Shrinkage Estimators for the Nonlinear Regression Model

by

S. E. Ahmed and C. J. Nicol
Department of Mathematics and Statistics
and Department of Economics
University of Regina
Regina, Saskatchewan, S4S 0A2
CANADA

Internet: Chris.Nicol@URegina.CA;
http://www.econ.uregina.ca/nicolc

May 26, 1999.
Revised, September 21, 1999.

Ahmed acknowledges support from the Natural Sciences and Engineering Research Council of Canada. Nicol acknowledges support from the Social Sciences and Humanities Research Council of Canada. The usual disclaimer applies.

Abstract

In this paper, we discuss various large sample estimation techniques in a nonlinear regression model. We propose estimators on the basis of preliminary tests of significance and James-Stein rule. The properties of these estimators are studied in the problem of estimating regression coefficients in the multiple 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 over a range of economically meaningful parameter values. This indicates that this estimator can be usefully employed in important practical situations.

Keywords: nonlinear regression, restricted estimation, shrinkage and pretest estimators, quadratic bias and risk, simulation.

1. Introduction

For many years, it has been known that shrinkage techniques yield estimators which are superior in terms of risk than the maximum likelihood estimator (MLE) over the entire parameter space (Gruber, 1998, p. 1). Gruber provides a recent starting point which surveys this extensive literature. Be this as it may, these estimators have not found extensive use, owing to the absence of means to compute confidence bands. Recently, however, Kazimi and Brownstone (1999) have proposed confidence bands based on bootstrap techniques. They find that "...simple percentile bootstrap confidence bands perform well enough to support empirical applications of shrinkage estimators." (Kazimi and Brownstone, 1999, p. 99).

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 would be little practical purpose in developing such estimators, however, without the prospect of their being a useful addition to the applied econometrician or statistician's repertoire. We consider the application of shrinkage estimation to the nonlinear regression model. Shrinkage estimators of the James and Stein type (Stein, 1956; James and Stein, 1961) are presented which have superior performance in terms of bias and risk over other estimators considered, under a variety of conditions. In what follows, three such estimators are developed. Their ADB and ADR properties are then analysed.

Nonlinear Least Squares

To fix ideas, consider the regression model,

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

where $x[\boldsymbol{\beta}]$ is an $n \times 1$ vector of elements, $x_i[\boldsymbol{\beta}] = x[z_i; \boldsymbol{\beta}]$, being nonlinear in the parameter vector, $\boldsymbol{\beta}$, and z_i is a $k \times 1$ vector in k-dimensional Euclidean space. The functional form of x_i is identical over all observations, $i = 1, \ldots, n$. The vector, $\boldsymbol{\beta}$ is $p \times 1$. The elements of ϵ 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 $\mathrm{SS}[\boldsymbol{\beta}] = [\mathbf{y} - x(\boldsymbol{\beta})]^T[\mathbf{y} - x(\boldsymbol{\beta})]$ by choice of $\boldsymbol{\beta}$ yields the set of first-order conditions,

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

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

Under the following regularity conditions:

A1 The fluctuations, ϵ_i, in the model (1.1) are independent and identically distributed random variables with a continuous distribution function F on the real line $\Re = (-\infty, +\infty)$; We do not make specific assumptions about the functional form of F, although we do assume that $\mathrm{VAR}[\boldsymbol{\epsilon}_i] = \sigma^2, \sigma^2 > 0, \forall i = 1, \ldots, n$;

A2 For all i, $x_i[\boldsymbol{\beta}]$ is a continuous function of $\boldsymbol{\beta}$, for $\boldsymbol{\beta} \in \mathbf{B}$;

A3 $\mathbf{B}$ is a closed and bounded subset of p-dimensional Euclidean space;

A4 (a) $n^{-1}S_n[\boldsymbol{\beta}, \boldsymbol{\beta}_1]$ converges uniformly for all $\boldsymbol{\beta}, \boldsymbol{\beta}_1 \in \mathbf{B}$ to a continuous function, $S[\boldsymbol{\beta}, \boldsymbol{\beta}_1]$; and

A4 (b) $D[\boldsymbol{\beta}, \boldsymbol{\beta}^*] = 0$ if and only if $\boldsymbol{\beta} = \boldsymbol{\beta}^*$, that is, $D[\boldsymbol{\beta}, \boldsymbol{\beta}^*]$ is positive definite. where $S_n[\boldsymbol{\beta}, \boldsymbol{\beta}_1] = \sum_{i=1}^n x_i[\boldsymbol{\beta}]x_i[\boldsymbol{\beta}_1] = x[\boldsymbol{\beta}]^T x[\boldsymbol{\beta}_1]$, and $D_n[\boldsymbol{\beta}, \boldsymbol{\beta}_1] = \sum_{i=1}^n [x_i(\boldsymbol{\beta}) - x_i(\boldsymbol{\beta}_1)]^2 = [x(\boldsymbol{\beta}) - x(\boldsymbol{\beta}_1)]^T [x(\boldsymbol{\beta}) - x(\boldsymbol{\beta}_1)]$,

$\boldsymbol{\beta}$ can be consistently estimated[1] by $\hat{\boldsymbol{\beta}}$ using the Gauss-Newton regression (GNR) of $\{y - x[\hat{\boldsymbol{\beta}}^i]\}$ on $X[\hat{\boldsymbol{\beta}}^i]$, where $\hat{\boldsymbol{\beta}}^i$ is a set of starting values for $\boldsymbol{\beta}$. Let $\hat{d}^i$ be the estimates of the coefficients on $X[\hat{\boldsymbol{\beta}}^i]$ at iteration i. Then $\hat{\boldsymbol{\beta}}^{i+1} = \hat{\boldsymbol{\beta}}^i + \hat{d}^i$, 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 $\hat{d}^*$, where the "$*$" indicates the estimator of the deviation vector on the last iteration. This variance-covariance estimator can be denoted

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

where $\hat{\sigma}^2 = e^T e/n$, and $e = y - x[\hat{\boldsymbol{\beta}}]$. It should also be noted that $\hat{\boldsymbol{\beta}}$ is asymptotically normally distributed, given the following additional regularity conditions

A5 $\boldsymbol{\beta}^*$ is an interior point of $\mathbf{B}$. Let $\mathbf{B}^*$ be an open neighbourhood of $\boldsymbol{\beta}^*$ in $\mathbf{B}$.

A6 The first and second derivatives, $\partial x_i[\boldsymbol{\beta}]/\partial \boldsymbol{\beta}_r$ and $\partial^2 x_i[\boldsymbol{\beta}]/\partial \boldsymbol{\beta}_r \partial \boldsymbol{\beta}_s$, $(r, s = 1, \ldots, p)$ exist and are continuous for all $\boldsymbol{\beta} \in \mathbf{B}$.

A7 $(1/n)\sum_i^n (\partial^2 x_i(\boldsymbol{\beta})/\partial \boldsymbol{\beta})(\partial^2 x_i(\boldsymbol{\beta})/\partial \boldsymbol{\beta})^T [= n^{-1}\mathbf{F}_.^T(\boldsymbol{\beta})\mathbf{F}_.(\boldsymbol{\beta})]$ converges to some matrix, $\boldsymbol{\Omega}(\boldsymbol{\beta})$ uniformly in $\boldsymbol{\beta}$ for $\boldsymbol{\beta} \in \mathbf{B}^*$.

A8 $(1/n)\sum_i^n [\partial^2 x_i(\boldsymbol{\beta})/\partial \boldsymbol{\beta}_r \partial \boldsymbol{\beta}_s]^2$ converges uniformly in $\boldsymbol{\beta}$ for $\boldsymbol{\beta} \in \mathbf{B}^*$, $(r, s = 1, \ldots, p)$.

A9 $\boldsymbol{\Omega} = \boldsymbol{\Omega}(\boldsymbol{\beta}^*)$ is nonsingular.

(see Seber and Wild, 1985, Chapter 12.2.3).

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

$$\mathbf{y} - x[\hat{\boldsymbol{\beta}}] = X_1[\hat{\boldsymbol{\beta}}]d_1 + X_2[\hat{\boldsymbol{\beta}}]d_2 + \mu \tag{1.4}$$

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

[1] See Seber and Wild (1989) Chapter 12.1–12.2 for a detailed discussion of the role played by each of these assumptions.

For later developments, we also denote the partitioned matrix

$$\mathbf{X}[\boldsymbol{\beta}] = \{\mathbf{X}_1[\boldsymbol{\beta}]|\mathbf{X}_2[\boldsymbol{\beta}]\} \tag{1.5}$$

and the product matrix,

$$\mathbf{X}[\boldsymbol{\beta}]^T\mathbf{X}[\boldsymbol{\beta}] = \begin{bmatrix} \mathbf{X}_1[\boldsymbol{\beta}]^T\mathbf{X}_1[\boldsymbol{\beta}] & \mathbf{X}_1[\boldsymbol{\beta}]^T\mathbf{X}_2[\boldsymbol{\beta}] \\ \mathbf{X}_2[\boldsymbol{\beta}]^T\mathbf{X}_1[\boldsymbol{\beta}] & \mathbf{X}_2[\boldsymbol{\beta}]^T\mathbf{X}_2[\boldsymbol{\beta}] \end{bmatrix} \tag{1.6}$$

We are interested in the estimation of the parameter vector, $\boldsymbol{\beta}_1$, in the presence of some sort of *non-sample information* (NSI) about $\boldsymbol{\beta}_2$.

