Computational Statistics and Data Analysis 56 (2012) 3309–3321

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

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

An application of shrinkage estimation to the nonlinear regression model

S. Ejaz Ahmed [a,*], Christopher J. Nicol [b]

[a] *Department of Mathematics and Statistics, University of Windsor, Windsor, Ontario, L9B 3P4, Canada*

[b] *Department of Economics, University of Lethbridge, Lethbridge, Alberta, T1K 3M4, Canada*

ARTICLE INFO

Article history:
Received 15 January 2010
Received in revised form 20 July 2010
Accepted 21 July 2010
Available online 1 August 2010

Keywords:
Nonlinear regression
Restricted estimation
Shrinkage and pre-test estimators
Quadratic bias and risk
Simulation

ABSTRACT

Various large sample estimation techniques in a nonlinear regression model are presented. These estimators are based around preliminary tests of significance, and the James–Stein rule. The properties of these estimators are studied when estimating regression coefficients in the multiple nonlinear regression model when it is *a priori* suspected that the coefficients may be restricted to a subspace. A simulation based on a demand for money model shows the superiority of the positive-part shrinkage estimator, in terms of standard measures of asymptotic distributional quadratic bias and risk measures, over a range of economically meaningful parameter values. Further work remains in analysing the use of these estimators in economic applications, relative to the inferential approach which is best to use in these circumstances.

1. Introduction

For many years, it has been known that shrinkage techniques yield estimators which are superior in terms of risk to the maximum likelihood estimator (MLE) over the entire parameter space. Gruber (1998) and Saleh (2006) provide recent starting points which survey this extensive literature. Until recently, these estimators have only been used to a limited extent in empirical applications, in part, owing to the computational burden of and competing alternatives for purposes of statistical inference, which will be expanded on below. However, with improvements in computing capability, and clear advantages to the use of prior information in certain applications, this is changing. For example, numerous cases of shrinkage estimation have appeared in applications involving the real estate market, where appraisers' expert knowledge can be very informative (Knight et al., 1993b), or in hedonic housing pricing models (Knight et al., 1993a; Stevenson, 2001; Bao and Wan, 2007), where real estate experts' knowledge and expertise often yield precise information regarding certain parameter values.

As noted above, the past limited use of shrinkage estimation in economics applications is partly owed to the existence of competing methods of conducting statistical inference in such cases, none of which are as straightforward as the usual methods of inference. One could, for example, employ an empirical Bayes approach to the computation of standard errors of these shrinkage estimators, as was seen in, for example, Maddala et al. (1997). Alternatively, Kazimi and Brownstone (1999) proposed confidence bands for shrinkage estimators using a simple percentile bootstrapping method. They find that "…simple percentile bootstrap confidence bands perform well enough to support empirical applications of shrinkage estimators" (Kazimi and Brownstone, 1999, p. 99), although there remain issues in using bootstrapping methods in this way, for the typical type of econometric model one encounters, where bootstrap sub-sampling is required to ensure consistency, and does not yield consistency in all cases. More recently, Wan et al. (2003) have proposed the use of mean squared error matrices with a class of shrinkage estimators, for the purposes of construction of confidence ellipsoids. Further progress with these aspects is postponed for future research, which may also follow up the work in Ahmed et al. (2009).

* Corresponding author. Tel.: +1 519 253 300; fax: +1 519 971 3649.
E-mail addresses: seahmed@uwindsor.ca (S.E. Ahmed), nicolc@uleth.ca (C.J. Nicol).

0167-9473/$ – see front matter
doi:10.1016/j.csda.2010.07.022

Of course, there is also the more general issue regarding the gains to be obtained through the use of shrinkage estimators. While it is often true that restricted estimators can offer a substantial mean squared error (MSE) gain over unrestricted estimators, there is still a concern that point estimators are less desirable to use when *non-sample information* (NSI) is incorrect. The advantage of the shrinkage approach is, therefore, that NSI is incorporated into estimation to the extent that it appears to be true, given sample information. We therefore view the use of shrinkage estimators as an attractive trade-off in the context of numerous applications in economics in particular.

Shrinkage estimators have been developed for many situations, including the linear regression model. However, many of the models the econometrician wishes to estimate are nonlinear, and often include regressors for which instrumental variables estimation would be necessary, to yield consistent estimators. There are, however, some examples of the use of shrinkage estimation to nonlinear settings. Cases in point include applications to the probit model (Adkins and Hill, 1989), as well as to the Box–Cox transformation (Kim and Hill, 1995), and the Poisson regression model (Sapra, 2003). These papers provide theoretical results indicating the superior performance in terms of risk of certain shrinkage estimators over unrestricted estimation, and also show the range of the parameter space over which this occurs for the particular applications in question.

In this paper, we consider the application of shrinkage estimation to a general form of the nonlinear regression model. We show that shrinkage estimators of the James and Stein type have superior performance in terms of asymptotic bias and risk over other estimators considered, under a variety of conditions. In what follows, three such estimators are developed. Their *asymptotic distributional quadratic bias* (ADB) *and asymptotic distributional quadratic risks* (ADR) properties are then analysed. Our approach is to apply this analysis in the context of the most widely used shrinkage estimators which have received attention in the empirical literature, although all three of these estimators have not always been compared for these purposes, at the same time.

