Title
A Semi-Parametric Basis for Combining Estimation Problems Under Quadratic Loss

Authors
Judge, George G.
Mittelhammer, Ron C

Publication Date
2003

A Semi-Parametric Basis for Combining Estimation Problems Under Quadratic Loss

George G. Judge[1] and Ron C. Mittelhammer[2]

University of California, Berkeley and Washington State University

Abstract

When there is uncertainty concerning the appropriate statistical model to use in representing the data sampling process and corresponding estimators, we consider a basis for optimally combining estimation problems. In the context of the multivariate linear statistical model, we consider a semi-parametric Stein-like (SPSL) estimator, $\bar{\beta}(\hat{\alpha})$, that shrinks to a random data-dependent vector and, under quadratic loss, has superior performance relative to the conventional least squares estimator. The relationship of the SPSL estimator to the family of Stein estimators is noted and risk dominance extensions between correlated estimators are demonstrated. As an application we consider the problem of a possibly ill-conditioned design matrix and devise a corresponding SPSL estimator. Asymptotic and analytic finite sample risk properties of the estimator are demonstrated. An extensive sampling experiment is used to investigate finite sample performance over a wide range of data sampling processes to illustrate the robustness of the estimator for an array of symmetric and skewed distributions. Bootstrapping procedures are used to develop confidence sets and a basis for inference.

Keywords: Stein-like shrinkage, quadratic loss, ill-conditioned design, semiparametric estimation and inference, data dependent shrinkage vector, asymptotic and finite sample risk.

AMS 1991 Classifications: Primary 62E20

JEL Classifications: Cl0, C24

---

[1] George G. Judge is Professor in the Graduate School, 207 Giannini Hall, University of California, Berkeley, e-mail: judge@are.berkeley.edu,

[2] Ron C. Mittelhammer is Professor of Statistics and Agricultural and Resource Economics at Washington State University, email: mittelha@wsu.edu

[*]The comments and suggestions of M.E. Bock, G. Savin, A. Ullah, the Berkeley econometric research group and three anonymous referees are gratefully acknowledged. We thank Tae-Hwan Kim for his helpful correspondence on the risk dominance issue.

1. INTRODUCTION

In the social sciences much empirical research proceeds in the context of partial-incomplete subject matter theories and data based on experimental designs not devised by or known to the analyst. This leads to uncertainty concerning the statistical model that is appropriate for describing the data sampling process compatible with the observed sample of data. Uncertainty regarding the appropriate statistical model in turn leads to uncertainty regarding appropriate estimation and inference methods. In empirical practice test statistics, tuning parameters, and sometimes magic are invoked to identify a single statistical model on which to base estimation and inference. Selecting one particular statistical model suffers from the possibility that a wrong choice may be made, resulting in a loss of estimation and inference accuracy. Moreover, the validity of eliminating statistical model uncertainty through the specification of a particular parametric formulation depends on information that one generally does not possess.

As one basis for identifying model-estimator uncertainty, Stein (1955) demonstrated the inadmissibility of the conventional maximum likelihood estimator $\boldsymbol{\delta}^{ML}(\mathbf{y}) = \hat{\boldsymbol{\beta}}$ when estimating the multivariate normal mean $\boldsymbol{\beta}$ under quadratic loss. Following this result as a basis for coping with estimator uncertainty James and Stein (1961) and Baranchik (1964) combined the k variate estimator $\hat{\boldsymbol{\beta}}$ with a k dimensional fixed null vector and demonstrated, under the assumption of normality, risk dominating Stein Rule (SR) estimators such as

$$\boldsymbol{\delta}^{S}(\mathbf{y}) = \left(1 - a / \left\|\hat{\boldsymbol{\beta}}\right\|^2\right)\hat{\boldsymbol{\beta}} \ , \tag{1.1}$$

when $(k-2) \le a \le 2(k-2)$. A very general class of estimators that improves on $\hat{\boldsymbol{\beta}}$ follows from Judge and Bock, (1978), Stein (1981) and Brandwein and Strawderman (1991). For the general multivariate normal case the class of pseudo-Bayes-Stein rules having risk less than that of $\hat{\boldsymbol{\beta}}$ is very large (see for example Judge and Bock (1978)). Making use of Stein-like estimators, Sclove, *et al.* (1972)

demonstrated the non-optimality of preliminary test estimators as a basis for dealing with model uncertainty.

In an orthonormal k mean context, Lindley (1962) suggested shrinking $\hat{\boldsymbol{\beta}}$ toward the grand mean estimator and demonstrated the risk dominance of the Stein estimator when $0 \le \mathrm{a} \le 2(\mathrm{k}-3)$. Green and Strawderman (1991) considered a parametric statistical model setting where $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ are independent k-dimensional normally distributed data-based estimators with known covariance matrices $\sigma^2\mathbf{I}_k$ and $\tau^2\mathbf{I}_k$, and demonstrated that the best linear combination of the independent random vector-estimators, under quadratic loss, yields the risk dominating estimator $\boldsymbol{\delta}^{GS}\left(\hat{\boldsymbol{\beta}},\tilde{\boldsymbol{\beta}}\right)=\left(1-\left(k-2\right)\sigma^2 / \| \hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}} \|^2\right)\left(\hat{\boldsymbol{\beta}}-\tilde{\boldsymbol{\beta}}\right)+\tilde{\boldsymbol{\beta}}$. Given this base, Kim and White (2001), provide an expression for the asymptotic risk and bias of Green and Strawderman (GS) Stein-type estimators when the estimators are correlated and demonstrated, for a particular application, shrinkage rules that have smaller asymptotic risk.

Given the uncertainty underlying the model discovery, estimation and inference tasks, and Stein-like possibilities for coping with it, we consider the statistical implications of combining related estimation problems, where the alternative estimators encompassed by alternative models exhibit distinct and dissimilar sampling properties. In the context of the multivariate linear statistical model we demonstrate a data-based *semi-parametric Stein-like* (SPSL) estimator that combines estimation problems by shrinking a base estimator to a plausible alternative estimator. Asymptotic and finite sample risk results are demonstrated and the relationship of the SPSL estimator to the family of Stein Rule (SR) estimators is discussed along with risk dominance properties under normality. As an application of the SPSL estimator we demonstrate the implications of combining two alternative linear statistical models whose associated estimators differ markedly in their bias and precision sampling characteristics. Sampling experiments are used to illustrate the superior finite sampling performance of the SPSL estimator for a variety of normal and non-normal sampling distributions. Bootstrap procedures are used to define and illustrate confidence set performance and a basis for inference.

2. STATISTICAL MODEL AND SEMI-PARAMETRIC STEIN-LIKE ESTIMATOR

Consider the problem of estimating the k dimensional location parameter vector $\boldsymbol{\beta}$ when one observes an n dimensional sample vector $\mathbf{y}$ such that

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2.1}$$

where $\mathbf{X}$ is an $(n \times k)$ matrix of rank k, and $\boldsymbol{\varepsilon}$ is an n dimensional random vector such that $\mathrm{E}[\boldsymbol{\varepsilon}]=\mathbf{0}$ and $\mathrm{cov}(\boldsymbol{\varepsilon})=\sigma^2\mathbf{I}_n$. The scale parameter σ^2 may either be known or unknown and no error distribution assumption need be made other than the existence of second order moments. The objective is to estimate the unknown location vector by some estimator $\boldsymbol{\delta}(\mathbf{y})$ when performance is evaluated by a squared error loss measure $L(\boldsymbol{\beta}, \boldsymbol{\delta}(\mathbf{y})) = \| \boldsymbol{\beta} - \boldsymbol{\delta}(\mathbf{y}) \|^2$. Assuming the usual regularity conditions underlying the linear model, the conventional least squares (LS) estimator is distributed with a mean of $\boldsymbol{\beta}$ and covariance matrix of $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ as $\boldsymbol{\delta}^{\mathrm{LS}}(\mathbf{y}) = \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \sim \left(\boldsymbol{\beta}, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)$, and under quadratic loss, is a minimax estimator with constant risk $\rho(\boldsymbol{\beta}, \hat{\boldsymbol{\beta}}) = \sigma^2 tr(\mathbf{X}'\mathbf{X})^{-1}$.

Assume that in addition to $\hat{\boldsymbol{\beta}}$, an alternative statistical model and corresponding possibly biased data based competing estimator is available,

$$\tilde{\boldsymbol{\beta}} \sim (\boldsymbol{\beta} + \boldsymbol{\gamma}, \boldsymbol{\Phi}), \tag{2.1}$$

where $\boldsymbol{\gamma}$ is a $(k \times 1)$ bias vector and $\boldsymbol{\Phi}$ is a positive definite covariance matrix. We allow the estimators to be correlated and let the covariance matrix of $\left[\hat{\boldsymbol{\beta}}' \vdots \tilde{\boldsymbol{\beta}}'\right]$ be defined by

$$\mathrm{cov}\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \tilde{\boldsymbol{\beta}} \end{bmatrix} = \left[\begin{array}{c|c} \sigma^2(\mathbf{X}'\mathbf{X})^{-1} & \boldsymbol{\Sigma} \\ \hline \boldsymbol{\Sigma}' & \boldsymbol{\Phi} \end{array}\right]. \tag{2.2}$$

Our objective is to identify a weighted linear combination of the two estimators with smaller expected quadratic risk than the LS estimator $\hat{\beta}$. Toward this end, define a new estimator as

$$\bar{\beta}(\alpha) = \alpha\hat{\beta} + (1-\alpha)\tilde{\beta}. \tag{2.3}$$

The quadratic risk or mean square error (MSE) of $\bar{\beta}(\alpha)$ is given by

$$\begin{aligned} MSE(\bar{\beta}(\alpha)) &= E\left[\left[\alpha(\hat{\beta}-\beta)+(1-\alpha)(\tilde{\beta}-\beta)\right]'\left[\alpha(\hat{\beta}-\beta)+(1-\alpha)(\tilde{\beta}-\beta)\right]\right] \\ &= \alpha^2 tr\left(\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)+(1-\alpha)^2\left[tr(\mathbf{\Phi})+\gamma'\gamma\right]+2\alpha(1-\alpha)tr(\mathbf{\Sigma}) \end{aligned} \tag{2.4}$$

In order to minimize $MSE(\bar{\beta}(\alpha))$, the first order necessary condition for α implies

$$\alpha_* = 1 - \frac{\sigma^2 tr(\mathbf{X}'\mathbf{X})^{-1} - tr(\mathbf{\Sigma})}{\gamma'\gamma + \sigma^2 tr(\mathbf{X}'\mathbf{X})^{-1} + tr(\mathbf{\Phi}) - 2tr(\mathbf{\Sigma})} \tag{2.5}$$

Because $\partial^2 MSE(\bar{\beta}(\alpha))/\partial\alpha^2 > 0$ whenever $\hat{\beta}$ and $\tilde{\beta}$ are not perfectly correlated, the optimal weighted linear combination estimator, $\bar{\beta}(\alpha_*) = \alpha_*\hat{\beta} + (1-\alpha_*)\tilde{\beta}$ will, under quadratic loss, be superior to LS.