Non-sample Information

We are primarily interested in the estimation of the regression parameter sub-vector $\boldsymbol{\beta_1}$ when the NSI or *uncertain prior information* (UPI) about $\boldsymbol{\beta_2}$ is readily available. This situation may arise when there is over-modelling and one wishes to remove the irrelevant part of the model, (1.1), and increase the efficiency of estimating $\boldsymbol{\beta}$. For example, alluding to the simulation example which will follow, suppose that there is uncertainty over the functional form of a money demand relationship. The functional form could be linear or nonlinear in parameters, and some of the elements of z_i could perhaps be excluded. The possible linearity in parameters and exclusion restrictions on elements of z_i (comprising the NSI) could be ignored. However, this NSI may be used to estimate the parameters of interest. Thus, the regression parameter vector can be partitioned, and it is plausible that $\boldsymbol{\beta_2}$ is *near* to some specified $\boldsymbol{\beta_2^o}$ which, without loss of generality, may be set to a null vector. The NSI may thus be formulated as

$$\text{NSI} : \boldsymbol{\beta_2} = \mathbf{0} \tag{1.7}$$

In this investigation we are interested in the robust statistical estimation of the parameter vector $\boldsymbol{\beta}$ when the UPI or NSI in (1.7) is available. In the present study, emphasis is on a situation where the sample size is large, while the parameter vector $\boldsymbol{\beta_2}$ is taken to be close to $\boldsymbol{\beta_2^o}$, in this case, 0. We shall study the large sample properties of the proposed estimators, in the light of a quadratic loss function.

The plan of the paper is as follows. In section 2, along with preliminary notation and basic assumptions, the estimators are formally introduced. Expressions for the asymptotic distributional quadratic bias (ADB) and asymptotic distributional quadratic risks (ADR) of the estimators under local alternatives are then obtained in section 3. Comparative risk performances are then provided in section 4. Section 5 provides simulation results for the proposed estimators based on an example involving a demand for money function. Section 6 concludes. Throughout this paper, boldface symbols will represent vector/matrix quantities.

2. Background and Proposed Estimation

Let $\mathbf{C}[\boldsymbol{\beta}] = \mathbf{X}[\boldsymbol{\beta}]^T\mathbf{X}[\boldsymbol{\beta}]$, then denote the product matrix decomposition, (1.6), as follows

$$\mathbf{C}[\boldsymbol{\beta}] = \begin{bmatrix} \mathbf{C}_{11}[\boldsymbol{\beta}] & \mathbf{C}_{12}[\boldsymbol{\beta}] \\ \mathbf{C}_{21}[\boldsymbol{\beta}] & \mathbf{C}_{22}[\boldsymbol{\beta}] \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1[\boldsymbol{\beta}]^T\mathbf{X}_1[\boldsymbol{\beta}] & \mathbf{X}_1[\boldsymbol{\beta}]^T\mathbf{X}_2[\boldsymbol{\beta}] \\ \mathbf{X}_2[\boldsymbol{\beta}]^T\mathbf{X}_1[\boldsymbol{\beta}] & \mathbf{X}_2[\boldsymbol{\beta}]^T\mathbf{X}_2[\boldsymbol{\beta}] \end{bmatrix} = \mathbf{X}[\boldsymbol{\beta}]^T\mathbf{X}[\boldsymbol{\beta}], \tag{2.1}$$

We assume that $\frac{1}{n}\mathbf{C}[\boldsymbol{\beta}] \to \mathbf{Q}[\boldsymbol{\beta}]$ as $n \to \infty$, where $\mathbf{Q}[\boldsymbol{\beta}]$ is a positive definite matrix decomposed as

$$\mathbf{Q}[\boldsymbol{\beta}] = \begin{bmatrix} \mathbf{Q}_{11}[\boldsymbol{\beta}] & \mathbf{Q}_{12}[\boldsymbol{\beta}] \\ \mathbf{Q}_{21}[\boldsymbol{\beta}] & \mathbf{Q}_{22}[\boldsymbol{\beta}] \end{bmatrix}, \quad \mathbf{Q}_{jk}[\boldsymbol{\beta}] = \lim_{n\to\infty} \frac{1}{n}\,\mathbf{C}_{jk}[\boldsymbol{\beta}], \quad j,k = 1,2. \tag{2.2}$$

The unrestricted nonlinear least squares (URN) estimator, $\hat{\boldsymbol{\beta}}$, is obtained as the solution to the system of equations, $[y - x(\hat{\boldsymbol{\beta}})]^T X[\hat{\boldsymbol{\beta}}] = 0$, which is derived from differentiation of the sum of squares function, $S[\boldsymbol{\beta}] = [y - x(\boldsymbol{\beta})]^T[y - x(\boldsymbol{\beta})]$. Given the regularity conditions stated in Section 1, $S[\boldsymbol{\beta}]$ is minimised at $\hat{\boldsymbol{\beta}}$. Note that the URN vector, $\hat{\boldsymbol{\beta}}$, is based on sample data only, and does not incorporate the NSI. However, it may be advantageous to use the available NSI to obtain an improved estimator of $\boldsymbol{\beta}$. The statistical objective is therefore to estimate the parameter vector $\boldsymbol{\beta}$, when NSI is available.

Unrestricted Nonlinear Estimation

The URN estimator, $\hat{\boldsymbol{\beta}}$, can be obtained in a variety of ways, depending on the nature of the problem at hand. A consistent estimator of the variance-covariance matrix of this estimator is given by the variance-covariance estimator in the GNR, (1.4). This variance-covariance matrix is used in construction of the improved (shrinkage) estimators which follow.

Restricted Nonlinear Estimation

The restricted nonlinear least squares (REN) estimator can be obtained by a variety of methods also. Depending on the nature of the NSI, the REN could be a model linear in parameters, or a model nonlinear in parameters but with fewer elements in z_i. In the example which follows in Section 5, both types of NSI are considered. In any event, the REN estimator, $\tilde{\boldsymbol{\beta}}$, is the solution to the system of equations,

$$\{\mathbf{y} - x[\tilde{\boldsymbol{\beta}}]\}^T X[\tilde{\boldsymbol{\beta}}] = 0 \tag{2.3}$$

in which the conditions $\boldsymbol{\beta_2} = 0$ have been imposed as constraints.

It is well known that, for the linear regression model, an estimator subject to linear restrictions (the restricted estimator) is more efficient (or, at least no less efficient) than the unrestricted estimator. However, the restricted estimator will, in general, be biased and inconsistent (Kiefer and Skoog (1984)). Thus, the imposition of false restrictions on some of the parameters of a statistical model generally causes all of the parameters estimates to be biased and inconsistent. The bias does not disappear as the sample size gets larger. Applied econometricians frequently find themselves in this kind of situation. Similar considerations apply in the context of nonlinear models. The econometrician still wishes to estimate $\boldsymbol{\beta}$ without knowing whether or not the NSI, $\boldsymbol{\beta_2} = \mathbf{0}$, is true.

Pre-test Nonlinear Estimation

A natural first estimator to define is the *pre-test or preliminary test nonlinear* (PTN) estimator, which is the URN when an appropriately constructed test statistic lies in a critical region, and takes the value of REN otherwise:

$$\begin{aligned}\hat{\boldsymbol{\beta}}^P &= \begin{cases} \tilde{\boldsymbol{\beta}} & \text{if} \quad \Lambda_n < \lambda_\alpha; \\ \hat{\boldsymbol{\beta}} & \text{if} \quad \Lambda_n \ge \lambda_\alpha, \end{cases} \\ &= \tilde{\boldsymbol{\beta}} \cdot I[\Lambda_n \le \lambda_\alpha] + \hat{\boldsymbol{\beta}} \cdot I[\Lambda_n > \lambda_\alpha].\end{aligned}$$

Here Λ_n is an appropriate test statistic for the null hypothesis $\boldsymbol{\beta_2} = \mathbf{0}$ and λ_α is the critical value for a test of size α given by the null distribution of Λ_n. Thus, $\hat{\boldsymbol{\beta}}^P$ will be the REN estimator, $\tilde{\boldsymbol{\beta}}$, when Λ_n test does not reject the null hypothesis that the restrictions are satisfied, and will be the URN estimator, $\hat{\boldsymbol{\beta}}$ when the test fails to reject that hypothesis.

