345
2: BFGS*	0.00001	57	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
3: BFGS-SC*	0.00001	1	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
4: DFP*	0.00001	150	Coefficient	-2.013	0.453	-0.115	-60.304	-0.363	0.651	0.0323	11.627	0.200
			Standard Error	1.071	0.206	0.068	65.050	0.303	0.140	0.033	27.879	0.476
5: Newton	0.00001	150	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
	0.001	134	Coefficient	0.154	0.148	0.138	-60.245	-0.875	0.453	0.034	2.135	1.202
			Standard Error	0.177	0.082	0.039	125.621	0.495	0.236	0.064	1.387	0.452
	0.1	16	Coefficient	-0.086	0.114	0.071	-14.095	-0.896	0.396	0.011	2.211	1.316
			Standard Error	0.118	0.059	0.015	164.111	0.448	0.261	0.083	1.281	0.372
	1.0	13	Coefficient	0.086	0.114	-0.071	-14.095	-0.896	0.396	0.011	2.211	1.316
			Standard Error	0.119	0.059	0.015	194.625	0.449	0.261	0.0985	1.288	0.373
6: PRCG*	0.00001	150	Coefficient	0.117	0.087	-0.134	-57.145	-1.029	0.322	0.033	3.229	0.980
			Standard Error	0.132	0.110	0.023	160.908	0.609	0.383	0.082	5.990	1.151

* indicates same results for all values of _oprteps.

Table B8: Dataset 9, *Gauss* 5 results.

Coefficient Estimates				$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	_oprteps	Iterations										
1: Steep*	0.00001	150	Coefficient	0.105	-0.123	-0.072	-31.254	-0.809	0.431	0.019	1.736	1.437
			Standard Error	0.146	0.067	0.015	137.396	0.449	0.239	0.070	0.915	0.334
2: BFGS*	0.00001	65	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
3: BFGS-SC*	0.00001	1	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
4: DFP*	0.00001	150	Coefficient	1.819	0.446	-0.122	-61.274	-0.363	0.646	0.033	18.626	0.125
			Standard Error	0.870	0.162	0.064	64.172	0.301	0.138	0.033	74.657	0.500
5: Newton*	0.00001	22	Coefficient	0.086	-0.116	-0.071	-19.980	-0.904	0.391	0.014	2.253	1.290
			Standard Error	0.114	0.057	0.015	147.587	0.436	0.259	0.075	1.274	0.362
6: PRCG	0.00001	150	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
	0.1	150	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
	1.0	150	Coefficient	0.081	-0.113	0.071	-16.504	-0.935	0.378	0.013	2.472	1.236
			Standard Error	0.109	0.057	0.014	186.950	0.439	0.268	0.095	1.496	0.384

* indicates same results for all values of _oprteps.

Table B9: Dataset 10, *Gauss* 5 results.

Coefficients Estimates				$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	_oprteps	Iterations										
1: Steepest Descent*	0.00001	150	Coefficient	0.170	-0.152	0.137	-62.455	-0.828	0.464	0.035	1.986	1.220
			Standard Error	0.191	0.074	0.039	113.734	0.494	0.217	0.058	1.131	0.396
2: BFGS*	0.00001	29	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
3: BFGS-SC*	0.00001	1	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
4: DFP*	0.00001	150	Coefficient	1.742	0.439	-0.121	-56.969	-0.342	0.642	0.031	7.573	0.303
			Standard Error	1.555	0.198	0.061	68.276	0.296	0.144	0.035	14.237	0.567
5: Newton	0.00001	150	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
	0.001	39	Coefficient	-0.089	-0.115	-0.072	-28.476	-0.904	0.391	0.018	2.243	1.278
			Standard Error	0.114	0.059	0.015	154.898	0.431	0.255	0.078	1.249	0.354
	0.1	15	Coefficient	0.161	-0.148	-0.138	-60.681	-0.872	0.447	0.035	2.267	1.135
			Standard Error	0.175	0.078	0.038	118.902	0.494	0.232	0.060	1.453	0.436
	1.0	15	Coefficient	-0.089	0.115	-0.072	-28.476	-0.904	0.391	0.018	2.243	1.278
			Standard Error	0.114	0.059	0.015	155.825	0.431	0.255	0.079	1.248	0.354
6: PRCG*	0.00001	150	Coefficient	0.088	0.114	-0.072	-27.608	-0.914	0.387	0.018	2.312	1.261
			Standard Error	0.114	0.060	0.021	153.456	0.431	0.259	0.078	1.311	0.361

* indicates same results for all values of _oprteps.