2.1. Estimating the Optimal α

Since relative to the theoretically optimal α defined in (2.5),

$$\begin{aligned} E\left[(\hat{\beta}-\tilde{\beta})'(\hat{\beta}-\tilde{\beta})\right] &= E\left[(\hat{\beta}-\beta)'(\hat{\beta}-\beta)\right]+E\left[(\tilde{\beta}-\beta)'(\tilde{\beta}-\beta)\right]-2E\left[(\tilde{\beta}-\beta)'(\hat{\beta}-\beta)\right] \\ &= \left[\sigma^2 tr(\mathbf{X}'\mathbf{X})^{-1}\right]+\left[\gamma'\gamma + tr(\mathbf{\Phi})\right]-2\left[tr(\mathbf{\Sigma})\right] \end{aligned} \tag{2.6}$$

and

$$E\left[(\hat{\beta}-\beta)'(\hat{\beta}-\tilde{\beta})\right] = \sigma^2 tr(\mathbf{X}'\mathbf{X})^{-1} - tr(\mathbf{\Sigma}), \tag{2.7}$$

it follows that

$$\alpha_* = 1 - \frac{E\left[\left(\hat{\beta}-\beta\right)'\left(\hat{\beta}-\tilde{\beta}\right)\right]}{E\left[\left(\hat{\beta}-\tilde{\beta}\right)'\left(\hat{\beta}-\tilde{\beta}\right)\right]}. \tag{2.8}$$

It is apparent that $\left(\hat{\beta}-\tilde{\beta}\right)'\left(\hat{\beta}-\tilde{\beta}\right)$ is an unbiased estimator of the expectation term appearing in the denominator of the α_* expression, and it is also consistent under the usual regularity conditions. Regarding the numerator expectation in the expression for α_* in (2.8), substituting the usual unbiased and consistent estimator $S^2 = (n-k)^{-1}\left\|\mathbf{Y}-\mathbf{X}\hat{\beta}\right\|^2$ for σ^2 and an unbiased and/or consistent estimator, $\hat{\mathbf{\Sigma}}$, for $\mathbf{\Sigma}$, defines an estimator of the optimal α weight in the form

$$\hat{\alpha}_* = 1 - \frac{S^2 tr\left(\mathbf{X}'\mathbf{X}\right)^{-1} - tr\left(\hat{\mathbf{\Sigma}}\right)}{\left(\hat{\beta}-\tilde{\beta}\right)'\left(\hat{\beta}-\tilde{\beta}\right)} \tag{2.9}$$

which yields the SPSL estimator

$$\overline{\beta}(\hat{\alpha}) = \hat{\beta} - \frac{\hat{a}}{\left\|\hat{\beta}-\tilde{\beta}\right\|^2}\left(\hat{\beta}-\tilde{\beta}\right) . \tag{2.10}$$

where $\hat{a} = S^2 tr\left(\mathbf{X}'\mathbf{X}\right)^{-1} - tr\left(\hat{\mathbf{\Sigma}}\right)$ acts as an estimate of $a = \sigma^2 tr\left(\mathbf{X}'\mathbf{X}\right)^{-1} - tr\left(\mathbf{\Sigma}\right)$. The estimator, $\overline{\boldsymbol{\beta}}(\hat{\alpha})$, is in the general form of the Stein-rule family of estimators, where shrinkage of the base estimator $\hat{\beta}$ is toward the alternative estimator $\tilde{\beta}$. The estimator is drawn towards the alternative estimator when the variance of the least squares estimator is higher, and drawn towards the least squares estimator when the alternative estimator has higher variance, higher bias, or is more highly correlated with the LS estimator

2.2. First Order Asymptotics

Based on regularity conditions no more stringent than the typical types of conditions assumed for establishing asymptotic properties of the LS estimator, the SPSL estimator also achieves consistency and asymptotic normality. Assume the familiar regularity conditions

$$S^2\left(n^{-1}\mathbf{X}'\mathbf{X}\right)^{-1}\xrightarrow{p}\sigma^2\mathbf{Q}^{-1} \text{ and } n^{-1/2}\mathbf{X}'\boldsymbol{\varepsilon}\xrightarrow{\text{d}}N\left(\mathbf{0},\sigma^2\mathbf{Q}\right) \tag{2.11}$$

so that $\left(\hat{\beta}-\beta\right)$ is $O_p\left(n^{-1/2}\right)$ and $n^{1/2}\left(\hat{\beta}-\beta\right)\xrightarrow{d}N\left(0,\sigma^2\mathbf{Q}^{-1}\right)$, where $n^{-1}\mathbf{X}'\mathbf{X}\to\mathbf{Q}$. Also assume analog conditions on the alternative estimator $\tilde{\beta}$ so that, allowing the bias term to change with n, $\left(\tilde{\beta}-\beta-\gamma_n\right)$ is $O_p\left(n^{-1/2}\right)$, $n\hat{\mathbf{\Phi}}\xrightarrow{\text{p}}\mathbf{\Phi}_0$ and $n\hat{\mathbf{\Sigma}}\xrightarrow{\text{p}}\mathbf{\Sigma}_0$ where $\mathbf{\Phi}_0$ and $\mathbf{\Sigma}_0$ are finite limiting covariance matrices, and $n^{1/2}\left(\tilde{\beta}-\beta-\gamma_n\right)\xrightarrow{d}N\left(\mathbf{0},\mathbf{\Phi}_0\right)$. Given that $\gamma_n\to\gamma_0$, so that $\left(\tilde{\beta}-\hat{\beta}\right)\xrightarrow{p}\gamma_0$, consistency follows from Slutsky's theorems, as[3]

$$p\lim\left(\overline{\beta}\left(\hat{\alpha}\right)\right)=p\lim\left(\hat{\beta}\right)+p\lim\left(\frac{\left[S^2tr\left(\mathbf{X}'\mathbf{X}\right)^{-1}-tr\left(\hat{\mathbf{\Sigma}}\right)\right]}{\left(\hat{\beta}-\tilde{\beta}\right)'\left(\hat{\beta}-\tilde{\beta}\right)}\left(\tilde{\beta}-\hat{\beta}\right)\right)=\beta+p\lim\left(o_p\left(1\right)\right)=\beta.. \tag{2.12}$$

Asymptotic normality follows when $\gamma_0\neq\mathbf{0}$ by first rewriting the SPSL estimator as

$$\begin{aligned} n^{1/2}\left(\overline{\beta}\left(\hat{\alpha}\right)-\beta\right)&=n^{1/2}\left(\hat{\beta}-\beta\right)+n^{1/2}\left[\frac{\left[S^2tr\left(\mathbf{X}'\mathbf{X}\right)^{-1}-tr\left(\hat{\mathbf{\Sigma}}\right)\right]}{\left(\hat{\beta}-\tilde{\beta}\right)'\left(\hat{\beta}-\tilde{\beta}\right)}\left(\tilde{\beta}-\hat{\beta}\right)\right]\\ &=n^{1/2}\left(\hat{\beta}-\beta\right)+\left[\frac{O_p\left(n^{-1/2}\right)}{O_p\left(1\right)}\right]O_p\left(1\right)=n^{1/2}\left(\hat{\beta}-\beta\right)+o_p\left(1\right). \end{aligned} \tag{2.13}$$

Thus $n^{1/2}\left(\overline{\beta}\left(\hat{\alpha}\right)-\beta\right)$ *and* $n^{1/2}\left(\hat{\beta}-\beta\right)$ have the same $N\left(\mathbf{0},\,\sigma^2\mathbf{Q}^{-1}\right)$ limiting distribution. If $\gamma_0=\mathbf{0}$, the limiting distribution of $n^{1/2}\left(\overline{\beta}\left(\hat{\alpha}\right)-\beta\right)$ will be dependent on the joint limiting distribution of

---

[3] Consistency is immediate if $\gamma_0\neq 0$, because $p\lim\left(\overline{\beta}\left(\hat{\alpha}\right)\right)=p\lim\left(\hat{\beta}\right)+\left(\frac{0}{\gamma_0'\gamma_0}\gamma_0\right)=\beta.$ If $\gamma_0=\mathbf{0}$, then

$$p\lim\left(\overline{\beta}\left(\hat{\alpha}\right)\right)=p\lim\left(\hat{\beta}\right)+p\lim\left(\frac{\left[S^2tr\left(n^{-1}\mathbf{X}'\mathbf{X}\right)^{-1}-tr\left(n\hat{\mathbf{\Sigma}}\right)\right]}{n^{1/2}\left(\hat{\beta}-\tilde{\beta}\right)'n^{1/2}\left(\hat{\beta}-\tilde{\beta}\right)}\left(\tilde{\beta}-\hat{\beta}\right)\right)=\beta+p\lim\left(O_p\left(1\right)\cdot o_p\left(1\right)\right)=\beta \text{ given that}$$

$n^{1/2}\left(\hat{\beta}-\tilde{\beta}\right)$ is $O_p\left(1\right)$.