Large Sample Tests

Since URN and REN are consistent estimators (the latter only when the NSI is true), and not unbiased in general, an appropriate large-sample test statistic, Λ_n, for $\boldsymbol{\beta_2}= \mathbf{0}$ is required. In this subsection we consider such a test statistic for the null hypothesis that

$$H_o : \boldsymbol{\beta_2} = \mathbf{0} \quad \text{against} \quad \boldsymbol{\beta_2} \neq \mathbf{0}. \tag{2.4}$$

An appropriate test statistic for (2.4) can be defined as follows:

$$\Lambda_n = \hat{\boldsymbol{\beta}}_{\mathbf{2}}^{T} [\hat{\sigma}^2 (\mathbf{X}_2^T(\hat{\boldsymbol{\beta}})\mathbf{M}_1(\hat{\boldsymbol{\beta}})\mathbf{X}_2(\hat{\boldsymbol{\beta}}))^{-1}]^{-1} \hat{\boldsymbol{\beta}}_{\mathbf{2}} \tag{2.5}$$

where $\mathbf{M}_1(\hat{\boldsymbol{\beta}}) = \mathbf{I_n} - \mathbf{X}_1(\hat{\boldsymbol{\beta}})(\mathbf{X}_1^T(\hat{\boldsymbol{\beta}})\mathbf{X}_1(\hat{\boldsymbol{\beta}}))^{-1}\mathbf{X}_1^T(\hat{\boldsymbol{\beta}})$ is the orthogonal projection off the span of the columns of $\mathbf{X}_1(\hat{\boldsymbol{\beta}})$. This means that if $(\mathbf{X}^T(\hat{\boldsymbol{\beta}})\mathbf{X}(\hat{\boldsymbol{\beta}}))^{-1}$ is partitioned in the same way as $\mathbf{X}^T(\hat{\boldsymbol{\beta}})\mathbf{X}(\hat{\boldsymbol{\beta}})$, then the lower right block of the of the partitioned inverse is $\mathbf{X}_2^T(\hat{\boldsymbol{\beta}})\mathbf{M}_1(\hat{\boldsymbol{\beta}})\mathbf{X}_2(\hat{\boldsymbol{\beta}}))^{-1}$. Using the earlier notation in (2.1), we define Λ_n as

$$\Lambda_n = \hat{\sigma}^{-2} \hat{\boldsymbol{\beta}}_{\mathbf{2}}^{T} \mathbf{C}_{22.1}(\hat{\boldsymbol{\beta}}) \hat{\boldsymbol{\beta}}_{\mathbf{2}}, \tag{2.6}$$

where $\mathbf{C}_{22.1}(\hat{\boldsymbol{\beta}}) = \mathbf{C}_{22}(\hat{\boldsymbol{\beta}}) - \mathbf{C}_{21}(\hat{\boldsymbol{\beta}})[\mathbf{C}_{11}(\hat{\boldsymbol{\beta}})]^{-1}\mathbf{C}_{12}(\hat{\boldsymbol{\beta}})$. Under the regularity conditions, **A1**–**A9**, assumed earlier the test statistic, Λ_n, is asymptotically distributed as $\chi(r)^2$, under the null hypothesis, where r are the degrees of freedom, or number of restrictions on $\tilde{\boldsymbol{\beta}}$.

James-Stein-type Nonlinear Estimation

We can define the *James-Stein-type nonlinear* (JSN) estimator by

$$\hat{\boldsymbol{\beta}}^{JS} = \tilde{\boldsymbol{\beta}} + \{1 - c\Lambda_n^{-1}\}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}), \tag{2.7}$$

where $c =$ is a *shrinkage constant* chosen in an interval such that $\hat{\boldsymbol{\beta}}^{JS}$ dominates $\hat{\boldsymbol{\beta}}$. In particular, the value of c which minimises the risk function of $\hat{\boldsymbol{\beta}}^{JS}$ is $(n-p)(p-2)/[r(n-p+2)]$ (Judge and Bock, 1978, p. 179) for the finite-sample case, and $(p-2)$ for the asymptotic case.

A well-known difficulty with the (traditional) James-Stein estimator is its tendency to "over-shrink" the resulting estimator "beyond" the unrestricted or maximum likelihood estimator, reversing the sign of the latter. This can occur when Λ_n is very small relative to c, thereby yielding a "shrinkage factor", c/Λ_n which is greater than unity in absolute value. To moderate this effect, the *positive-part James-Stein-type* estimator has been suggested in the literature. Analogously, we define the *positive-part James-Stein-type nonlinear* (PJSN) estimator, as described below.

Positive-part James-Stein-type Nonlinear Estimation

Now, we truncate the JSN in relation (2.7) with its positive-part and to obtain the PJSN as follows:

$$\hat{\boldsymbol{\beta}}^{JSP} = \tilde{\boldsymbol{\beta}} + \{1 - c\Lambda_n^{-1}\}^{+}(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}}), \tag{2.8}$$

where we define the notation $z^{+} = max(0, z)$. This adjustment controls for the over-shrinking problem inherent in $\hat{\boldsymbol{\beta}}^{JS}$.

3. Asymptotic Distributional Results: Bias

The normal theory of $\hat{\boldsymbol{\beta}}^{P}$ and $\hat{\boldsymbol{\beta}}^{JS}$ was considered by Saleh and Han (1990) and Ali and Saleh (1993). Ghosh et. al. (1989) provided the empirical Bayes solution to this problem. The asymptotic properties of least squares variants of $\hat{\boldsymbol{\beta}}^{P}$ and $\hat{\boldsymbol{\beta}}^{JS}$, compared with $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ were considered by Saleh and Sen (1987) and Ahmed (1997). However, the properties of the proposed nonlinear variants of the estimators, $\hat{\boldsymbol{\beta}}^{P}$, $\hat{\boldsymbol{\beta}}^{JS}$ and $\hat{\boldsymbol{\beta}}^{JSP}$ are not available for the problem under consideration. In this paper we shall attempt to provide a comprehensive study of this problem.

Since the test based on Λ_n is consistent against fixed alternatives, all the estimators based on either Stein-rule or the preliminary-test approach become asymptotically isomorphic to $\hat{\boldsymbol{\beta}}$ as $n \to \infty$. Hence we will investigate asymptotic bias and risk under local alternatives, and compare the relative performance of the estimators, URN, REN, PTN, JSN and PJSN. Specifically, we consider a sequence $\{K_n\}$ of local alternatives defined by

$$K_n : \boldsymbol{\beta_2^n} = \boldsymbol{\beta_2^o} + \frac{\boldsymbol{\delta}}{n^{\frac{1}{2}}}, \tag{3.1}$$

where $\boldsymbol{\delta}$ is a real fixed vector and $\boldsymbol{\delta} = (\delta_1, \cdots, \delta_{(r)})^T \in \Re^{(k-p)}$. Note that $\boldsymbol{\delta} = \mathbf{0}$ implies $\boldsymbol{\beta_2^n} = \boldsymbol{\beta_2^o}$, so (2.4) is a special case of $\{K_n\}$. Under local alternatives, $\{K_n\}$, the following theorem facilitates computation of the ADB and, later, the ADR of the estimators outlined above.

Theorem 3.1 Under $\{K_n\}$ and the usual regularity conditions, as n increases, Λ_n follows a non-central χ^2 distribution with r degrees of freedom and non-centrality parameter:

$$\Delta = \frac{\boldsymbol{\delta}^T \mathbf{Q}_{22.1}(\boldsymbol{\beta^o})\boldsymbol{\delta}}{\sigma^2}, \quad \mathbf{Q}_{22.1}(\boldsymbol{\beta^o}) = \lim_{n\to\infty} \frac{1}{n}\, \mathbf{C}_{22.1}(\boldsymbol{\beta^o}) \tag{3.2}$$

Using Theorem 3.1, and applying results from Judge and Bock (1978) given in Appendix B, the ADB and ADR expressions will be presented in the the following respective theorems.

The ADB of an estimator $\boldsymbol{\beta}^*$ is defined as

$$ADB(\boldsymbol{\beta_1}^*) = \lim_{n\to\infty} E\{n^{\frac{1}{2}}(\boldsymbol{\beta}^* - \boldsymbol{\beta})\}. \tag{3.3}$$

Theorem 3.2 Using the above definition of ADB, under $\{K_n\}$ in (3.1) and the assumed regularity conditions, as $n \to \infty$,

$$\begin{aligned}
\mathrm{ADB}(\hat{\boldsymbol{\beta}}) &= \mathbf{0}, \\
\mathrm{ADB}(\tilde{\boldsymbol{\beta}}) &= \mathbf{Q}_{11}^{-1}(\boldsymbol{\beta}^{o})\mathbf{Q}_{12}(\boldsymbol{\beta}^{o})\boldsymbol{\delta} \\
\mathrm{ADB}(\hat{\boldsymbol{\beta}}^{P}) &= \mathbf{Q}_{11}^{-1}(\boldsymbol{\beta}^{o})\mathbf{Q}_{12}(\boldsymbol{\beta}^{o})\boldsymbol{\delta} H_{k+2}(\chi^2_{k,\alpha};\Delta) \\
\mathrm{ADB}(\hat{\boldsymbol{\beta}}^{JS}) &= (k-2)\mathbf{Q}_{11}^{-1}(\boldsymbol{\beta}^{o})\mathbf{Q}_{12}(\boldsymbol{\beta}^{o})\boldsymbol{\delta} E\left(\chi^{-2}_{k+2}(\Delta)\right) \\
\mathrm{ADB}(\hat{\boldsymbol{\beta}}^{JSP}) &= \mathbf{Q}_{11}^{-1}(\boldsymbol{\beta}^{o})\mathbf{Q}_{12}(\boldsymbol{\beta}^{o})\boldsymbol{\delta}\left[H_{k+2}(k-2;\Delta) + E\left\{\chi^{-2}_{k+2}(\Delta) I\left(\chi^{-2}_{k+2}(\Delta)\right) > (k+2)\right\}\right],
\end{aligned}$$

The notation $H_\nu(x\,;\Delta)$ indicates the noncentral χ^2 distribution function, with non-centrality parameter Δ and ν degrees of freedom. Further, $E\left(\chi_\nu^{-2j}(\Delta)\right) = \int_0^\infty x^{-2j} d\phi_\nu(x\,;\Delta)$.

