Journal of Statistical Planning and Inference 139 (2009) 1276 – 1285

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Parameter estimation in semi-linear models using a maximal invariant likelihood function

Jahar L. Bhowmik[a,b], Maxwell L. King[c,*]

[a] Faculty of Life and Social Sciences, Swinburne University of Technology, VIC 3122, Australia
[b] Department of Econometrics, Monash University, VIC 3800, Australia
[c] Department of Econometrics and Business Statistics, Monash University, VIC 3800, Australia

ARTICLE INFO

Article history:
Received 16 September 2004
Received in revised form
4 March 2008
Accepted 15 July 2008
Available online 17 August 2008

Keywords:
Maximum likelihood estimation
Non-linear modelling
Simulation experiment
Two-step estimation

ABSTRACT

In this paper, we consider the problem of estimation of semi-linear regression models. Using invariance arguments, Bhowmik and King [2007. Maximal invariant likelihood based testing of semi-linear models. Statist. Papers 48, 357–383] derived the probability density function of the maximal invariant statistic for the non-linear component of these models. Using this density function as a likelihood function allows us to estimate these models in a two-step process. First the non-linear component parameters are estimated by maximising the maximal invariant likelihood function. Then the non-linear component, with the parameter values replaced by estimates, is treated as a regressor and ordinary least squares is used to estimate the remaining parameters. We report the results of a simulation study conducted to compare the accuracy of this approach with full maximum likelihood and maximum profile-marginal likelihood estimation. We find maximising the maximal invariant likelihood function typically results in less biased and lower variance estimates than those from full maximum likelihood.

1. Introduction

A major difficulty with full maximum likelihood (FML) estimation of multiparameter models is that it can result in poor estimates in some circumstances. There is a problem of potentially biased estimates arising from the joint estimation of multiple parameters. A good example is the estimate of the variance of the disturbances in the classical linear regression model. In this case, the maximum likelihood estimator is known to be biased and a simple correction is needed to make it unbiased in small samples. This is because the regression coefficients are nuisance parameters when it comes to estimating the variance. For further discussion of the problems of joint estimation of multiple parameters, see Neyman and Scott (1948), Anderson (1970) and Cox and Hinkley (1974). There is a vast amount of literature on the satisfactory handling of nuisance parameters, see for example, Fraser (1967), Kalbfleisch and Sprott (1970, 1973), Bellhouse (1978), King (1983), Barndorff-Nielsen (1983), Lehmann (1986), Cox and Reid (1987), Tunnicliffe Wilson (1989), McCullagh and Tibshirani (1990), Ara and King (1993, 1995), Ara (1995), and Laskar and King (1998, 2001).

One approach that has received a good deal of attention in the literature is the concept of the marginal likelihood which was first introduced by Fraser (1967) and further developed by Kalbfleisch and Sprott (1970). The main idea is to transform the data vector to another random vector, a subvector of which has a likelihood (marginal likelihood) that involves only the parameters of interest and the remainder of which contains no information about those parameters. There is a lot of evidence in the literature that the use of marginal likelihood methods can produce more accurate estimates and, in particular, less biased estimates. See for

* Corresponding author. Tel.: +61 3 9905 2449; fax: +61 3 9905 8039.
E-mail address: Max.King@adm.monash.edu.au (M.L. King).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2008.07.011

example Cooper and Thompson (1977), Kitanidis and Vomvoris (1983), Kitanidis (1983, 1987), Hoeksema and Kitanidis (1985), Kitanidis and Lane (1985), Corduas (1986), Tunnicliffe Wilson (1989), Bellhouse (1991), Shephard (1993), Ara (1995), Ara and King (1993, 1995), Laskar and King (1998) and Rahman and King (1998).

The use of invariance arguments has been a useful method for dealing with some of the problems caused by nuisance parameters, particularly for hypothesis testing. The approach involves noting that the testing problem is invariant to a certain class of transformations on the observed data vector and then requiring the chosen test to also be invariant to such transformations. A key device for test construction is the maximal invariant statistic. It is a vector function of the data vector that takes the same value for data vectors that can be connected by a transformation and different values for those data vectors that cannot be connected by a transformation. Thus the class of all invariant test statistics corresponds to the class of functions of the maximal invariant. This allows us to treat the maximal invariant as the observed data when designing a new test. The density function of the maximal invariant can be treated as a likelihood for this purpose. This function is known as the maximal invariant likelihood (MIL) function.