$n^{1/2}\left(\hat{\beta}-\beta\right)$ and $n^{1/2}\left(\tilde{\beta}-\hat{\beta}\right)$ through the relation

$$n^{1/2}\left(\bar{\beta}(\hat{\alpha})-\beta\right)\xrightarrow{d} n^{1/2}\left(\hat{\beta}-\beta\right)+\left[\frac{\left[\sigma^2 tr\left(\mathbf{Q}\right)^{-1}-tr\left(\mathbf{\Sigma}\right)\right]}{n^{1/2}\left(\hat{\beta}-\tilde{\beta}\right)' n^{1/2}\left(\hat{\beta}-\tilde{\beta}\right)} n^{1/2}\left(\tilde{\beta}-\hat{\beta}\right)\right].$$

3. ASYMPTOTIC AND FINITE SAMPLE RISK PERFORMANCES

In order to indicate the potential finite and asymptotic risk performance of the semi-parametric estimator (2.10), we prove a general risk dominance theorem and identify an important relationship between the SPSL estimator and a risk-dominating SR estimator. In particular the result encompasses: *i*) shrinkage toward an estimator that may be asymptotically biased, *ii*) the case where the joint distribution of the estimators may be singular and, *iii*) a result that applies to finite samples, and can be extended to asymptotic results.

3.1. SR Sampling Characteristics and Dominance

Relating to the SPSL estimator (2.10), let the distribution of the estimators $\hat{\beta}$ and $\tilde{\beta}$ be

$$\mathbf{U}=\begin{bmatrix}\mathbf{U}_1\\ \mathbf{U}_2\end{bmatrix}=\begin{bmatrix}\hat{\beta}-\beta\\ \tilde{\beta}-\beta\end{bmatrix}\sim \mathrm{N}(\xi,\Psi)=\mathrm{N}\left[\begin{pmatrix}\mathbf{0}\\ \gamma\end{pmatrix},\left(\begin{array}{c|c}\mathrm{A} & \Sigma\\ \hline \Sigma' & \mathbf{\Phi}\end{array}\right)\right] \tag{3.1}$$

where $\mathbf{U}$ is a $2k\times 1$ random vector and $\mathbf{A}$ and $\mathbf{\Phi}$ are positive definite matrices. Let $\mathbf{J}\equiv[\mathbf{I}\,\vdots-\mathbf{I}]$, and define

$$\mathbf{V}=\mathbf{JU}=\mathbf{U}_1-\mathbf{U}_2\sim \mathrm{N}(-\gamma,\mathbf{J\Psi J}')=\mathrm{N}(-\gamma,\mathrm{A}-\mathbf{\Sigma}-\mathbf{\Sigma}'+\mathbf{\Phi}), \tag{3.2}$$

where we assume that $\mathbf{A}-\mathbf{\Sigma}-\mathbf{\Sigma}'+\mathbf{\Phi}$ is positive definite. Using these definitions we define an SR-type estimator, which is akin to the SPSL estimator in (2.10), as

$$\hat{\delta}(\hat{\beta},\tilde{\beta};\mathrm{c})=\hat{\beta}-\frac{\mathrm{c}}{\left\|\hat{\beta}-\tilde{\beta}\right\|^2}(\hat{\beta}-\tilde{\beta})\,. \tag{3.3}$$

Let $\Xi \equiv \mathbf{J}\ \Psi\mathbf{J}'$, represent Ξ in terms of Cholesky factors as $\Xi = \mathbf{PP}'$ and define

$\mathbf{Z} = \mathbf{P}^{-1}(\mathbf{U}_1 - \mathbf{U}_2) \sim \mathrm{N}(\mu, \mathbf{I}_k)$ where $\mu = \mathbf{P}^{-1}(-\gamma)$ and $\mathbf{R} = \mathbf{P}'\mathbf{P}$. It follows that

$$\hat{\delta}(\hat{\beta},\tilde{\beta};c) - \beta = \hat{\beta} - \beta - \frac{c}{\left\|\hat{\beta}-\tilde{\beta}\right\|^2}(\hat{\beta}-\tilde{\beta}) = \mathbf{U}_1 - \frac{c}{\left\|\mathbf{U}_1 - \mathbf{U}_2\right\|^2}(\mathbf{U}_1 - \mathbf{U}_2) = \hat{\delta}(\mathbf{U}_1, \mathbf{U}_2; c) \quad (3.4)$$

Based on the representation in (3.4), the mean squared error (MSE) of $\hat{\delta}(\hat{\beta},\tilde{\beta};c)$ is

$$\begin{aligned} \mathrm{MSE}(\hat{\delta}(\hat{\beta},\tilde{\beta};c)) &= \mathrm{E}(\hat{\delta}(\hat{\beta},\tilde{\beta};c)-\beta)'(\hat{\delta}(\hat{\beta},\tilde{\beta};c)-\beta) = \mathrm{E}(\hat{\delta}(\mathbf{U}_1,\mathbf{U}_2;c)'\hat{\delta}(\mathbf{U}_1,\mathbf{U}_2;c)) \\ &= \mathrm{tr}(\mathbf{A}) - 2c\ \mathrm{E}\left[\frac{\mathbf{U}_1'\mathbf{V}}{\mathbf{V}'\mathbf{V}}\right] + c^2\mathrm{E}\left[\frac{1}{\mathbf{V}'\mathbf{V}}\right] \\ &= \mathrm{tr}(\mathbf{A}) - 2c\underbrace{\mathrm{E}\left[\frac{\mathbf{U}_1'\mathbf{PZ}}{\mathbf{Z}'\mathbf{RZ}}\right]}_{\eta} + c^2\underbrace{\mathrm{E}\left[\frac{1}{\mathbf{Z}'\mathbf{RZ}}\right]}_{\omega} \\ &= \mathrm{tr}(\mathbf{A}) - 2c\eta + c^2\omega. \end{aligned} \quad (3.5)$$

There is a range of c-values for which $\hat{\delta}(\hat{\beta},\tilde{\beta};c)$ dominates $\hat{\beta}$ in MSE, where $\mathrm{MSE}\left(\hat{\beta}\right) = \mathrm{tr}(\mathbf{A})$, iff there exist nonzero values of c such that $-2c\eta + c^2\omega < 0$. Assuming the existence, and hence positivity, of ω and assuming that η exists and is nonzero, the MSE-dominating range of c is given by

$$c \in \left(\min\{0, 2\eta/\omega\}, \max\{0, 2\eta/\omega\}\right). \quad (3.6)$$

It is clear from (3.5) that the MSE-minimizing choice of the constant c, and the associated minimum MSE of the SPSL estimator, is given by

$$c_* = \eta/\omega \Rightarrow MSE\left(\hat{\delta}(\hat{\beta},\tilde{\beta};c_*)\right) = tr(\mathrm{A}) - \left(\eta^2/\omega\right). \quad (3.7)$$

We emphasize, subject to the aforementioned existence conditions, that both (3.6) and (3.7) apply whether or not the data sampling process is normally distributed. In effect, so long as the MSE of the estimator $\hat{\delta}(\hat{\beta},\tilde{\beta};c_*)$ exists, the estimator is never worse than the base estimator in MSE and will represent a MSE improvement, as is generally the case, when $\eta \neq 0$. Adding the normality assumption

(3.1) allows *sufficient* conditions for the existence of the MSE components ω *and* η to be stated and proved as follows:

Theorem: Under normality, $k \geq 3 \Rightarrow 0 < \omega < \infty$ and $k \geq 5 \Rightarrow |\eta| < \infty$.

MSE Existence Proof:

- $k \geq 3 \Rightarrow 0 < \omega < \infty$

$$\omega = \mathrm{E}\left[\frac{1}{\mathbf{Z'RZ}}\right] = \mathrm{E}\left[\left(\frac{\mathbf{Z'Z}}{\mathbf{Z'RZ}}\right)\left(\frac{1}{\mathbf{Z'Z}}\right)\right] \in \left(\frac{1}{\lambda_L}\mathrm{E}\left[\frac{1}{\mathbf{Z'Z}}\right], \frac{1}{\lambda_S}\mathrm{E}\left[\frac{1}{\mathbf{Z'Z}}\right]\right)$$

where λ_L and λ_S are the positive and finite largest and smallest eigenvalues of the positive definite matrix **R**. Note that $\mathbf{Z'Z} \sim \chi^2(\mathrm{k},\lambda)$, where the noncentrality $\lambda = \mu'\mu/2$, and thus

$$\omega \in \left(\frac{1}{\lambda_L}\mathrm{E}\left[\frac{1}{\chi^2(\mathrm{k},\lambda)}\right], \frac{1}{\lambda_S}\mathrm{E}\left[\frac{1}{\chi^2(\mathrm{k},\lambda)}\right]\right).$$

Note that $0 < \mathrm{E}[(\chi^2(\mathrm{k},\lambda))^{-1}] < \infty \text{ if } \mathrm{k} \geq 3$ because the expectation is a Poisson(λ)-weighted sum of reciprocal expectations defined by $\mathrm{E}\left[(\chi^2_{\mathrm{k}+2\mathrm{j}})^{-1}\right]$, for $\mathrm{j} \geq 0$, and $\mathrm{E}\left[(\chi^2_{\mathrm{k}+2\mathrm{j}})^{-1}\right] = 1/(\mathrm{k}-2+2\mathrm{j})$ (Judge and Bock, p. 315, Theorems A.2.18 and A.2.21). Thus, ω is positive and finite.