For the special case of $\mathbf{Q}_{12}(\boldsymbol{\beta}^{o}) = \mathbf{0}$, all the estimators are unbiased and hence they are equivalent to each other with respect to the ADB measure. Due to this fact, we will confine ourselves to the situation where $\mathbf{Q}_{12}(\boldsymbol{\beta}^{o}) \neq \mathbf{0}$, and the remaining discussions follow. In this case, $\hat{\boldsymbol{\beta}}$ is the only asymptotically unbiased estimator of $\boldsymbol{\beta}$, since it is unrelated to the NSI. Furthermore, in order to analyse these bias functions, first we transform them into scalar (quadratic) forms. Thus, we define the *quadratic* ADB (QADB) of an estimator $\boldsymbol{\beta}^*$ of parameter vector $\boldsymbol{\beta}$ by

$$\mathrm{QADB}(\boldsymbol{\beta}^*) = [\mathrm{ADB}(\boldsymbol{\beta}^*)]^T\ \mathbf{Q}_{11.2}(\boldsymbol{\beta}^{o})\ [\mathrm{ADB}(\boldsymbol{\beta}^*)] \tag{3.4}$$

Let $\mathbf{Q}_{11.2}(\boldsymbol{\beta}^{o}) = \mathbf{Q}_{11}(\boldsymbol{\beta}^{o}) - \mathbf{Q}_{12}(\boldsymbol{\beta}^{o})\mathbf{Q}_{22}^{-1}(\boldsymbol{\beta}^{o})\mathbf{Q}_{21}(\boldsymbol{\beta}^{o})$. Then, we can define the QADB for the various estimators as follows:

$$\begin{aligned}
\mathrm{QADB}(\hat{\boldsymbol{\beta}}) =& 0, \\
\mathrm{QADB}(\tilde{\boldsymbol{\beta}}) =& \Delta^\star, \quad \Delta^\star = \boldsymbol{\delta}^T \mathbf{Q}^\star(\boldsymbol{\beta}^{o})\boldsymbol{\delta}, \\
& \mathbf{Q}^\star(\boldsymbol{\beta}^{o}) = \mathbf{Q}_{21}(\boldsymbol{\beta}^{o})\mathbf{Q}_{11}^{-1}(\boldsymbol{\beta}^{o})\mathbf{Q}_{11.2}(\boldsymbol{\beta}^{o})\mathbf{Q}_{11}^{-1}(\boldsymbol{\beta}^{o})\mathbf{Q}_{12}(\boldsymbol{\beta}^{o}) \\
\mathrm{QADB}(\hat{\boldsymbol{\beta}}^{P}) =& \Delta^\star[H_{k+2}(\chi^2_{k,\alpha};\Delta)]^2, \\
\mathrm{QADB}(\hat{\boldsymbol{\beta}}^{JS}) =& (k-2)^2\Delta^\star[E(\chi^{-2}_{k+2}(\Delta))]^2, \\
\mathrm{QADB}(\hat{\boldsymbol{\beta}}^{JSP}) =& \Delta^\star\left[H_{k+2}(k-2;\Delta) + E\{\chi^{-2}_{k+2}(\Delta) I(\chi^{2}_{k+2}(\Delta)) > (k-2))\}\right]^2,
\end{aligned} \tag{3.5}$$

Evidently, the QADB of $\tilde{\boldsymbol{\beta}}$ is an unbounded function of $\Delta^\star$. The magnitude of its bias will depend on the quantity Δ^*.

In order to provide a meaningful comparison of the bias functions of the other estimators, we state the following theorem:

Theorem 3.3 (Courant-Fisher, see Gruber (1998), p. 205) If $\mathbf{B}$ and $\mathbf{D}$ are two positive semi-definite matrices with $\mathbf{D}$ nonsingular, both of order $(m \times m)$, then

$$ch_{\min}(\mathbf{B}\mathbf{D}^{-1}) \leq \frac{\mathbf{x}^T\mathbf{B}\mathbf{x}}{\mathbf{x}^T\mathbf{D}\mathbf{x}} \leq ch_{\max}(\mathbf{B}\mathbf{D}^{-1}) \tag{3.6}$$

where $ch_{\min}(\cdot)$ and $ch_{\max}(\cdot)$ are the smallest and largest eigenvalues of $(\cdot)$ respectively, and $\mathbf{x}$ is a column vector of order $(m \times 1)$. We note that the above lower and upper bounds are equal to the infimum and supremum, respectively, of the ratio $\mathbf{x}^T\mathbf{B}\mathbf{x}/\mathbf{x}^T\mathbf{D}\mathbf{x}$ for $\mathbf{x} \neq \mathbf{0}$. Also, for $\mathbf{D} = \mathbf{I}$, the ratio is known as the Rayleigh quotient for matrix $\mathbf{B}$. As a consequence of Theorem 3.3, we have

$$ch_{\min}(\mathbf{Q}^\star(\boldsymbol{\beta}^o)\mathbf{Q}_{22.1}^{-1}) \leq \frac{\Delta^\star}{\Delta} \leq ch_{\max}(\mathbf{Q}^\star(\boldsymbol{\beta}^o)\mathbf{Q}_{22.1}^{-1}) \tag{3.7}$$

The quadratic bias of $\hat{\boldsymbol{\beta}}^P$ is a function of Δ and α. For fixed α, this function begins at zero, increases to a point, then decreases gradually to zero. As a function of α for fixed Δ, it is a decreasing function of $\alpha \in [0, 1)$, with a maximum value at $\alpha = 0$ and zero at $\alpha = 1$. On the other hand, the quadratic bias of $\hat{\boldsymbol{\beta}}^{JS}$ starts from zero at $\Delta = 0$, increases to a point, then decreases towards zero, since $E\left(\chi_{k+2}^{-2}(\Delta)\right)$ is a decreasing, log-convex function of Δ. The quadratic bias curve of $\hat{\boldsymbol{\beta}}^{JSP}$ remains below the curve of $\hat{\boldsymbol{\beta}}^{JS}$ for all values of Δ.

4. Asymptotic Distributional Results: Risk

For the purposes of ADR and loss, we confine our treatment to the case of a loss function of the following form:

$$L(\boldsymbol{\beta}^*, \boldsymbol{\beta}; \mathbf{W}) = a^* n(\boldsymbol{\beta}^* - \boldsymbol{\beta})^T \mathbf{W}(\boldsymbol{\beta}^* - \boldsymbol{\beta}), \tag{3.8}$$

where $\mathbf{W}$ is positive semi-definite weighting matrix and a^* is a positive scalar constant. Such functions are generally called *weighted loss functions*. Then, the expected loss function is defined:

$$E[L(\boldsymbol{\beta}^*, \boldsymbol{\beta}; \mathbf{W})] = R(\boldsymbol{\beta}^*, \boldsymbol{\beta}; \mathbf{W}) \equiv R(\boldsymbol{\beta}^*, \boldsymbol{\beta}) \equiv R(\boldsymbol{\beta}^*), \tag{3.9}$$

which is called the *risk function*. The risk function can be rewritten as

$$\begin{aligned} R(\boldsymbol{\beta}^*, \boldsymbol{\beta}; \mathbf{W}) =& nE\{(\boldsymbol{\beta}^* - \boldsymbol{\beta})^T \mathbf{W}(\boldsymbol{\beta}^* - \boldsymbol{\beta})\} \\ =& n \operatorname{trace}\left[\mathbf{W}\{E(\boldsymbol{\beta}^* - \boldsymbol{\beta})(\boldsymbol{\beta}^* - \boldsymbol{\beta})^T\}\right] \\ =& \operatorname{trace}(\mathbf{W}\boldsymbol{\Gamma}), \end{aligned} \tag{3.10}$$

where $\boldsymbol{\Gamma}$ is the asymptotic covariance matrix of $\boldsymbol{\beta}^*$ and $a^* = 1$.

Further, $\boldsymbol{\beta}^*$ will be termed an *inadmissible estimator* of $\boldsymbol{\beta}$ if there exists an alternative estimator, $\boldsymbol{\beta}^{**}$, such that

$$\mathcal{R}(\boldsymbol{\beta}^{**}, \boldsymbol{\beta}) \leq \mathcal{R}(\boldsymbol{\beta}^*, \boldsymbol{\beta}) \quad \text{for all} \quad (\boldsymbol{\beta}, \mathbf{W}), \tag{3.11}$$

with strict inequality for some $\boldsymbol{\beta}$. We also say that $\boldsymbol{\beta}^{**}$ dominates $\boldsymbol{\beta}^*$. If, instead of (3.11) holding for every n, we have

$$\lim_{n\to\infty} \mathcal{R}(\boldsymbol{\beta}^{**}, \boldsymbol{\beta}) \leq \lim_{n\to\infty} \mathcal{R}(\boldsymbol{\beta}^*, \boldsymbol{\beta}) \quad \text{for all} \quad \boldsymbol{\beta}, \tag{3.12}$$

with strict inequality for some $\boldsymbol{\beta}$, then $\boldsymbol{\beta}^*$ is termed an *asymptotically inadmissible estimator* of $\boldsymbol{\beta}$. In practice, the expression in (3.12) may be difficult to obtain. Hence, we consider the ADR for the sequence $\{K_n\}$ of local alternatives defined in relation (3.1). Suppose that, under local alternatives, $n^{\frac{1}{2}}(\boldsymbol{\beta}^* - \boldsymbol{\beta})$ has a limiting distribution given by

$$F(\mathbf{y}) = \lim_{n\to\infty} P\{\sqrt{n}(\boldsymbol{\beta}^* - \boldsymbol{\beta}) \leq \mathbf{y}\}, \tag{3.13}$$

which is called the *asymptotic distribution function* ADF of $\boldsymbol{\beta}^*$. Further, let

$$\boldsymbol{\Gamma}^* = \int\int\cdots\int \mathbf{y}\mathbf{y}^T dF(\mathbf{y}), \tag{3.14}$$

be the dispersion matrix which is obtained from the ADF, (3.13). The ADR may then be defined as

$$R(\boldsymbol{\beta}^*;\boldsymbol{\beta}) = \text{trace}(\mathbf{W}\boldsymbol{\Gamma}^*). \tag{3.15}$$

An estimator $\boldsymbol{\beta}^*$ is then said to dominate an estimator $\boldsymbol{\beta}^o$ asymptotically if, $R(\boldsymbol{\beta}^*;\boldsymbol{\beta}) \leq R(\boldsymbol{\beta}^o;\boldsymbol{\beta})$. If, in addition, $R(\boldsymbol{\beta}^*;\boldsymbol{\beta}) < R(\boldsymbol{\beta}^o;\boldsymbol{\beta})$ for at least some $(\boldsymbol{\beta},\mathbf{W})$, then $\boldsymbol{\beta}^*$ strictly dominates $\boldsymbol{\beta}^o$.