Nonlinear least squares

To fix ideas, consider the regression model,

$$\mathbf{y} = \mathbf{x}[\boldsymbol{\beta}] + \boldsymbol{\epsilon}, \tag{1.1}$$

where $\mathbf{x}[\boldsymbol{\beta}]$ is an $n \times 1$ vector of elements, $x_t[\boldsymbol{\beta}] = x[\mathbf{z}_t; \boldsymbol{\beta}]$, being nonlinear in the $p \times 1$ parameter vector, $\boldsymbol{\beta}$, and $\mathbf{z}_t$ is a $1 \times m$ vector in m-dimensional Euclidean space. $\mathbf{z}_t$ is thus a vector of variables on which the nonlinear function, $x_t[\boldsymbol{\beta}]$, depends. Variables which might comprise elements of $\mathbf{z}_t$ are described in a simulation application, in Section 5. The functional form of $x_t[\boldsymbol{\beta}]$ is identical over all observations, $t = 1, \ldots, n$. The elements of $\boldsymbol{\epsilon}$ are unknown errors or fluctuations, with the properties assumed below.

The objective is to estimate the parameter vector, $\boldsymbol{\beta}$, using nonlinear least squares. The minimisation of $S[\boldsymbol{\beta}] = [\mathbf{y} - \mathbf{x}(\boldsymbol{\beta})]^T[\mathbf{y} - \mathbf{x}(\boldsymbol{\beta})]$ by choice of $\boldsymbol{\beta}$ yields the set of first-order conditions,

$$\{\mathbf{y} - \mathbf{x}[\hat{\boldsymbol{\beta}}]\}^T \mathbf{X}[\hat{\boldsymbol{\beta}}] = 0, \tag{1.2}$$

where $\mathbf{X}[\hat{\boldsymbol{\beta}}] = \partial \mathbf{x}[\boldsymbol{\beta}]/\partial \boldsymbol{\beta} \,|_{\hat{\beta}}$, is the $n \times p$ matrix of derivatives of $\mathbf{x}[\boldsymbol{\beta}]$, evaluated at the nonlinear least squares estimator, $\hat{\boldsymbol{\beta}}$.

Under the regularity conditions described in Appendix (Seber and Wild, 1989), $\boldsymbol{\beta}$ can be consistently estimated by $\hat{\boldsymbol{\beta}}$ using the Gauss–Newton regression (GNR) of $\{\mathbf{y} - \mathbf{x}[\hat{\boldsymbol{\beta}}^i]\}$ on $\mathbf{X}[\hat{\boldsymbol{\beta}}^i]$, where $\hat{\boldsymbol{\beta}}^i$ is a set of starting values for $\boldsymbol{\beta}$. Let $\hat{\mathbf{d}}^i$ be the estimates of the coefficients on $\mathbf{X}[\hat{\boldsymbol{\beta}}^i]$ at iteration i. The Gauss–Newton Regression was introduced in Davidson and MacKinnon (1993), pp. 176–208. The GNR can be a regression performed in the context of a model which is nonlinear in parameters. The nonlinear model is linearised using a Taylor-series expansion at a point which represents a consistent estimator of the parameters of interest. In the process, nonlinear restrictions are incorporated in the derivation of matrices of derivatives which have to be evaluated at this consistent estimator. The result is a "regression model" which is linear in artificial variables dependent on the consistent estimator. Use of the GNR then yields a consistent *and efficient* estimator of the parameters of interest.

Then $\hat{\boldsymbol{\beta}}^{i+1} = \hat{\boldsymbol{\beta}}^i + \hat{\mathbf{d}}^i$, and this iteration procedure continues until an appropriate convergence criterion is satisfied, yielding $\hat{\boldsymbol{\beta}}$ at the final iteration. A consistent estimator of the variance–covariance matrix of $\hat{\boldsymbol{\beta}}$ is given by the estimator of the variance–covariance of $\hat{\mathbf{d}}^*$, where the "$*$" indicates the estimator of the deviation vector on the last iteration. This variance–covariance estimator can be denoted

$$\widehat{\mathrm{VAR}}[\hat{\boldsymbol{\beta}}] = \hat{\sigma}^2\{\mathbf{X}[\hat{\boldsymbol{\beta}}]^T\mathbf{X}[\hat{\boldsymbol{\beta}}]\}^{-1}, \tag{1.3}$$

where $\hat{\sigma}^2 = \mathbf{e}^T\mathbf{e}/n$, and $\mathbf{e} = \mathbf{y} - \mathbf{x}[\hat{\boldsymbol{\beta}}]$. It should also be noted that $\hat{\boldsymbol{\beta}}$ is asymptotically normally distributed, given a further set of regularity conditions, also detailed in Appendix.

Statement of the problem

Suppose that $\boldsymbol{\beta}$ can be partitioned such that $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^T \mid \boldsymbol{\beta}_2^T]^T$. The sub-vectors, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are assumed to have dimensions $q \times 1$ and $r \times 1$ respectively, and $p = q + r$. The associated GNR required to estimate $\boldsymbol{\beta}$ will then be

$$\mathbf{y} - \mathbf{x}[\hat{\boldsymbol{\beta}}] = \mathbf{X}_1[\hat{\boldsymbol{\beta}}]\mathbf{d}_1 + \mathbf{X}_2[\hat{\boldsymbol{\beta}}]\mathbf{d}_2 + \boldsymbol{\mu}, \tag{1.4}$$

where $X_i[\hat{\boldsymbol{\beta}}] = \partial \mathbf{x}_i[\boldsymbol{\beta}]/\partial \boldsymbol{\beta}_i \,|_{\hat{\beta}_i}$, $i = 1, 2$ are $n \times q$ and $n \times r$ matrices of derivatives of $\mathbf{x}_i[\boldsymbol{\beta}]$ with respect to $\boldsymbol{\beta}_i$, $i = 1, 2$ respectively, evaluated at the nonlinear least squares (NLS) estimator.