- $k \geq 5 \Rightarrow |\eta| < \infty$

Note that $\eta^2 \leq \mathrm{E}\left[(\mathbf{U}_1'\mathbf{PZ})^2\right]\mathrm{E}\left[((\mathbf{Z'RZ})^2)^{-1}\right]$ by the Cauchy-Schwarz inequality. Because $\mathbf{U}_1$ *and* $\mathbf{Z}$ are both normally distributed and **P** is fixed and finite, $\mathrm{E}\left[(\mathbf{U}_1'\mathbf{PZ})^2\right] < \infty$. Along the lines of the preceding proof,

$$\frac{1}{\lambda_L^2}\mathrm{E}\left[\frac{1}{(\mathbf{Z'Z})^2}\right] \leq \mathrm{E}\left[\frac{1}{(\mathbf{Z'RZ})^2}\right] \leq \frac{1}{\lambda_S^2}\mathrm{E}\left[\frac{1}{(\mathbf{Z'Z})^2}\right]$$

where $(\mathbf{Z'Z})^2 \sim (\chi^2(\mathrm{k},\lambda))^2$ and λ_L and λ_S are the positive and finite largest and smallest eigenvalues of the positive definite matrix **R**. Note further that $0 < \mathrm{E}\left[\left(\chi^2(\mathrm{k},\lambda)^2\right)^{-1}\right] < \infty \text{ if } \mathrm{k} \geq 5$ because the expectation is a Poisson(λ)-weighted sum of the values of $\mathrm{E}\left[\frac{1}{(\chi^2_{k+2j})^2}\right] = \frac{1}{(k-2+2j)(k-4+2j)}$ for $j \geq 0$ (Judge and Bock, p. 315, Theorems A.2.20 and A.2.21). Thus η is finite. □

3.2. Feasible Optimal SR Estimation

The optimal SR-type estimator $\hat{\delta}(\hat{\beta}, \tilde{\beta}; c_*)$ is empirically infeasible since it depends on the unknown value of c_* in (3.7). An estimator of c_* can be defined, in the absence of parametric assumptions, by substituting a consistent estimator for the unknown ratio of expectations that defines c_*.[4] Expanding both η and ω in a Taylor series around the expectations of their numerator and denominator components as defined in (3.5), and then suppressing second order and higher terms obtains

$$c_* = \eta / \omega = E\left[\frac{\mathbf{U}_1'\mathbf{V}}{\mathbf{V}'\mathbf{V}}\right] \Big/ E\left[\frac{1}{\mathbf{V}'\mathbf{V}}\right] \approx E\left[\left(\hat{\beta} - \beta\right)' \left(\hat{\beta} - \tilde{\beta}\right)\right] = \sigma^2 tr\left(\mathbf{X}'\mathbf{X}\right)^{-1} - tr\left(\boldsymbol{\Sigma}\right) = a \,. \qquad (3.8)$$

Given the regularity conditions assumed in section 2.2, a consistent estimator of c_* is given by $\hat{c}_* = S^2 tr\left(\mathbf{X}'\mathbf{X}\right)^{-1} - tr\left(\hat{\boldsymbol{\Sigma}}\right)$, resulting in the feasible optimal SR estimator $\hat{\delta}(\hat{\beta}, \tilde{\beta}; \hat{c}_*)$.

3.3. Remarks on SPSL-SR Relationships

Remark 1

The SPSL estimator can now be motivated as an approximate semiparametric optimal SR estimator, $\hat{\delta}(\hat{\beta}, \tilde{\beta}; \hat{c}_*)$. In particular, in its estimated form (2.10), the original SPSL estimator objective of defining a weighted linear combination minimum MSE estimator leads to a first order semiparametric optimal SR estimator. This is apparent upon comparing the first order representation of c_* in (3.8) with

[4] Kim and White (2001) demonstrate a consistent estimator of c_* under the specific parametric assumption of a normally distributed data sampling process. Exploiting specific functional representations of c_* derived from specific parametric distribution assumptions, if actually known to hold, could suggest more efficient estimators of c_* than would be possible in a purely semiparametric context.

the representation of the SPSL estimator given in (2.10) and noting that the SPSL equals the feasible optimal SR, $\hat{\delta}(\hat{\beta},\tilde{\beta};\hat{c}_*)$, where $\hat{c}_* = \hat{a}$.

Remark 2

In a multivariate normal case, and in the application illustrated in section 4 ahead, the data sampling process can be such that the estimators of σ^2 and $\mathbf{\Sigma}$, and thus c_*, are independent of the base and alternative estimators. In these cases the finite sample MSE of the SPSL estimator (2.10) is

$$MSE\left(\overline{\beta}\left(\hat{\alpha}\right)\right) = tr(\mathbf{A}) - 2a\eta + E\left(\hat{a}^2\right)\omega \approx tr(\mathbf{A}) - a^2 \Big/ E\left[\left\|\hat{\beta}-\tilde{\beta}\right\|^2\right]. \tag{3.9}$$

The approximation indicated in (3.9) is based on the analogous type of first order approximation underlying (3.8). In this case the SPSL estimator acts like a SR-type estimator that achieves dominance of the base estimator $\hat{\beta}$ at least to the first order of approximation.

Remark 3

The SPSL estimator is based on an MSE-minimizing estimation objective that is designed to provide MSE improvements over the base estimator in semiparametric contexts in which the data sampling distribution is unspecified and unknown. The SPSL estimation objective is applicable to non-normal data sampling cases, where the SR estimator is not known to provide MSE improvement guarantees. In this sense, the SPSL is an extension of the SR methodology to semiparametric sampling contexts.

Remark 4

We note that in the Green and Strawderman (1991) orthonormal statistical model under normality with covariance matrix assumptions $\mathbf{A} = \sigma^2\mathbf{I}$, $\mathbf{\Phi} = \tau^2\mathbf{I}$, and $\mathbf{\Sigma} = \mathbf{0}$ inserted in (3.1), it follows that $\eta = (k-2)\sigma^2\omega$, and thus the condition $k \geq 3$ is sufficient for a risk dominating choices of c to exist, with the optimal c given by $\eta/\omega = (k-2)\sigma^2$. However, in the more general context of nonzero correlation between the base and alternative estimators, the value of η/ω remains a function of the degree of correlation, and moreover $k \geq 3$ is no longer sufficient to guarantee risk dominance of the SR estimator relative to the base estimator.

4. AN APPLICATION

The estimator proposed in section 3 is applicable in a wide variety of data sampling contexts. Situations involving a structural equation and a potential violation of the orthogonality condition, as well as possible violations in the standard assumptions relating to the noise process itself, are but two examples. To illustrate an application of the SPSL estimator, we focus on generating mean square error improvements relative to the conventional least squares estimator for the linear statistical model in cases that may involve the possibility of ill-conditioned design matrices. In this context, consider as a provisional alternative the polar case of a linear statistical model in which the columns of **X** are orthogonal. This results in a corresponding coordinate-by-coordinate set of LS estimators of the parameters and represents a data dependent alternative estimator that can be combined with the base LS estimator. This alternative estimator differs significantly from the conventional estimator in terms of sampling attributes. For example, when the design matrix is actually nonorthogonal, the coordinate-wise least squares (CLS) estimator is biased, but it has the smallest variances of all subset least squares estimator possibilities.

The CLS estimator can be defined as $\tilde{\boldsymbol{\beta}} = \left[diag\left(\mathbf{X}'\mathbf{X}\right)\right]^{-1} \mathbf{X}'\mathbf{y}$, where $diag\left(\mathbf{X}'\mathbf{X}\right)$ denotes a $\left(k \times k\right)$ diagonal matrix. It is straightforward to show that the covariance matrix of the CLS estimator is given by $\text{cov}\left(\tilde{\boldsymbol{\beta}}\right) = \sigma^2 \left[diag\left(\mathbf{X}'\mathbf{X}\right)\right]^{-1} \mathbf{X}'\mathbf{X} \left[diag\left(\mathbf{X}'\mathbf{X}\right)\right]^{-1} = \boldsymbol{\Phi}$, the bias vector of the CLS estimator is $bias\left(\tilde{\boldsymbol{\beta}}\right) = \left[\left[diag\left(\mathbf{X}'\mathbf{X}\right)\right]^{-1} \mathbf{X}'\mathbf{X} - \mathbf{I}_k \right] \boldsymbol{\beta} = \boldsymbol{\gamma}$, and $\text{cov}\left(\hat{\boldsymbol{\beta}}, \tilde{\boldsymbol{\beta}}\right) = \sigma^2 \left[diag\left(\mathbf{X}'\mathbf{X}\right)\right]^{-1} = \boldsymbol{\Sigma}$. It follows that the CLS estimator is superior to the LS estimator in MSE iff $\boldsymbol{\gamma}'\boldsymbol{\gamma}/\sigma^2 < tr\left(\mathbf{X}'\mathbf{X}\right)^{-1} - tr\left[\left[diag\left(\mathbf{X}'\mathbf{X}\right)\right]^{-1} \right]$.

Upon substituting the preceding definitions of $\boldsymbol{\Phi}, \boldsymbol{\Sigma}$, and $\boldsymbol{\gamma}$ into (2.5), the optimal weight for combining the estimators $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ is

$$\alpha_* = 1 - \frac{\sigma^2 tr\left[(\mathbf{X}'\mathbf{X})^{-1} - \left[diag(\mathbf{X}'\mathbf{X}) \right]^{-1} \right]}{\sigma^2 tr\left[(\mathbf{X}'\mathbf{X})^{-1} - \left[diag(\mathbf{X}'\mathbf{X}) \right]^{-1} \right] + \boldsymbol{\beta}'\left[\mathbf{A}(\mathbf{X}) - \mathbf{I}_k \right]' \left[\mathbf{A}(\mathbf{X}) - \mathbf{I}_k \right] \boldsymbol{\beta}} \tag{4.1}$$

where $\mathbf{A}(\mathbf{X}) \equiv \left[diag(\mathbf{X}'\mathbf{X}) \right]^{-1} \mathbf{X}'\mathbf{X}$.

4.1 On Estimating the Optimal α

The SPSL estimator in the current application is a special case of (2.10) in which $\boldsymbol{\Sigma} = \sigma^2 \left[diag(\mathbf{X}'\mathbf{X}) \right]^{-1}$, which can be unbiasedly and consistently estimated by $\hat{\boldsymbol{\Sigma}} = S^2 \left[diag(\mathbf{X}'\mathbf{X}) \right]^{-1}$. This leads to the following estimator of the optimal combining weight,

$$\hat{\alpha}_* = 1 - \frac{S^2 tr\left[(\mathbf{X}'\mathbf{X})^{-1} - \left[diag(\mathbf{X}'\mathbf{X}) \right]^{-1} \right]}{\left\| \tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}} \right\|^2}, \tag{4.2}$$

and the estimated optimal combination (convex in this case) of $\hat{\boldsymbol{\beta}}$ and $\tilde{\boldsymbol{\beta}}$ given by

$$\bar{\mathbf{b}}(\hat{\alpha}_*) = \hat{\boldsymbol{\beta}} - \left(\frac{S^2 tr\left[(\mathbf{X}'\mathbf{X})^{-1} - \left[diag(\mathbf{X}'\mathbf{X}) \right]^{-1} \right]}{\left\| \tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}} \right\|^2} \right) \left(\hat{\boldsymbol{\beta}} - \tilde{\boldsymbol{\beta}} \right) \tag{4.3}$$

which is a SPSL estimator of the type identified in (2.10) and (3.9).