Under local alternatives as described in (3.1) and the usual regularity conditions, with $a^* = \sigma^{-2}$, we obtain the ADR functions of the proposed estimators by virtue of the following theorem:

Theorem 3.4

$$\begin{aligned}
\text{ADR}(\hat{\boldsymbol{\beta}}) &= trace(\mathbf{W}\mathbf{Q}_{11.2}^{-1}[\boldsymbol{\beta^o}]),\\
\text{ADR}(\tilde{\boldsymbol{\beta}}) &= trace(\mathbf{W}\mathbf{Q}_{11}^{-1}[\boldsymbol{\beta^o}]) + \boldsymbol{\delta}^T\mathbf{Q}^o\boldsymbol{\delta},\\
\text{ADR}(\hat{\boldsymbol{\beta}}^P) &= trace(\mathbf{W}\mathbf{Q}_{11.2}^{-1}[\boldsymbol{\beta^o}]) - trace(\mathbf{W}\mathbf{Q}_{11.2}^{-1}[\boldsymbol{\beta^o}])H_{k+2}(\chi^2_{k,\alpha};\Delta) +\\
&\quad \boldsymbol{\delta}^T\mathbf{Q}^o\boldsymbol{\delta}[2H_{k+2}(\chi^2_{k,\alpha};\Delta) - \Phi_{k+4}(\chi^2_{k,\alpha};\Delta)]\\
\text{ADR}(\hat{\boldsymbol{\beta}}^{JS}) &= ADR(\hat{\boldsymbol{\beta}}) + \boldsymbol{\delta}^T\mathbf{Q}^o\boldsymbol{\delta}(k^2-4)E\left(\chi^{-4}_{k+4}(\Delta)\right) -\\
&\quad (k-2)trace(\mathbf{W}\mathbf{Q}_{22.1}^{-1}[\boldsymbol{\beta^o}])\{2E\left(\chi^{-2}_{k+2}(\Delta)\right) - (k-2)E\left(\chi^{-2}_{k+4}(\Delta)\right)\},\\
\text{ADR}(\hat{\boldsymbol{\beta}}^{JSP}) &= \text{ADR}(\hat{\boldsymbol{\beta}}^{JS}) + (k-2)trace(\mathbf{W}\mathbf{Q}_{22.1}^{-1}[\boldsymbol{\beta^o}])\Big[2E\left\{\chi^{-2}_{k+2}(\Delta)I(\chi^2_{k+2}(\Delta) \leq (k-2)\right\} -\\
&\quad (k-2)E\left\{\chi^{-4}_{k+2}(\Delta)I(\chi^2_{k+2}(\Delta) \leq (k-2)\right\}\Big] - trace(\mathbf{W}\mathbf{Q}_{22.1}^{-1}[\boldsymbol{\beta^o}])H_{k+2}(k-2;\Delta) +\\
&\quad \boldsymbol{\delta}^T\mathbf{Q}^o\boldsymbol{\delta}\left\{2H_{+2}(k-2;\Delta) - H_{k+4}(k-2;\Delta)\right\} -\\
&\quad (k-2)\boldsymbol{\delta}^T\mathbf{Q}^o\boldsymbol{\delta}\Big[2E\{\chi^{-2}_{k+2}(\Delta))I(\chi^2_{k+2}(\Delta) \leq (k-2)\} -\\
&\quad 2E\{\chi^{-2}_{k+4}(\Delta)I(\chi^2_{k+4}(\Delta) \leq (k-2)\} +\\
&\quad (k-2)E\{\chi^{-4}_{k+4}(\Delta)I(\chi^2_{k+4}(\Delta) \leq (k-2)\}\Big],
\end{aligned}$$

where

$$\mathbf{Q}^o = \mathbf{Q}_{21}(\boldsymbol{\beta^o})\mathbf{Q}_{11}^{-1}(\boldsymbol{\beta^o})\mathbf{W}\mathbf{Q}_{11}^{-1}(\boldsymbol{\beta^o})\mathbf{Q}_{12}(\boldsymbol{\beta^o}), \tag{3.16}$$

Proof. The above relations are obtained by the same arguments as in Section 4.3 of Judge and Bock (1978).

4. Comparisons of ADR Functions

Again, here we discard the case where $\mathbf{Q}_{12}(\boldsymbol{\beta^o}) = \mathbf{0}$. In this situation, $\mathbf{Q}_{11.2}(\boldsymbol{\beta^o}) = \mathbf{Q}_{11}(\boldsymbol{\beta^o})$. Then the ADR of all the estimators are reduced to the ADR of $\hat{\boldsymbol{\beta}}$. Hence, all the estimators are ADR equivalent.

The ADR of $\hat{\boldsymbol{\beta}}$ is unrelated to the NSI and hence does not depend on $\boldsymbol{\delta}$, but the other estimators are functions of $\boldsymbol{\delta}$. With this in mind, we consider a special choice of $\mathbf{W} = \mathbf{Q}_{11.2}(\boldsymbol{\beta^o})$, thereby giving the ADR expressions in Theorem 3.3 the interpretation of a loss function in the metric of the Mahalanobis distance. For such a special choice of $\mathbf{W}$ we have $trace(\mathbf{W}\mathbf{Q}_{11.2}^{-1}[\boldsymbol{\beta^o}]) = p$ and $trace(\mathbf{W}\mathbf{Q}_{11}^{-1}[\boldsymbol{\beta^o}]) = p - trace(\mathbf{Q}^\diamond)$, where $\mathbf{Q}^\diamond = \mathbf{Q}_{12}(\boldsymbol{\beta^o})\mathbf{Q}_{22}^{-1}(\boldsymbol{\beta^o})\mathbf{Q}_{21}(\boldsymbol{\beta^o})\mathbf{Q}_{11}^{-1}(\boldsymbol{\beta^o})$. Thus, we note that the ADR$(\hat{\boldsymbol{\beta}}) = p$, which is constant and independent of $\boldsymbol{\delta} \in \Re_p$.

Now, we compare $\tilde{\boldsymbol{\beta}}$ with $\hat{\boldsymbol{\beta}}$ when the NSI is correct (that is, the null hypothesis is true). In such a case, we have ADR$(\hat{\boldsymbol{\beta}})$ − ADR$(\tilde{\boldsymbol{\beta}}) = trace(\mathbf{Q}^\diamond) > 0$. Hence, when the restriction is correctly specified $\tilde{\boldsymbol{\beta}}$ strictly dominates $\hat{\boldsymbol{\beta}}$. However, when $\boldsymbol{\delta}$ moves away from the null vector, the ADR of $\tilde{\boldsymbol{\beta}}$ monotonically increases and goes to ∞ as $\boldsymbol{\delta}^T\mathbf{Q}^o\boldsymbol{\delta} \to \infty$. Thus, $\tilde{\boldsymbol{\beta}}$ may not behave well when the assumed pivot is different from the specified value of $\boldsymbol{\beta}_2$.

It can also be seen that ADR$(\tilde{\boldsymbol{\beta}}) \leq$ ADR$(\hat{\boldsymbol{\beta}})$ if $\boldsymbol{\delta}^T\mathbf{Q}^o\boldsymbol{\delta} \leq trace(\mathbf{Q}^\diamond)$. Further, by the Courant-Fisher Theorem

$$ch_{min}(\mathbf{Q}^\diamond) \leq \frac{\boldsymbol{\delta}^T\mathbf{Q}^o\boldsymbol{\delta}}{\boldsymbol{\delta}'\mathbf{Q}_{22.1}(\boldsymbol{\beta^o})\boldsymbol{\delta}} \leq ch_{max}(\mathbf{Q}^\diamond). \tag{4.1}$$

Thus, ADR$(\tilde{\boldsymbol{\beta}})$ intersects ADR$(\hat{\boldsymbol{\beta}})$ between the bounds given by

$$\Delta_{\max} = \frac{trace(\mathbf{Q}^\diamond)}{ch_{min}(\mathbf{Q}^\diamond)} \quad \text{and} \quad \Delta_{\min} = \frac{trace(\mathbf{Q}^\diamond)}{ch_{max}(\mathbf{Q}^\diamond)} \tag{4.2}$$

Thus, for

$$\Delta \in \left[0,\ \frac{trace(\mathbf{Q}^\diamond)}{ch_{max}(\mathbf{Q}^\diamond)}\right] \tag{4.3}$$

$\tilde{\boldsymbol{\beta}}$ has smaller risk than that of $\hat{\boldsymbol{\beta}}$. Alternatively, for

$$\Delta \in \left(\frac{trace(\mathbf{Q}^\diamond)}{ch_{min}(\mathbf{Q}^\diamond)},\ \infty\right) \tag{4.4}$$

$\hat{\boldsymbol{\beta}}$ has smaller ADR. Clearly, when Δ moves away from the null vector beyond the value $trace(\mathbf{Q}^\diamond)/ch_{min}(\mathbf{Q}^\diamond)$, the ADR of $\tilde{\boldsymbol{\beta}}$ increases and becomes unbounded. This indicates that the performance of $\tilde{\boldsymbol{\beta}}$ will depend strongly on the reliability of the NSI. The performance of $\hat{\boldsymbol{\beta}}$ is always steady throughout $\Delta \in [0, \infty)$.