Ara (1995) showed that the marginal likelihood function and the likelihood of the maximal invariant statistic are equivalent in the case of non-spherical disturbances in the linear regression model. In the context of a linear regression model with a non-linear additive component, Bhowmik and King (2007) derived a MIL function for the non-linear component. The purpose of this paper is to compare maximum MIL (MMIL) and FML approaches to the estimation of parameters in the non-linear component. Using the MIL function to estimate these parameters results in a two-step process. First the non-linear component parameters are estimated by maximising the MIL. Then the non-linear component, with the parameter values replaced by estimates, is treated as a regressor and ordinary least squares is used to estimate the remaining parameters. Alternatively, the full likelihood of the complete model can be maximised to obtain the standard maximum likelihood estimates. The MMIL estimator might be expected to be superior to the full likelihood estimator given the evidence in the literature outlined above. For a slightly more specific semi-linear model, we also derive a profile likelihood from a marginal joint likelihood which we call a profile-marginal likelihood (PML) function which can be used in place of the MIL in the first step of the two-step estimation process.

The plan of this paper is as follows. In Section 2 we derive the likelihood functions (full likelihood, MIL and PML) for the different non-linear models. Monte Carlo experiments to investigate the performance of different ML estimators in the context of non-linear parameters are outlined and reported in Section 3. Finally, concluding remarks are made in Section 4.

2. Theory

Consider the following semi-linear model:

$$y = X_1\beta_1 + g(\beta_2) + u, \quad u \sim \mathrm{N}(0, \sigma^2 I_n) \tag{2.1}$$

where y is an $n \times 1$ vector, X_1 is an $n \times q$ non-stochastic matrix, β_1 is a $q \times 1$ parameter vector and $g(\beta_2)$ is a known non-linear function of the $s \times 1$ parameter vector β_2. Our interest is in estimating (and testing) β_2. This problem is invariant to transformations of the form

$$y \to \gamma_0 y + X_1\gamma \tag{2.2}$$

where γ_0 is a positive scalar and γ is a $q \times 1$ vector. Note that $w = z/(z'z)^{1/2}$ is a maximal invariant statistic where $z = Py$, P is any $m \times n$ matrix such that $PP' = I_m$ and $P'P = M_1$, $m = n - q$ and $M_1 = I - X_1(X_1'X_1)^{-1}X_1'$.

To find the density function of w, Bhowmik and King (2007) first transformed z to m-dimensional polar co-ordinates $(r, \theta_1, \theta_2, \ldots, \theta_{m-1})$ via

$$\begin{aligned} z_1 &= r\cos\theta_1 \\ z_j &= r\left(\prod_{k=1}^{j-1}\sin\theta_k\right)\cos\theta_j \quad \text{for } 2 \leqslant j \leqslant m-1 \\ z_m &= r\prod_{k=1}^{m-1}\sin\theta_k \end{aligned} \tag{2.3}$$

where $0 \leqslant r < \infty$, $0 \leqslant \theta_k \leqslant \pi$, for $k = 1, 2, \ldots, (m-2)$ and $0 \leqslant \theta_{m-1} \leqslant 2\pi$. Observing that $z = rw$, the problem therefore becomes one of integrating out r from the joint density function of $(r, \theta_1, \theta_2, \ldots, \theta_{m-1})$. This results in the marginal density

$$f(\theta_1, \theta_2, \ldots, \theta_{m-1}) = (2\pi)^{-m/2}\prod_{k=1}^{m-2}\sin\ \theta_k^{m-1-k}\ \exp(b(w,\beta_2))\int_0^\infty \exp\left\{-\frac{1}{2}(\lambda - a(w,\beta_2))^2\right\}\lambda^{m-1}\,\mathrm{d}\lambda \tag{2.4}$$