4.2 Sampling Properties in a Semiparametric Context

Noting that $\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}} = \left[\mathbf{A}(\mathbf{X}) - \mathbf{I}_k \right] \hat{\boldsymbol{\beta}}$, it follows from the definition of the SPSL estimator in (4.3) that the bias of the SPSL estimator is defined by

$$bias\left(\bar{\mathbf{b}}(\hat{\alpha}_*) \right) = E\left[\left(\frac{S^2 tr\left[(\mathbf{X}'\mathbf{X})^{-1} - \left[diag(\mathbf{X}'\mathbf{X}) \right]^{-1} \right]}{\hat{\boldsymbol{\beta}}'\left[\mathbf{A}(\mathbf{X}) - \mathbf{I}_k \right]' \left[\mathbf{A}(\mathbf{X}) - \mathbf{I}_k \right] \hat{\boldsymbol{\beta}}} \right) \left[\mathbf{A}(\mathbf{X}) - \mathbf{I}_k \right] \hat{\boldsymbol{\beta}} \right] \tag{4.4}$$

The bias is a highly nonlinear function of the data, and the expectation defining the bias cannot be calculated in the absence of knowledge about the probability distribution underlying the data sampling process. The covariance matrix, and thus also the representation of the finite sample MSE, presents a similar problem. An estimate of both the bias and covariance matrix can be based on Efron's bootstrap, and as we demonstrate in Section 5 that this estimate can be used to good effect for defining SPSL estimator-based testing and confidence interval generating procedures.

Since this is a special case of the general case demonstrated in Section 2.2, and because the CLS estimator is biased for all n, and in the limit, if **X** is not orthogonal, it follows directly that the SPSL estimator is a consistent estimator of $\boldsymbol{\beta}$. Furthermore the SPSL estimator has a first order asymptotic normal limiting distribution that is identical to the limiting distribution of the $\hat{\boldsymbol{\beta}}$ estimator,

$n^{1/2}\left(\overline{\mathbf{b}}(\hat{\alpha}_*)-\boldsymbol{\beta}\right)\xrightarrow{d} N\left(\mathbf{0},\, \sigma^2\mathbf{Q}^{-1}\right)$.

4.3 Finite Sample Characteristics Under Normality

The MSE of the estimated optimal SPSL in the current application is defined by (3.9), with $\hat{a}=S^2 tr\left[(\mathbf{X}'\mathbf{X})^{-1}-\left(diag(\mathbf{X}'\mathbf{X})\right)^{-1}\right]$ and $a=\sigma^2 tr\left[(\mathbf{X}'\mathbf{X})^{-1}-\left(diag(\mathbf{X}'\mathbf{X})\right)^{-1}\right]$. Assuming that a multivariate normal data sampling process underlies the linear model specification, so that $\hat{\boldsymbol{\beta}}\sim N\left(\boldsymbol{\beta},\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right)$, $S^2\sim Gamma\left((n-k)/2,\, \sigma^2/(n-k)\right)$, and $\hat{\boldsymbol{\beta}}$ and S^2 are independent, the MSE can be represented as

$$MSE\left(\overline{\mathbf{b}}(\hat{\alpha}_*)\right)=\sigma^2 tr(\mathbf{X}'\mathbf{X})^{-1}-2a\eta+\left(\frac{n-k+2}{n-k}\right)a^2\omega \qquad (4.5)$$

where

$$\omega=E\left[\left(\hat{\boldsymbol{\beta}}'\left[\mathbf{A}(\mathbf{X})-\mathbf{I}_k\right]'\left[\mathbf{A}(\mathbf{X})-\mathbf{I}_k\right]\hat{\boldsymbol{\beta}}\right)^{-1}\right] \text{ and } \eta=E\left[\frac{\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)'\left[\mathbf{A}(\mathbf{X})-\mathbf{I}_k\right]\hat{\boldsymbol{\beta}}}{\hat{\boldsymbol{\beta}}'\left[\mathbf{A}(\mathbf{X})-\mathbf{I}_k\right]'\left[\mathbf{A}(\mathbf{X})-\mathbf{I}_k\right]\hat{\boldsymbol{\beta}}}\right]. \qquad (4.6)$$

It is then apparent that the estimated SPSL estimator will be MSE superior to the base estimator $\hat{\boldsymbol{\beta}}$ iff

$$\left(\frac{2(n-k)}{n-k+2}\right)\left(\frac{\eta}{a\omega}\right) > 1. \tag{4.7}$$

While the evaluation of (4.7) depends on the unknown value of $\eta/(a\omega)$, the first order approximation of $\eta/\omega = a$ derived from applying (3.8) results in the MSE superiority condition $\left(\frac{2(n-k)}{n-k+2}\right) > 1$.

Thus, so long as $n \geq 3$, and given that the MSE exists (a sufficient condition for which is $k \geq 5$, as noted in section 3), the SPSL will be superior to at least the first order of approximation.

5. SAMPLING EXPERIMENTS

The definition of the data sampling process begins with the general linear model form

$$Y_i = \mathbf{X}[i,.]\boldsymbol{\beta} + \varepsilon_i = \sum_{j=1}^{5} \beta_j X[i,j] + \varepsilon_i, \text{ for i = 1, …, n.} \tag{5.1}$$

The five-element vector of unknown parameters, $\boldsymbol{\beta}$, is arbitrarily set equal to the vector $[.1 \quad -.1 \quad .2 \quad -.2 \quad .3]'$. The first column of the $n \times 5$ matrix $\mathbf{X}$ is a column of unit values and the remaining columns of the $\mathbf{X}[i,.]'s$ are generated independently from a four-dimensional normal distribution having a mean vector of ones, standard deviations all equal to one, and various levels of pairwise intercorrelation. The outcomes of the ε_i 's were generated independently based on various normal, uniform, beta, and exponential probability distributions, all defined to have zero means over a range of standard deviations. The possible choices for each of the data sampling process characteristics are summarized in Table 5.1. Each of the 800 experiments was repeated m=50,000 times for purposes of calculating stable empirical quadratic risk estimates. The estimated α weights identified in (4.2) were used in the definition of the SPSL estimator.

Table 5.1. Monte Carlo Data Sampling Characteristics

Sampling Characteristic	Possibility Set
n	$\{15, 25, 50, 100\}$
$Corr(x_i, x_j)$	$\{0, .25, .5, .6, .7, .8, .85, .9, .95, .99\}$
$Var(\varepsilon_i)^{1/2}$	$\{.1, .25, 1, 5, 10\}$
$f(\varepsilon)$	Normal, Uniform, Beta, Exponential
$\boldsymbol{\beta}$	$[.1, -.1, .2, -.2, .3]'$

5.1 MSE Sampling Results

The relative empirical quadratic risks of the SPSL, true ideal convex $\bar{\mathbf{b}}(\alpha_*)$ and LS estimators are displayed graphically for the Normal and Exponential distributed data sampling processes in Figures 1 and 2. In the figures, the scenarios are displayed in successive blocks of 10 increasing **X**-correlation levels, each block successively interacted with 5 increasing standard deviation values, and each of the resulting 50 correlation-standard deviation combinations interacted with 4 increasing sample sizes. This resulted in 200 scenarios for each of the sampling distributions. The abscissa values provide a coding for the scenario numbers. The actual numerical MSE results by scenario number for all 800 experiments are presented for all of the sampling distributions in a table that is available from the authors.

When comparing like scenarios, the empirical MSE results were remarkably stable across the different sampling distributions. Figures 1 and 2 depict what were most often the largest differences between results of the four sampling distributions, with the Beta and Uniform distribution results generally more similar to the Normal than the Exponential distribution results. However differences in relative expected quadratic risk were generally very small across the sampling distributions, and dissipated to negligible levels as the data sample size increased.

There are some distinct sampling behavioral patterns that are quite apparent from the empirical sampling results. First of all, except for negligible violations in a few cases of sampling from the Exponential distribution when there was high model fit and small sample sizes, the SPSL estimator empirically dominates the LS estimator in terms of MSE. Given the finite sampling results of Section 4, this was expected when sampling from the normal distribution. What was not expected was the general nature of the superior sampling performance of the SPSL estimator over a wide range of data sampling processes. In cases where model fit was low and/or condition numbers are high, the quadratic risk improvements afforded by SPSL estimator relative to the LS estimator are very substantial.

When model fit decreases, improvements of the SPSL estimator risk relative to LS risk increased monotonically. As data sample sizes increase, the MSE gap between the SPSL and LS estimators decreases, although for many cases the relative risk gains of the SPSL estimator remain substantial. The SPSL estimator exhibits its greatest risk gains over LS risk when the correlations between the X's are at their highest levels.

Comparing the sampling behavior of the $\overline{\mathbf{b}}(\alpha_*)$ and SPSL estimators provides an indication of the cost of not knowing the true value of α_* in (4.2) and having to estimate its value from the sample data. The sampling results indicate that the efficiency gap is considerable when estimates of $\hat{\beta}$, $\tilde{\beta}$ and S^2, used in the construction of the estimator of α_*, are highly variable. The gap narrows as sample size increases. The observed size of the efficiency gap provides motivation for additional work to improve the accuracy of the α_* estimate.

Overall, the SPSL estimator yields substantial risk improvements over the traditional LS estimator in models plagued by poorly conditioned data and poorly fitting models. Moreover, the estimator appears robust to the form of the data sampling distribution, and acts effectively as a semi-parametric variant of Stein-rule type methods.