In an effort to compare the statistical properties of $\hat{\boldsymbol{\beta}}^P$ with $\hat{\boldsymbol{\beta}}$, note that $\Phi_{k+4}(\chi^2_{k,\alpha}; \Delta) \leq H_{k+2}(\chi^2_{k,\alpha}; \Delta) \leq \Phi_{k+2}(\chi^2_{k,\alpha}; 0) = 1 - \alpha$, for $\alpha \in (0, 1)$ and $\Delta > 0$. The left hand side of the above relation converges to 0 as $\Delta \to \infty$. Also, as $||\boldsymbol{\delta}|| \to \infty \Rightarrow \Delta \to \infty$, then $\Phi_{k+4}(\chi^2_{k,\alpha}; \Delta)$, $\boldsymbol{\delta}'\mathbf{Q}^o\boldsymbol{\delta}H_{k+2}(\chi^2_{k,\alpha}; \Delta)$ and $\boldsymbol{\delta}'\mathbf{Q}^o\boldsymbol{\delta}\Phi_{k+4}(\chi^2_{k,\alpha}; \Delta)$ approach 0, and the risk of $\hat{\boldsymbol{\beta}}^P$ approaches the risk of $\hat{\boldsymbol{\beta}}$. The risk of $\hat{\boldsymbol{\beta}}^P$ is smaller than the risk of $\hat{\boldsymbol{\beta}}$ near the null hypothesis which keeps on increasing crosses the risk of $\hat{\boldsymbol{\beta}}$, reaches maximum then decreases monotonically to the risk of $\hat{\boldsymbol{\beta}}$. Hence a preliminary test approach controls the magnitude of the risk. The dominating condition is given by

$$\text{ADR}(\hat{\boldsymbol{\beta}}^P) \leq \text{ADR}(\hat{\boldsymbol{\beta}}) \quad \text{if} \quad \boldsymbol{\delta}^T \mathbf{Q}^o \boldsymbol{\delta} \leq \frac{trace(\mathbf{W}\mathbf{Q}_{11.2}^{-1}[\boldsymbol{\beta}^o])h(\boldsymbol{\delta})}{2h(\boldsymbol{\delta}) - g(\boldsymbol{\delta})}, \tag{4.5}$$

where

$$h(\boldsymbol{\delta}) = H_{k+2}(\chi^2_{k,\alpha}; \Delta), \quad g(\boldsymbol{\delta}) = \Phi_{k+4}(\chi^2_{k,\alpha}; \Delta). \tag{4.6}$$

There are points in the parameter space for which $\hat{\boldsymbol{\beta}}^P$ is inferior to $\hat{\boldsymbol{\beta}}$ and a sufficient condition for this is that

$$\boldsymbol{\delta}^T \mathbf{Q}^o \boldsymbol{\delta} \in (\frac{trace(\mathbf{W}\mathbf{Q}_{11.2}^{-1}[\boldsymbol{\beta}^o])h(\boldsymbol{\delta})}{2h(\boldsymbol{\delta}) - g(\boldsymbol{\delta})}, \infty) \tag{4.7}$$

Moreover, as α (the level of significance of the pre-test) tends to 1, $\text{ADR}(\hat{\boldsymbol{\beta}}^P)$ tends to $\text{ADR}(\hat{\boldsymbol{\beta}})$.

We find that the performance of the PTN estimator, which combines sample information with NSI depends heavily depend on the correctness of this NSI. The gain in risk can be substantial over unrestricted estimation when NSI is nearly correct. However, $\hat{\boldsymbol{\beta}}^P$ combines the NSI in a superior way to that of $\hat{\boldsymbol{\beta}}^R$, in the sense that the risk of $\hat{\boldsymbol{\beta}}^P$ is a bounded function of the NSI. Though we will later show that $\hat{\boldsymbol{\beta}}^P$ is an inadmissible estimator, it is quite robust with respect to ADR and does not require any extra condition on $\mathbf{Q}^\diamond$, besides the basic one that $\mathbf{Q}^\diamond \neq \mathbf{0}$.

The choice of significance level for the preliminary test is one of the factors that determines the shape of the risk function of the PTN estimator. Hence, the sampling properties of $\hat{\boldsymbol{\beta}}^P$ depend, among other factors, on the size of the test chosen for the pre-test. Unfortunately, this feature is often overlooked in applications. Since the size of test α is under the control of the researcher, there exists a statistical decision problem for choosing α. We refer to Brook (1976), Ahmed (1992) and others for a detailed discussion on the selection of α. Further discussion on this matter is beyond the scope of the present paper.

Remark: None of the three estimators, $\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\beta}}^R$ or $\hat{\boldsymbol{\beta}}^P$ is inadmissible with respect to each other. However, at $\boldsymbol{\delta} = \mathbf{0}$, the risks of the estimators may be ordered according to the magnitude of their risk as follows:

$$\hat{\boldsymbol{\beta}}^R \succ \hat{\boldsymbol{\beta}}^P \succ \hat{\boldsymbol{\beta}}, \tag{4.8}$$

where the notation $\succ$ stands for dominance.

Now, let us consider the JSN estimator in the case of the Mahalanobis distance metric of ADR discussed earlier. We compare $\hat{\boldsymbol{\beta}}^{JS}$ with $\hat{\boldsymbol{\beta}}$. The ADR expressions for these estimators were given in Theorem 3.4. These reveal that $\hat{\boldsymbol{\beta}}^{JS}$ will dominate $\hat{\boldsymbol{\beta}}$ if

$$\boldsymbol{\delta}^T \mathbf{Q}^o \boldsymbol{\delta} \leq \left[\frac{trace(\mathbf{Q}^\diamond)}{(k+2)}\right] \left[\frac{2E\left(\chi^{-2}_{k+2}(\Delta)\right) - (k-2)E\left(\chi^{-2}_{k+4}(\Delta)\right)}{E\left(\chi^{-4}_{k+4}(\Delta)\right)}\right]. \tag{4.9}$$

Then by the use of Courant-Fisher Theorem, the condition in relation (4.9) can be rewritten as:

$$\Delta \leq \left[\frac{trace(\mathbf{Q}^\diamond)}{ch_{max}(\mathbf{Q}^\diamond)}\right]\left[\frac{2E\left(\chi_{k+2}^{-2}(\Delta)\right)-(k-2)E\left(\chi_{k+4}^{-2}(\Delta)\right)}{(k^2-4)E\left(\chi_{k+4}^{-4}(\Delta)\right)}\right], \quad \forall \quad \Delta > 0, \tag{4.10}$$

which in turn requires that $p \geq 3$ and $ch_{max}(\mathbf{Q}^\diamond)/trace(\mathbf{Q}^\diamond) \leq 2/(k+2)$. Thus, $\text{ADR}(\hat{\boldsymbol{\beta}}^{JS}) \leq \text{ADR}(\hat{\boldsymbol{\beta}})$ if the following set of conditions is satisfied:

(i) $p_{\min} = min(q, k)$,

(ii) $ch_{max}(\mathbf{Q}^\diamond) < \frac{1}{2}trace(\mathbf{Q}^\diamond)$,

(iii) $0 \leq (k-2) \leq min\left\{\frac{2trace(\mathbf{Q}^\diamond)}{ch_{max}(\mathbf{Q}^\diamond)} - 4\ , 2(k-2)\right\}$.

Finally, we compare the ADR performance of $\hat{\boldsymbol{\beta}}^{JSP}$ and $\hat{\boldsymbol{\beta}}^{JS}$. We may conclude from the ADR relations for these estimators in Theorem 3.4 that

$$\frac{\text{ADQR}(\hat{\boldsymbol{\beta}}^{JSP})}{\text{ADR}(\hat{\boldsymbol{\beta}}^{JS})} \leq 1, \quad \text{for all} \quad \boldsymbol{\delta}, \tag{4.11}$$

with strict inequality for some $\boldsymbol{\delta}$. Therefore, $\hat{\boldsymbol{\beta}}^{JSP}$ asymptotically dominates $\hat{\boldsymbol{\beta}}^{JS}$ under local alternatives. Hence, $\hat{\boldsymbol{\beta}}^{JSP}$ is also superior to $\hat{\boldsymbol{\beta}}$, and thus:

$$ADR(\hat{\boldsymbol{\beta}}^{JSP}) \leq ADR(\hat{\boldsymbol{\beta}}^{JS}) \leq ADR(\hat{\boldsymbol{\beta}}). \tag{4.12}$$

We observed that the JSN and PJSN estimators combine the sample and NSI in superior way, since these estimators perform better than the $\hat{\boldsymbol{\beta}}$ regardless of the correctness of the NSI. However, the gain in ADR over $\hat{\boldsymbol{\beta}}$ is substantial when the NSI is *nearly* correct. We can also conclude that the proposed estimator, $\hat{\boldsymbol{\beta}}^{JSP}$, is asymptotically superior to $\hat{\boldsymbol{\beta}}^{JS}$ and hence to $\hat{\boldsymbol{\beta}}$. However, the important point here is not the improvement in sense of lowering the ADR by using the positive part of the $\hat{\boldsymbol{\beta}}^{JS}$. More importantly, the components of $\hat{\boldsymbol{\beta}}^{JSP}$ *have the same sign* as that of components of $\hat{\boldsymbol{\beta}}$. In other words, $\hat{\boldsymbol{\beta}}^{JSP}$ does not suffer from the standard problem of James-Stein-type estimators (such as $\hat{\boldsymbol{\beta}}^{JS}$), in that it does not over-shrink beyond $\hat{\boldsymbol{\beta}}$.