Table B10: Dataset 11, *Gauss* 5 results.

Coefficient Estimates				$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	_oprteps	Iterations										
1: Steepest Descent*	0.00001	150	Coefficient	0.108	0.127	0.071	-18.092	-0.780	0.450	0.013	1.555	1.480
			Standard Error	0.154	0.065	0.014	120.389	0.437	0.227	0.061	0.744	0.303
2: BFGS*	0.00001	27	Coefficient	1.935	0.450	0.110	-46.994	-0.316	0.645	0.025	99.894	0.023
			Standard Error	0.661	0.158	0.046	58.592	0.287	0.135	0.030	3595.458	0.829
3: BFGS-SC*	0.00001	1	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
4: DFP*	0.00001	150	Coefficient	1.935	0.450	0.110	-46.994	-0.316	0.645	0.025	92.147	0.025
			Standard Error	0.661	0.158	0.046	58.622	0.287	0.135	0.030	3319.922	0.899
5: Newton*	0.00001	10	Coefficient	0.136	0.127	0.136	-35.965	-0.957	0.404	0.022	2.561	1.073
			Standard Error	0.140	0.160	0.024	124.981	0.508	0.354	0.063	2.076	0.495
6: PRCG*	0.00001	150	Coefficient	0.085	-0.118	0.070	-6.998	-0.905	0.397	0.008	2.167	1.297
			Standard Error	0.123	0.061	0.027	132.466	0.427	0.254	0.067	1.107	0.325

* indicates same results for all values of _oprteps.

Table B11: Dataset 5, *Gauss* 3 results.

Coefficient Estimates			$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	Iterations										
1: Steepest Descent	150	Coefficient	0.153	-0.156	-0.140	-136.197	-0.984	0.439	0.073	1.978	1.236
		Standard Error	0.172	0.078	0.033	126.625	0.439	0.221	0.064	1.259	0.443
2: BFGS	62	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
3: BFGS-SC	1	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
4: DFP	150	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
5: Newton	11	Coefficient	-0.152	0.155	0.140	-136.081	-0.991	0.436	0.073	2.025	1.221
		Standard Error	0.169	0.078	0.032	127.704	0.439	0.223	0.065	1.315	0.451
6: PRCG	150	Coefficient	-0.146	-0.151	0.141	-135.445	-1.029	0.421	0.073	2.329	1.129
		Standard Error	0.154	0.076	0.032	134.591	0.444	0.233	0.068	1.794	0.525

Table B12: Dataset 7, *Gauss* 3 results.

Coefficient Estimates			$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	Iterations										
1: Steepest Descent	150	Coefficient	1.538	-1.175	-0.083	-103.345	-0.474	0.750	0.054	4.402	0.438
		Standard Error	0.583	0.441	0.022	62.183	0.257	0.118	0.032	3.830	0.361
2: BFGS	26	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
3: BFGS-SC	1	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
4: DFP	150	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
5: Newton	10	Coefficient	0.137	-0.039	-0.128	-201.127	-0.991	0.391	0.106	5.063	0.575
		Standard Error	0.122	0.044	0.017	166.789	0.471	0.163	0.085	6.120	0.576
6: PRCG	12	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-

Table B13: Dataset 8, *Gauss* 3 results.