5.2 Risk versus Parameter and Bias Norms

Another view of the estimation performance of the SPSL estimator is provided by examining relative (to LS) risk performance as a function of the parameter norm ($\beta'\beta$) and bias norm ($\gamma'\gamma$), as displayed in Figures 3 and 4, for samples of size n = 15. As expected, the graphs have many of the characteristics of the risk function for the traditional Stein-like estimator. As either of the norms $\boldsymbol{\beta}'\boldsymbol{\beta}$ or $\boldsymbol{\gamma}'\boldsymbol{\gamma}$ increase, the risk of the SPSL estimator increases and approaches the risk of the LS estimator. It is also interesting to note that as the correlation among the **X**-variables increases the performance of the SPSL estimator improves substantially relative to the LS estimator. Qualitatively, these results were stable over the range of distribution functions examined.

5.3. Testing and Confidence Interval Performance

Even in the absence of specific distributional assumptions, somewhere in the asymptotic sequence, tests and confidence estimation procedures based on the SPSL estimator will begin behaving correctly in terms of size and coverage. We find that if a bias correction is applied, and if a T-distribution is used for establishing critical values, then target size and coverage probabilities, and reasonably good test power, can be achieved even for sample sizes as low as $n = 25$. When the condition number of the design matrix is low and model fit is reasonably good, inference is defensible even for sample sizes of n = 15. In other cases for the smallest sample size, we find that a more computationally intensive nested bootstrap, which calculates critical values of the test statistic from its bootstrapped distribution rather than basing tests on the critical value of the t-distribution, represents a useful testing procedure. We use balanced bootstrapping resampling methods (Efron, 1979; Efron and Tibshirani, 1993; Graham, et. al., 1990) to calculate asymptotically valid variances, covariances, and bias measures (based on a bootstrapped expectation of (4.4)) for calculating outcomes of the "t-statistic"

$$T = \frac{\mathbf{c}\left[\overline{\mathbf{b}}(\hat{\alpha}_*) - \widehat{\mathbf{bias}}\left(\overline{\mathbf{b}}(\hat{\alpha}_*)\right)\right] - r}{\left(\mathbf{c}\,\widehat{\mathbf{Cov}}\left(\overline{\mathbf{b}}(\hat{\alpha}_*)\right)\mathbf{c}'\right)^{1/2}} \tag{5.1}$$

for testing hypotheses of the form $H_o : \mathbf{c}\boldsymbol{\beta} = r$. The bootstrapped LS-estimated residuals are inflated both to account for the inherent over-fitting caused by the least squares fit and to transform the estimated residuals to homoscedastic form (Shao and Tu, 1995), as

$$\hat{\mathbf{e}}_* = \left[diag\left(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\right]^{-1/2} \hat{\mathbf{e}}_{\mathbf{LS}}\,. \tag{5.2}$$

In sampling experiments based on 10,000 data sample replications, 1000 bootstrap resamples to calculate the bias correction terms and SPSL estimator covariance matrix were used to examine the size and power behavior of tests based on the bias-corrected T-statistic and the T-distribution. Representative Q-Q plots displayed in Figures 5 and 6 imply that the use of the tails of the T-distribution to form critical values of tests and to form confidence intervals should perform reasonably well in practice. These results are typical of the sampling results obtained in the case of sampling from a normal error distribution, with the Q-Q plots emulating the T-distribution ever more closely as the sample size increased, as expected. Moreover, the results were similar and robust, qualitatively, when the alternative error distributions (including Exponential, Beta, and Uniform, and additionally examining a T(3) distribution) were simulated. We also tracked the coverage probabilities of upper confidence bounds based on the bias-adjusted T-statistic and found they were quite accurate across all scenarios, especially for 95% and 99% target levels.

Overall, the simulated inference results suggest that the T-type statistic utilizing a bootstrapped bias correction and covariance matrix, and using critical values based on the usual Student-T distribution, exhibits reasonably accurate size and coverage behavior. Moreover, it is competitive with the usual T-type tests based on the LS estimator in terms of power, as illustrated in Figures 7 and 8. The SPSL estimator-based tests exhibit somewhat higher power than LS-based tests and confidence intervals of somewhat smaller length. This is especially true when sample sizes are small and the condition number of the design matrix is high.

We note that selected experimentation with the use of a more computationally intensive nested bootstrap procedure for conducting the "T-test" based on bootstrapped critical values of the test statistic suggested that size accuracy might be improved further for the smaller sample sizes. For example, based on 500 sample repetitions of n = 15, a primary bootstrap sample size of 500, and a nested bootstrap sample size of 200, the .90, .95, and .99 target level quantiles of the nested bootstrapped testing procedure's null distribution actually achieved levels of $\{.894, .942, .992\}$ for testing the significance of the β_2, while comparable observed quantiles based on the T-distribution were $\{.869,.927, .983\}$. Additional work on improving the accuracy of the bootstrap approximation may lead to further increases in the accuracy of inference for small sample sizes, whereas the Monte Carlo evidence suggests that for moderate and larger size samples, the less computationally intensive T-distribution approach may suffice.

6. SUMMARY AND IMPLICATIONS

In this paper, we continue the search started four decades ago by Lindley (1962) for new ways to think about combining data based estimators and thus the possibility of combining estimation problems. In the context of a multivariate linear statistical model, we consider methods for optimally combining, under quadratic loss, estimation problems involving data dependent estimators that have different sampling characteristics. The mean squared error minimizing weighted linear combination estimator that results is a natural basis for specifying a semiparametric estimator that is in the form of, and behaves much like, a Stein-type estimator. In seeking an optimally combined minimum risk estimator, the estimators may be correlated and in general the required set of regularity conditions are no more stringent than those assumed for the traditional LS estimator. Finite sample risk dominance results for the SPSL estimator are demonstrated and the relationship of these results with related Stein estimators is discussed. The general applicability of the SPSL estimator to a range of statistical models and data dependent estimators is emphasized.

As one application possibility we examined the problem of an estimator that combines, in an optimum quadratic loss sense, the LS estimator and the biased, low variance coordinate-wise LS estimator. Analytical finite sample results and sampling experiments provide evidence that the SPSL estimator performs well in both a finite and asymptotic sampling context, and over a range of normal and non-normal data sampling situations. Bootstrapping procedures provide bias-corrected tests and confidence set estimators that appears to work well in practice. In particular we suggest how "T-ratios," based on the SPSL estimator can be bias corrected and the covariance matrix can be estimated, so that inference can proceed in the usual way.

Although the estimation and inference implications of these semiparametric estimators are encouraging, some interesting questions remain. For example, we have used unbiased estimators of the numerator and denominator of the shrinkage parameter α in our application. Using a biased estimator of the shrinkage parameter, as would be the case upon using a minimum quadratic risk estimator of the unknown scale parameter, or requiring positivity, and hence convexity, of the combining weights may produce risk gains. In terms of inference, other alternatives such as nesting the bootstrap to generate data dependent critical values for test and confidence interval construction, rather than using critical values based on the Student T distribution, might produce more accurate size, coverage probabilities, and higher power. We have developed our formulations and evaluations in a two estimation problem context. The formulations can be extended to multiple estimation problems. Research on the statistical implications of these and other estimators combining possibilities for a range of statistical models and distance measures, is ongoing.

REFERENCES

Baranchik, A.M. (1964), "Multiple Regression and Estimation of the Mean of a Multivariate Normal Distribution," Technical Report 51, Stanford University, Dept. of Statistics.

Brandwein, A.C. and Strawderman, W.E. (1991), "Generalizations of the James Stein Estimators Under Spherical Symmetry," *Annals of Statistics*, 19:1639-1650.

Dey, D.K., Ghosh, M. and Strawderman, W.E. (1999), "On Estimation with Balanced Loss," Statistics and Probability Letters, 45:97-101.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics*, 7, 1-26.

Efron, B. and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, London: Chapman and Hall.

Graham, R.L., Hinkley, D.V., John, P.W.M., and Shi, S. (1990), "Balanced Design of Bootstrap Simulations", *J. Royal. Statist. Soc.* , B52, 185-202

Green, E. J., and Strawderman W. E. (1991), "James-Stein-Type Estimator for Combining Unbiased and Possibly Biased Estimators," *Journal of the American Statistical Association*, 86, 1001-1006.

James, W., and Stein, C. (1961), "Estimation With Quadratic Loss," *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability* (vol.1), Berkeley, CA: University of California Press, pp. 361-380.

Judge, G. G., and Bock, M. E. (1978), *The Statistical Implications of Pre-Test and Stein-Rule Estimators in Econometrics*, Amsterdam: North-Holland.

Kim, T. H., and White, H. (2001), "James-Stein type Estimators in Large Samples with application to the Least Absolute Deviation Estimator," *Journal of the American Statistical Association*, 96, 697-705.

Lindley, D. V. (1962), "Discussion of Professor Stein's Paper," *Journal of the Royal Statistical Society*, Series B, 24, 285-288.

Sclove, S.L., Morris, C. and Radhakrishman, (1972), "Non Optimality of Preliminary Test Estimators for the Multinormal Mean" *Annals of Mathematical Statistics*, 43:1481-1490.

Stein, C. M. (1981), "Estimation of the parameters of a multivariate normal distribution," *Annals of Statistic* 9, pp. 1135-1151.

------- (1962), "Confidence Sets for the Mean of a Multivariate Normal Distribution," *Journal of the Royal Statistical Society*, Series B, 24, 265-285.

------- (1955), "Inadmissibility of the Usual Estimator for the Mean of a Multivariate Distribution," in *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability (vol. 1)*, Berkeley, California: University of California Press, pp. 197-206.

Ullah, A. (1990), "Finite Sample Econometrics: A Unified Approach," in *Contributions to Economic Theory and Application: Essays in Honour of A. L. Nagar*, eds. R. A. L. Carter, J. Dutta, and A. Ullah, New York: Springer-Velag, pp. 242-292.

Ullah, A. and Ullah, S., (1978), "Double K-Class Estimators of Coefficients in Linear Regression", *Econometrica*, 46, 705-722.