5. A Simulated Application: The Demand for Money

One of the simplest examples to which the theoretical results in this paper can be applied is a money demand function with autoregressive errors of order 1 (AR(1)). Parkin and Bade (1992, p. 149) observe that the functional form of the demand for money which typically best fits macroeconomic data is logarithmic in the quantity of real money demanded, depending linearly on log real gross domestic product (GDP) and the level of the interest rate. Of course, there has been much empirical research in this area, employing more sophisticated functional forms than that indicated above. Hendry (1993) provides an exposition on various considerations which one ought to take into account when estimating models of this sort. In light of those considerations, we employ a model which includes a series of lags on all variables indicated above, and where the interest rate enters logarithmically. The basic model then takes the form

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon} \tag{5.1}$$

and

$$\boldsymbol{\epsilon} = \phi\boldsymbol{\epsilon}_{-1} + \boldsymbol{\mu} \tag{5.2}$$

where $\mathbf{Z}$ is an $n \times (p-1)$ matrix of regressors to be defined in more detail below, $\boldsymbol{\gamma}$ is a $(p-1) \times 1$ vector of parameters to be estimated, $\boldsymbol{\epsilon}_{-1}$ is $\boldsymbol{\epsilon}$ lagged one period and ϕ is the autoregressive parameter. The vector $\boldsymbol{\mu} \sim \mathrm{N}[\mathbf{0}, \sigma_\mu^2 I_n]$. Re-writing this in scalar form for observations $t \geq 2, \ldots, n$, allowing for up to $\ell = 8$ lags on each of the regressors in $\mathbf{Z}$, and adjusting for the autoregressive error structure yields a model which is nonlinear in $\boldsymbol{\gamma}$ and ϕ,

$$\begin{aligned} y_t =& (1-\phi)\gamma_1 + \phi y_{t-1} + \sum_{i=0}^{\ell}[(X_{2,t-i} - \phi X_{2,t-i-1})\gamma_{2,i}] + \\ & \sum_{i=0}^{\ell}[(X_{3,t-i} - \phi X_{3,t-i-1})\gamma_{3,i}] + \sum_{i=0}^{\ell}[(X_{4,t-i} - \phi X_{4,t-i-1})\gamma_{4,i}] + \mu_t \end{aligned} \tag{5.3}$$

In the above, we let $y_t = \ln \mathrm{M}_t^d, X_{2,t} = \ln \mathrm{GNP}_t, X_{3,t} = \ln \mathrm{P}_t$, and $X_{4,t} = \ln r_t$.

In the empirical literature dealing with the estimation of models which include a high order of lagged regressors, it is common to consider strategies which can reduce the number of parameters to be estimated, yet maintain a fairly long "memory" of past values which influence y_t. One means of achieving this objective is through the use of a polynomial distributed lag (PDL) process (Almon, 1965). In this simulation, the URN estimator, $\hat{\boldsymbol{\beta}}$, is as indicated in (5.3) above, while the REN estimator, *tilde*$\boldsymbol{\beta}$, is obtained by imposing the restrictions which yield a PDL structure. In particular, let $s = 2$ be the degree of the polynomial to which the parameters, $\gamma_{j,i}, j = 2, \ldots, 4$ and $i = 0, \ldots, \ell$ are restricted. Then $\gamma_{i,j}$ can be written

$$\gamma_{j,i} = \sum_{m=0}^{s} \alpha_{j,m} i^m, j = 2, \ldots, 4 \text{ and } i = 1, \ldots, \ell \tag{5.4}$$

For example, when $j = 2$ and $i = 3$, $\gamma_{2,3} = \alpha_{2,0} + 3 \cdot \alpha_{2,1} + 9 \cdot \alpha_{2,2}$. Thus, since $\ell = 8$, the total number of $\gamma_{j,i}$ parameters is twenty-seven, and these depend on only nine $\alpha_{j,m}$ parameters. This is in fact a set of r linear restrictions on the parameter vector $\boldsymbol{\beta} = [\boldsymbol{\gamma} \mid \phi]^T$. This NSI can be represented in the form

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r} \tag{5.5}$$

where $\mathbf{R}$ is a $q \times k$ matrix with $q < k$ linearly independent rows, and $\mathbf{r} = \mathbf{0}$. $\mathbf{R}$ has three diagonal blocks, $\mathbf{R}_j$, one for each set of parameters $\gamma_{2,i}, \gamma_{3,i}$ and $\gamma_{4,i}$ as follows Fomby, Hill and Johnson, 1984, p141)

$$\mathbf{R}_j = \begin{bmatrix} 1 & -8 & 28 & -56 & 70 & -56 & 28 & -8 & 1 \\ -1 & 6 & -14 & 14 & 0 & -14 & 14 & -6 & 1 \\ 4 & -17 & 22 & 1 & -20 & 1 & 22 & -17 & 4 \\ -4 & 11 & -4 & -9 & 0 & 9 & 4 & -11 & 4 \\ 14 & -21 & -11 & 9 & 18 & 9 & -11 & -21 & 14 \\ -14 & 7 & 13 & 9 & 0 & -9 & -13 & -7 & 14 \end{bmatrix} \tag{5.6}$$

with zeroes elsewhere in $\mathbf{R}$ and two additional columns of zeroes located in a way to allow γ_1 and ϕ to be free of restrictions, even in REN.

The simulation of URN, REN, JSN and PJSN was based around quarterly Canadian macroeconomic data extracted from the CANSIM database

http://datacenter.chass.utoronto.ca/cansim/

for the period from 1967–1998. Since the first eight observations were used to create the lagged observations, this yields a sample size of $n = 120$ data points. The CANSIM series numbers used were B1629 (M1B, currency and chequable deposits) for y_t, D15721 (real GDP at market prices, 1992 dollars) for $X_{2,t}$, D15721 and D15689 (current-valued real GDP at market prices) to obtain the implicit price deflator for $X_{3,t}$ and B14006 (the Bank of Canada rate) for $X_{4,t}$.

The foregoing data were used to obtain initial values for REN, which are given as the second column of Table 5.1. This also yielded a simulation experiment design value for $\sigma_\mu = 0.02$, being the standard error of the nonlinear regression with the actual data. Departures from the null hypothesis implicit in REN were based around what could be viewed as economically meaningful long-run values for the parameters of interest. That is, long-run response elasticities of demand for money of 1.00 with respect to price, 0.75 with respect to income and -0.90 with respect to interest rates. Short-run values consistent with these long-run values are given in column 3 of Table 5.1. The difference between these two vectors yields column 4 of Table 5.1, $\boldsymbol{\delta}$. For the purposes of the simulation experiment, departures from REN were estimated at 200 data points, with the 100'th data point representing the data-generating process implicit in column 3 of Table 5.1.

Given the nature of URN, which includes $y_{t-1}, \mathbf{C}[\boldsymbol{\beta}]$ changes with every replication of the simulation, unlike the standard linear or nonlinear regression model. In addition, it is well-known that the linear variants of JSN and PJSN only have their desirable properties (in terms of risk) when

$$\text{trace}[X^T X]^{-1}/\omega_{\max} > 2 \tag{5.7}$$

where $\omega_{\max}$ is the largest eigenvalue of $[X^T X]^{-1}$ (Judge and Bock, 1976 and Trivedi, 1978). We take these considerations into account when designing the simulation experiment, and in the calculation of the empirical risk functions for each estimator. In particular, we verify that (5.7) is met for each replication, and we normalise the computation of the empirical risk functions by $\mathbf{C}[\boldsymbol{\beta}]$.

The URN, REN, JSN and PJSN estimators were computed based on $N = 500$ replications at REN, and the data-generating processes implicit in Table 5.1, as previously described.

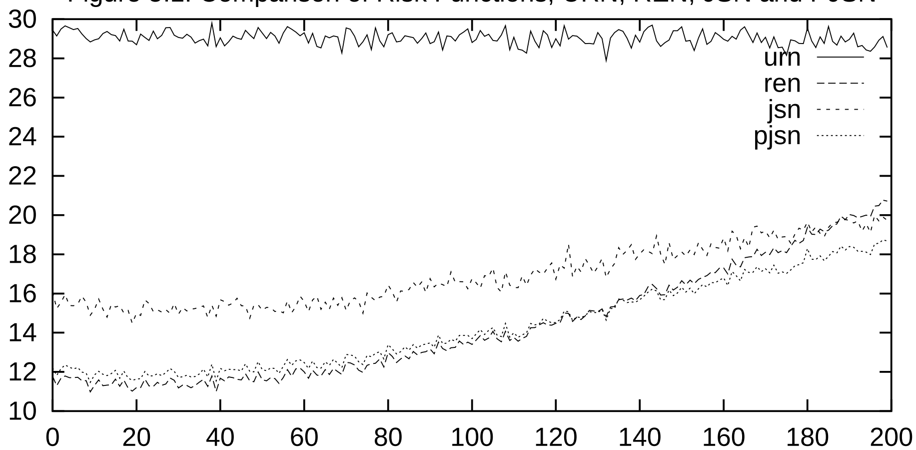
Table 5.1: URN and REN Parameter Values for Simulation Experiments.