Coefficient Estimates			$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	Iterations										
1: Steepest Descent	150	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
2: BFGS	42	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
3: BFGS-SC	1	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
4: DFP	150	Coefficient	1.966	0.447	0.115	-60.980	-0.363	0.648	0.033	14.822	0.158
		Standard Error	0.708	0.172	0.057	64.908	0.305	0.137	0.033	45.736	0.484
5: Newton	10	Coefficient	0.154	0.148	-0.138	-60.245	-0.875	0.453	0.034	2.135	1.202
		Standard Error	0.177	0.082	0.039	125.218	0.495	0.236	0.063	1.387	0.452
6: PRCG	150	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-

Table B14: Dataset 11, *Gauss* 3 results.

Coefficient Estimates			$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	Iterations										
1: Steepest Descent	150	Coefficient	-0.091	0.120	-0.070	-10.413	-0.871	0.411	0.009	1.977	1.348
		Standard Error	0.126	0.057	0.013	149.804	0.430	0.244	0.076	0.978	0.316
2: BFGS	27	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
3: BFGS-SC	1	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
4: DFP	150	Coefficient	1.895	0.386	0.127	-46.959	-0.307	0.639	0.025	7.190	0.316
		Standard Error	0.750	0.188	0.070	59.678	0.289	0.137	0.031	10.033	0.432
5: Newton	14	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
6: PRCG	150	Coefficient	-0.144	-0.157	-0.069	-22.715	-0.776	0.439	0.015	2.331	1.203
		Standard Error	0.257	0.067	0.015	150.104	0.738	0.260	0.075	1.625	0.399

Appendix C

Hamilton’s example 1 & 2 results.

Table C1: Example 1, *Gauss* 3.

Coefficient Estimates			$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	Iterations								
1: Steepest Descent	150	Coefficient	0.075	4.097e-6	4.667	0.308	0.196	1.554	1.047
		Standard Error	0.010	0.002	1.0	0.053	0.010	0.331	0.083
2: BFGS	40	Coefficient	0.074	-7.782e-11	4.747	0.307	0.196	1.758	1.027
		Standard Error	0.006	0.002	1.096	0.058	0.010	0.398	0.082
3: BFGS-SC	1	Coefficient	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-
4: DFP	37	Coefficient	0.074	2.696e-9	4.747	0.307	0.196	1.758	1.027
		Standard Error	0.006	0.002	1.096	0.058	0.010	0.398	0.082
5: Newton	19	Coefficient	-0.074	2.078e-12	4.747	0.307	0.196	1.758	1.027
		Standard Error	0.006	0.002	1.096	0.058	0.010	0.398	0.082
6: PRCG	150	Coefficient	0.093	-0.0002	4.262	0.309	0.197	0.987	1.131
		Standard Error	0.007	0.004	0.647	0.038	0.010	0.162	0.090

Table C2: Example 1, *Gauss* 5 results, oprteps= 0.00001.

Coefficient Estimates			$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	Iterations								
1: Steepest Descent	150	Coefficient	0.103	0.0002	3.966	0.286	0.214	0.774	1.177
		Standard Error	0.012	0.004	0.523	0.032	0.011	0.119	0.094
2: BFGS	20	Coefficient	0.080	-3.09e-9	4.387	0.276	0.213	1.549	1.042
		Standard Error	0.005	0.001	0.950	0.052	0.010	0.335	0.081
3: BFGS-SC	1	Coefficient	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-
4: DFP	41	Coefficient	0.080	-1.0e-9	4.387	0.276	0.213	1.549	1.042
		Standard Error	0.005	0.001	0.950	0.052	0.010	0.335	0.081
5: Newton	18	Coefficient	0.080	1.099e-9	4.387	0.276	0.213	1.549	1.042
		Standard Error	0.005	0.001	0.950	0.052	0.010	0.335	0.081
6: PRCG	150	Coefficient	-0.081	-7.298e-5	4.330	0.278	0.213	1.323	1.068
		Standard Error	0.021	0.002	0.882	0.047	0.010	0.295	0.083

Table C3: Example 2, *Gauss* 3.