Figure 1. MSE Relative to LS: Normal DSP

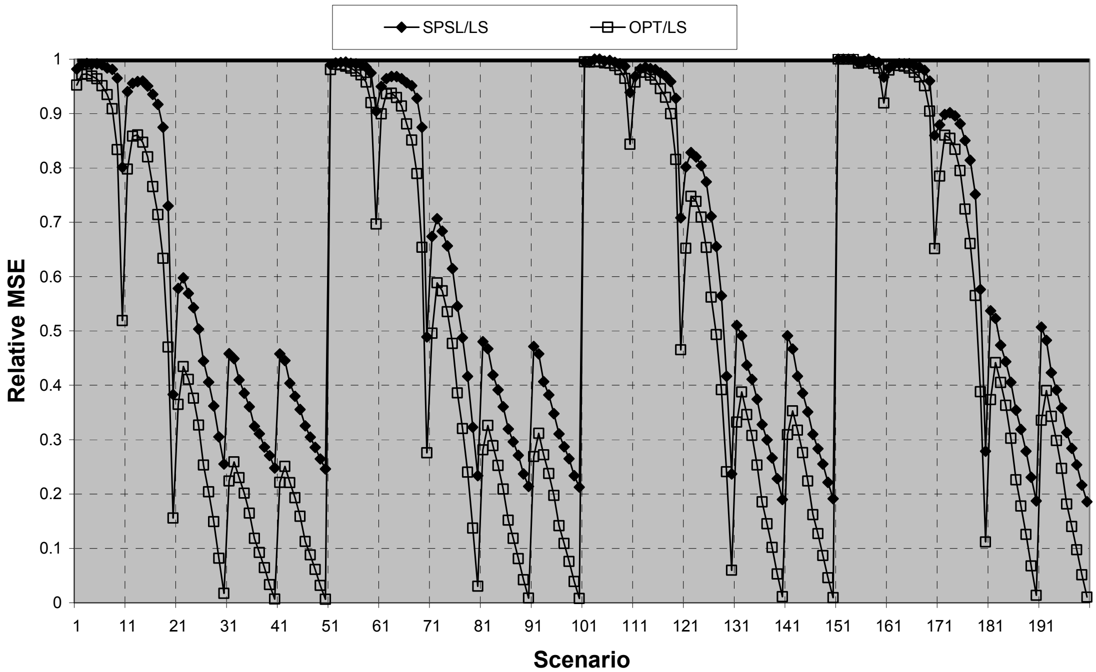

Figure 2. MSE Relative to LS: Exponential DSP

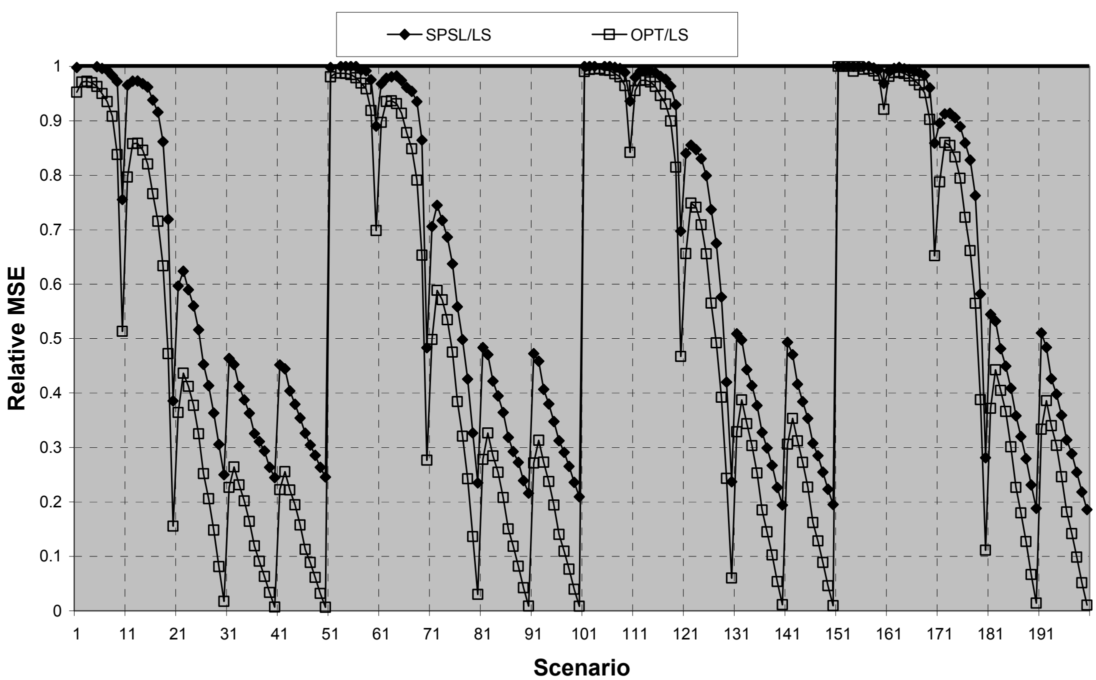

Figure 3. MSE Relative to LS, n = 15, for rho = 0, .75, .95, and .99

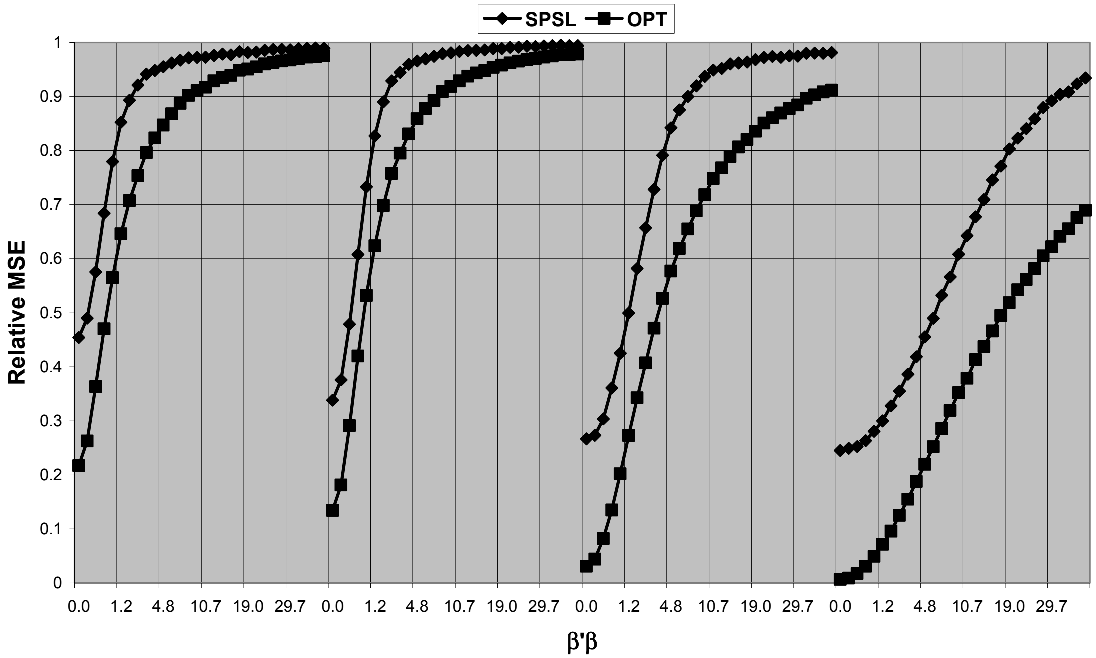

Figure 4. MSE Relative to LS vs Norm of CLS BIAS, n = 15

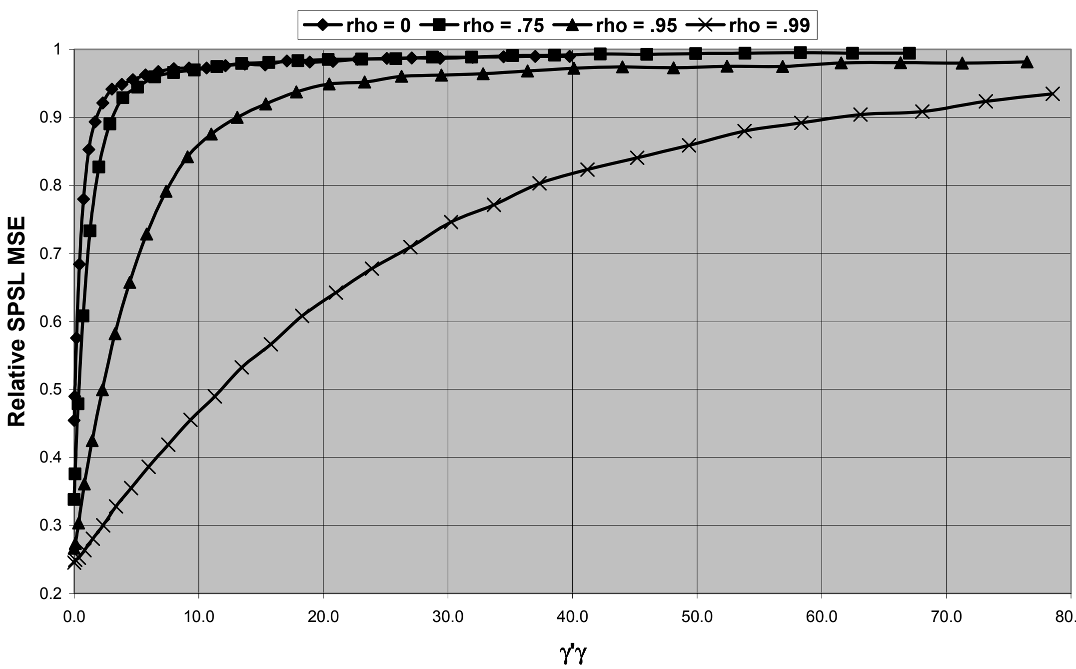

Figure 5. Q-Q Plot for B[2], stde = .1

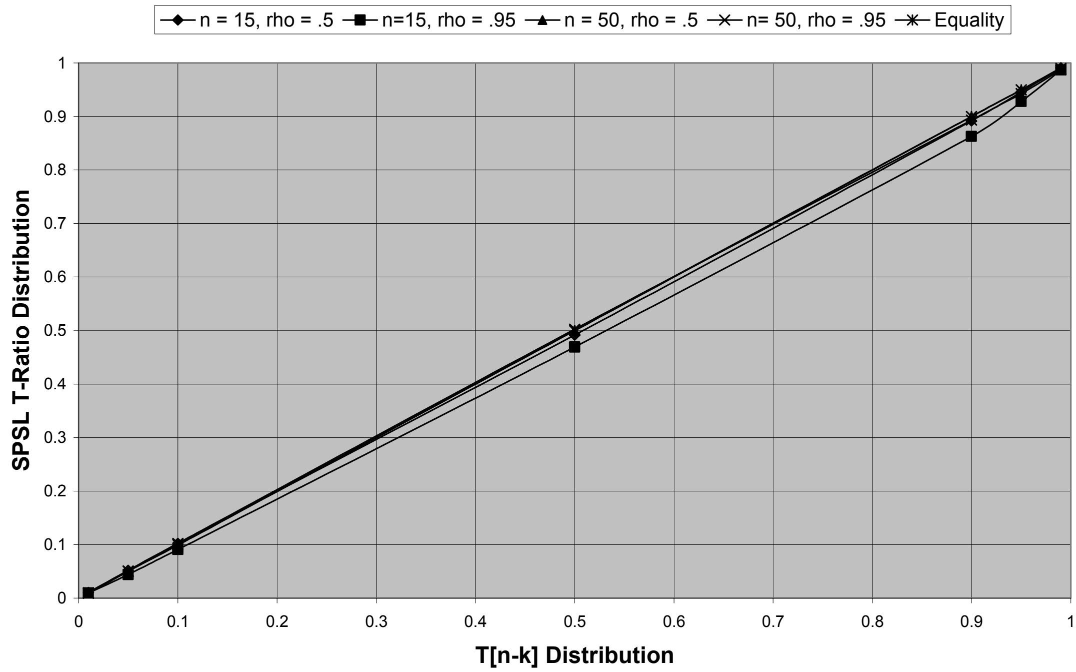

Figure 6. Q-Q Plot for B[5], stde = .25

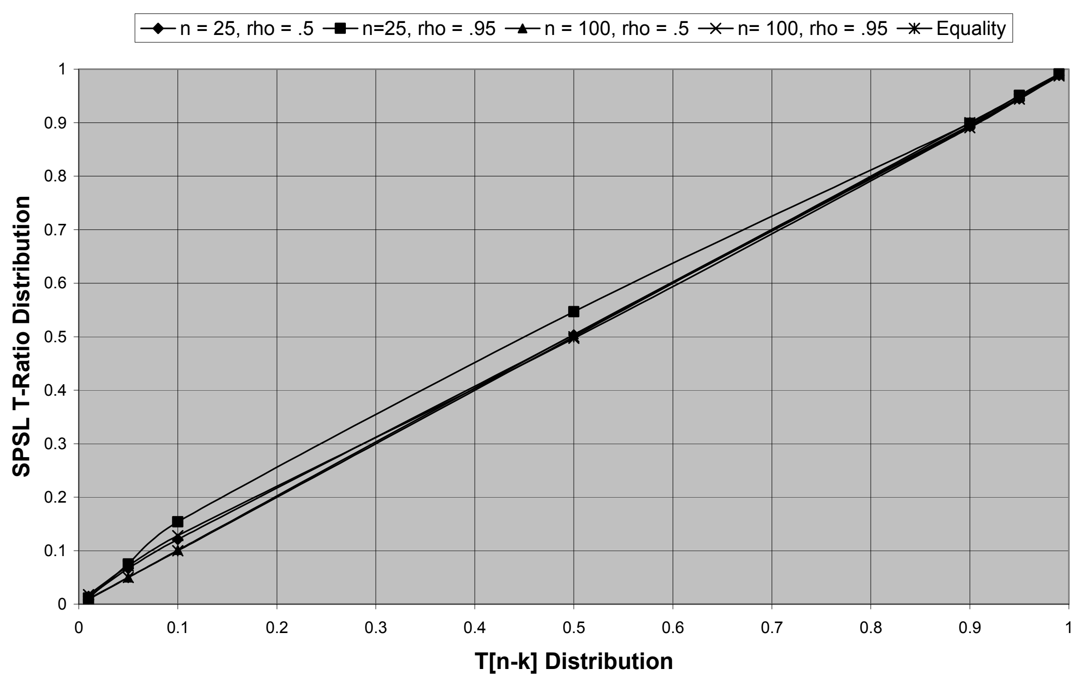

Figure 7. SPSL vs OLS Power for Ho: B[3] = r, n = 15, stde = .25