Parameter	URN	REN	Step
β_1	-1.9289	-1.9289	0.0000E+00
ϕ	0.8920	0.8920	0.0000E+00
$\beta_{2,0}$	0.4000	-0.0770	0.4770E-02
$\beta_{2,1}$	0.3300	-0.0720	0.4020E-02
$\beta_{2,2}$	0.1500	-0.0586	0.2086E-02
$\beta_{2,3}$	0.0500	-0.0367	0.8673E-03
$\beta_{2,4}$	0.0200	-0.0063	0.2634E-03
$\beta_{2,5}$	0.0125	0.0326	-.2005E-03
$\beta_{2,6}$	0.0125	0.0799	-.6744E-03
$\beta_{2,7}$	0.0125	0.1358	-.1233E-02
$\beta_{2,8}$	0.0125	0.2002	-.1877E-02
$\beta_{3,0}$	0.3500	0.1141	0.2359E-02
$\beta_{3,1}$	0.2000	0.2097	-.9723E-04
$\beta_{3,2}$	0.1500	0.2784	-.1284E-02
$\beta_{3,3}$	0.0300	0.3200	-.2900E-02
$\beta_{3,4}$	0.0100	0.3348	-.3248E-02
$\beta_{3,5}$	0.0025	0.3226	-.3201E-02
$\beta_{3,6}$	0.0025	0.2834	-.2809E-02
$\beta_{3,7}$	0.0025	0.2172	-.2147E-02
$\beta_{3,8}$	0.0025	0.1242	-.1217E-02
$\beta_{4,0}$	-0.6000	-0.0640	-.5360E-02
$\beta_{4,1}$	-0.1500	-0.0488	-.1012E-02
$\beta_{4,2}$	-0.0500	-0.0361	-.1390E-03
$\beta_{4,3}$	-0.0200	-0.0259	0.5899E-04
$\beta_{4,4}$	-0.0100	-0.0182	0.8213E-04
$\beta_{4,5}$	-0.0175	-0.0130	-.4464E-04
$\beta_{4,6}$	-0.0175	-0.0104	-.7131E-04
$\beta_{4,7}$	-0.0175	-0.0102	-.7289E-04
$\beta_{4,8}$	-0.0175	-0.0126	-.4937E-04

Note: The parameter values in the REN column are those implied by a second-order polynomial distributed lag structure. These values are based on the underlying values of: $\alpha_{2,0} = 0.077, \alpha_{2,1} = 0.0007, \alpha_{2,2} = 0.004, \alpha_{3,0} = 0.1141, \alpha_{3,1} = 0.1091, \alpha_{3,2} = -0.0135, \alpha_{4,0} = -0.064, \alpha_{4,1} = -0.0165, \alpha_{4,2} = -0.0013$. These values were obtained as the estimates from estimating the PDL model with the CANSIM data described earlier. They were then used to compute the values in the REN column using equation (5.4). The Step column indicates the incremental departures from REN, reaching the values in URN at step 100.

The ordinary least squares routines in SHAZAM, Version 8 (White, 1978) were used for the estimation, based on a one-step efficient nonlinear estimation algorithm. Elements of the vector, $\boldsymbol{\mu}$, were generated from a pseudo-random Normal distribution with mean zero and $\sigma_\mu = 0.02$ using the IMSL subroutine DRNNOA. The empirical risk functions for each estimator were then computed. These functions are plotted in Figure 5.1.

Figure 5.1: Comparison of Risk Functions, URN, REN, JSN and PJSN



As the earlier analytical results indicate, we see the REN estimator dominate all others for values of $\boldsymbol{\delta}$ close to zero. REN still dominates all estimators at the "economically meaningful" values associated with column 3 of Table 5.1, but deteriorates and is dominated by the PJSN estimator shortly thereafter. Clearly, for the range of parameter values implicit in this simulation, both JSN and PJSN dominate URN by a significant margin, and PJSN dominates REN for a range of parameter values which are economically meaningful.

An extension of these simulation results would be to consider the possibility of computing confidence intervals for the shrinkage estimators along the lines suggested in Kazimi and Brownstone (1999). This topic will be the subject of future research.

6. Concluding Remarks

For a general nonlinear regression model, we have considered various estimation strategies based on preliminary test and Stein-type estimation. It is concluded that the positive-part shrinkage estimator dominates the usual shrinkage estimator. At any rate, both shrinkage estimators perform well relative to the usual unrestricted nonlinear least squares estimator of the parameter vector in the entire parameter space. In contrast, the performance of the estimator based on a preliminary test rule lacks this property. The restricted nonlinear least squares estimator depends heavily on the quality of the NSI. The ADR of the restricted nonlinear least squares estimator is unbounded when the parameter moves far from the subspace of the restriction while $\hat{\boldsymbol{\beta}}^P$ provides good control on the magnitude of the ADR. It is exceedingly important to note that the shrinkage estimators have the smallest possible risk in most cases, as compared to other estimators except when the NSI is nearly correct. Further, the application of shrinkage estimators are subject to condition that $p \geq 3$. Therefore, we recommend the use of $\hat{\boldsymbol{\beta}}^P$ when $p < 3$. Finally, when $p \geq 3$, from the point of robust performance, use of all the estimators may be advocated leaning towards the positive part shrinkage estimator.

References

1. Ahmed, S. E. (1997). Asymptotic Shrinkage Estimation: Regression Case. Applied Statistical Sciences II (ed. M. Ahsanullah). Nova Science: New York, 113-143.

2. Ahmed, S. E.(1994). Improved estimation in a multivariate regression model. Journal of Computational Statistics and Data Analysis, **17**, 537-554.

3. Ahmed, S. E.(1992). Large-sample pooling procedure for correlation. The Statistician **41**, 425-438.

4. Ahmed, S. E. and A.K. Md. E. Saleh (1990). Estimation strategies for the intercept vector in a simple linear multivariate normal regression model. Computational Statistics and Data Analysis, **10**, 193-206.

5. Almon, S. (1965). The Distributed Lag Between Capital Appropriations and Expenditures. *Econometrica*, **33**, p. 178–196.

6. Bancroft, T.A. (1944). On biases in estimation due to the use of preliminary tests of significances. *Ann. Math. Statist.* **15**, 190-204.

7. Bancroft, T.A. and C.P. Han (1977). Inference based on conditional specification: A note and a bibliography. International Statistical Review, **45**, 117-127.

8. Brook, R. J. (1976). On the use of a regret function to set significance points in prior tests of estimation. *JASA* **71**, 126-131.

9. Cohen, A. (1968). A note on the admissibility of pooling in the analysis of variance. *Ann. Math. Statist.* **39**, 1744-1746.

10. Fomby, T.B, R. C. Hill and S.R. Johnson (1984). *Advanced Econometric Methods.* Springer-Verlag, New York.

11. Giles, J. A. and D. E. A. Giles (1993). Pre-test estimation and testing in econometrics: Recent developments. *Journal of Economic Review Surveys* **7**, 145-197.

12. Gruber, M.H.J. (1998), *Improving Efficiency by Shrinkage. The James-Stein and Ridge Regression Estimators.* Marcel Dekker, New York.

13. Han, C.P., C.V. Rao and J. Ravichandran (1988). Inference based on conditional specification: A second bibliography. Communications in Statistics–Theory and Methods **17**, 1945-1964.

14. Hendry, D.F. (1993), *Econometrics. Alchemy or Science?.* Blackwell, Oxford.

15. Judge, G. G. and M. E. Bock (1978). The statistical implication of pre-test and Stein-rule estimators in econometrics. Amsterdam:North Holland.

16. Kazimi, C. and Brownstone, D. (1999) Bootstrap Confidence Bands for Shrinkage Estimators. *Journal of Econometrics*, **90**, 99–127.

17. Parkin, M. and Bade, R. (1992), *Modern Macroeconomics, Third Edition.* Prentice Hall, Toronto.

18. Saleh, A.K.Md.E. and Sen, P.K. (1984). Least squares and rank order preliminary test estimation in general multivariate linear models. Proceedings of the Indian Statist. Inst. Golden Jubilee International Conference in Statistics: Application and New Directions, Calcutta, 1981, 237-253.

19. Sclove, S.L., Morris, C. and Radhakrishnan (1972). Non-optimality of preliminary test estimations of the mean of a multivariate normal distribution. *Ann. Math. Statist.* **43**, 1481-1490.

20. Seber, G.A.F. and Wild, C.J. (1989), *Nonlinear Regression.* Wiley, New York.

21. Sen, P.K. and Saleh, A.K.Md. (1987). On preliminary test and shrinkage M-estimation in linear models. *Ann. Statist.* **15**, 1580-1591.

22. Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proceeding of the fourth Berkeley symposium on Mathematical Statistics and Probability. University of California Press. volume 1, 197-206.

23. Trivedi, P.K. (1978). Estimation of a Distributed Lag Model Under Quadratic Loss. *Econometrica*, **46**, p. 1181–1192.

24. White, K.J. (1978). A General Computer Program for Econometric Methods – SHAZAM. *Econometrica*, **46**, p. 239–240.