Coefficient Estimates			$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	Iterations										
1: Steepest Descent	150	Coefficient	0.329	0.329	0.0003	7.861	0.896	1.200	0.794	2.991	0.909
		Standard Error	0.041	0.040	0.015	1.088	0.281	0.356	0.070	0.563	0.124
2: BFGS	45	Coefficient	0.325	0.321	3.336e-8	7.929	0.924	1.232	0.782	4.340	0.716
		Standard Error	0.045	0.034	0.013	1.242	0.315	0.399	0.066	1.176	0.143
3: BFGS-SC	1	Coefficient	-	-	-	-	-	-	-	-	-
		Standard Error	-	-	-	-	-	-	-	-	-
4: DFP	150	Coefficient	0.327	-0.318	0.0002	7.935	0.925	1.233	0.781	4.503	0.698
		Standard Error	0.047	0.043	0.013	1.257	0.320	0.403	0.066	1.294	0.148
5: Newton	15	Coefficient	0.325	-0.321	1.071e-8	7.929	0.924	1.232	0.782	4.340	0.716
		Standard Error	0.045	0.034	0.013	1.242	0.315	0.399	0.066	1.176	0.143
6: PRCG	150	Coefficient	0.326	0.323	-0.0005	7.912	0.916	1.223	0.786	3.827	0.778
		Standard Error	0.040	0.035	0.013	1.192	0.303	0.385	0.067	0.884	0.131

Table C4: Example 2, *Gauss* 5 results.

Coefficient Estimates				$\widetilde{g}_1$	$\widetilde{g}_2$	$\widetilde{g}_3$	$\widetilde{\alpha}_0$	$\widetilde{\alpha}_1$	$\widetilde{\alpha}_2$	$\widetilde{\alpha}_3$	$\widetilde{\zeta}$	$\widetilde{\sigma}$
Algorithm	_oprteps	Iterations										
1: Steepest Descent*	0.00001	150	Coefficient	0.398	0.334	-0.002	8.020	0.796	1.368	0.825	2.783	0.936
			Standard Error	0.038	0.069	0.021	0.995	0.261	0.328	0.072	0.512	0.124
2: BFGS	0.00001	24	Coefficient	0.394	0.321	-4.063e-7	8.155	0.806	1.391	0.834	4.073	0.735
			Standard Error	0.026	0.033	0.036	1.131	0.296	0.367	0.067	1.040	0.138
	0.001	22	Coefficient	0.394	0.321	1.141e-7	8.155	0.806	1.391	0.834	4.073	0.735
			Standard Error	0.026	0.033	0.034	1.131	0.296	0.367	0.067	1.039	0.138
	0.1	19	Coefficient	0.394	0.321	1.730e-6	8.155	0.806	1.391	0.834	4.073	0.735
			Standard Error	0.026	0.033	0.037	1.131	0.296	0.367	0.067	1.040	0.138
	1.0	53	Coefficient	0.394	0.321	1.976e-7	8.155	0.806	1.391	0.834	4.073	0.735
			Standard Error	0.026	0.033	0.037	1.131	0.296	0.367	0.067	1.040	0.138
3: BFGS-SC*	0.00001	1	Coefficient	-	-	-	-	-	-	-	-	-
			Standard Error	-	-	-	-	-	-	-	-	-
4: DFP*	0.00001	22	Coefficient	0.394	0.321	2.440e-6	8.155	0.806	1.391	0.834	4.073	0.735
			Standard Error	0.026	0.033	0.036	1.131	0.296	0.367	0.067	1.040	0.138
5: Newton*	0.00001	10	Coefficient	-0.394	0.321	-4.195e-8	8.155	0.806	1.391	0.834	4.073	0.735
			Standard Error	0.026	0.033	0.038	1.131	0.296	0.367	0.067	1.040	0.138
6: PRCG*	0.00001	150	Coefficient	0.394	-0.321	7.281e-7	8.147	0.806	1.389	0.833	3.928	0.753
			Standard Error	0.026	0.034	0.030	1.118	0.293	0.364	0.068	0.956	0.134

* indicates same results for all values of _oprteps.