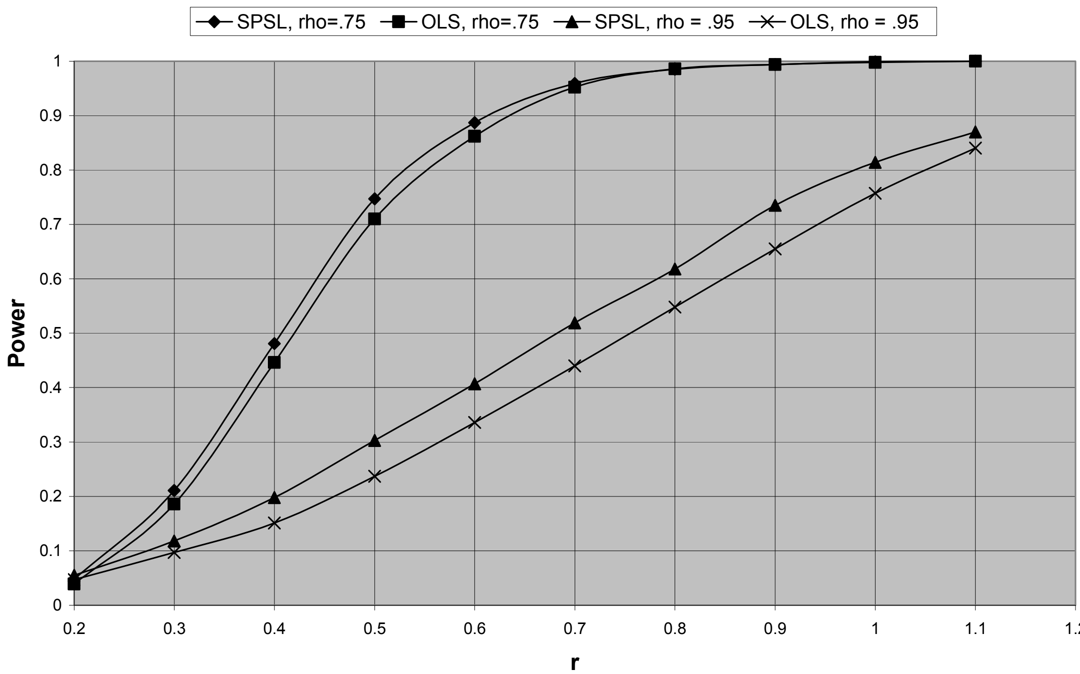

Figure 8. SPSL vs OLS Power for Ho: B[3] = r, n = 50, stde = .25

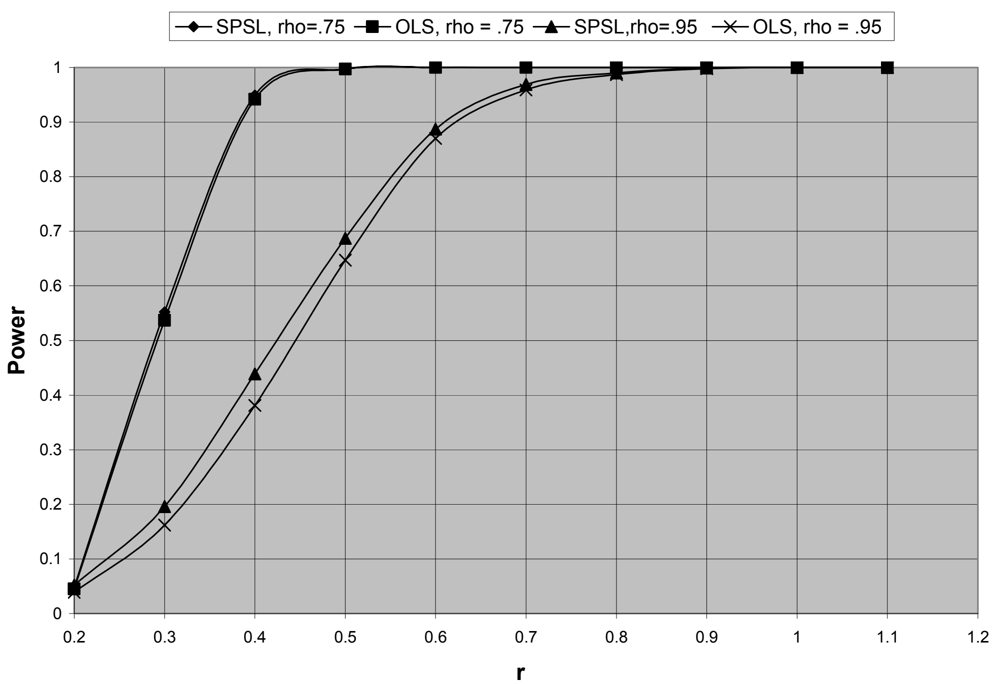

Figure Captions

Figure 1. MSE Relative to LS: Normal DSP

Figure 2. MSE Relative to LS: Exponential DSP

Figure 3. MSE Relative to LS, n=15, for rho = 0, .75, .95, and .99

Figure 4. MSE Relative to LS vs Norm of CLS Bias, n=15

Figure 5. Q-Q Plot for B[2], stde = .1

Figure 6. Q-Q Plot for B[5], stde = .25

Figure 7. SPSL estimator vs OLS Power for Ho: B[3] = r, n = 15, stde = .25

Figure 8. SPSL estimator vs OLS Power for Ho:B[3] = r, n = 50, stde = .25