The RLS Positive-Part Stein Estimator

Lee C. Adkins and R. Carter Hill

The RLS Stein-rule estimator of the classical normal linear regression model is formed by taking a linear combination of the least squares and restricted least squares estimators. Using a simple analytical device, we prove that the convex combination known as the RLS positive-part Stein estimator dominates the conventional version under weighted quadratic loss. Possible uses for the positive-part estimator in economic and agricultural economic research are discussed.

Key words: biased estimation, restricted least squares estimator, Stein-rule estimator.

Although it is well known that positive-part Stein rules dominate ordinary versions in many general settings, such a result has not been obtained for an important family of Stein estimators. We respond to the conjectures of Hill, Ziener, and White; and Mittelhammer and Young, who claim that the Stein rule, which is formed by taking a linear combination of OLS and RLS estimators, is dominated by the convex combination called the positive-part Stein rule. By using a simple analytical device, we prove that the positive-part rule dominates the conventional version under weighted quadratic loss.

In the next section the classical normal linear regression model and its estimators are presented. Then the canonical form for the Stein-rule estimator is derived and the desired result is deduced using the canonical form and results given in Judge and Bock. Finally, the properties and potential uses of the positive-part Stein rule are discussed.

The Model and Its Estimators

The classical normal linear regression model (CNLRM) is represented by

\[ y = X\beta + e \quad e \sim N(0, \sigma^2 I_T), \]

where \( y \) is a \( T \times 1 \) vector of observable random variables, \( X \) is a nonstochastic \( T \times K \) matrix of rank \( K \), \( \beta \) is a \( K \times 1 \) vector of unknown parameters, and \( e \) is a \( T \times 1 \) vector of unobservable normally and independently distributed random variables having zero mean and finite variance \( \sigma^2 \). The ordinary least squares (OLS) and maximum likelihood estimator of \( \beta \) is

\[ \hat{\beta} = (X'X)^{-1}X'y \sim N(\beta, \sigma^2(X'X)^{-1}), \]

and the minimum variance unbiased estimator of \( \sigma^2 \) is

\[ \hat{\sigma}^2 = \frac{1}{T-K}(y - X\hat{\beta})'(y - X\hat{\beta})/(T-K), \]

which is known as the Stein-rule estimator.

Judge and Bock (pp. 240–42) proposed a family of Stein-rule estimators which dominates the MLE of \( \beta \) in the CNLRM under weighted quadratic loss with weight matrix \( W \). Their estimator is a linear combination of the unrestricted and restricted MLEs and has the form

\[ \delta(b) = (1 - c/u)b + (c/u)b^*, \]

where \( u = (Rb - r)'(RS^{-1}R')^{-1}(Rb - r)/J\sigma^2 \sim F_{J,T-K,A} \) is the conventional F-statistic used to test the hypothesis restrictions \( H_0: R\beta = r, S = X'X \); \( R \) is a known \( J \times K \) nonstochastic matrix of rank \( J \); \( r \) is a \( J \times 1 \) vector of known constants; \( b^* = b - S^{-1}R'(RS^{-1}R')^{-1}(Rb - r) \) is the restricted least squares estimator (RLS); \( \lambda = (Rb - r)'(RS^{-1}R')^{-1}(Rb - r)/2\sigma^2 \) is the noncentrality parameter; and \( c = a(T - K)/J \). The estimator is minimax if the scalar \( a \) is chosen to lie within the interval \([0, a_{\text{max}}]\), where

\[ a_{\text{max}} = \left[2/(T - K + 2)\right]\lambda^{-1}\sigma^2 \text{tr}[(RS^{-1}R')^{-1}RS^{-1}WS^{-1}R' - 2], \]

and \( \lambda \) is the largest characteristic root of \([RS^{-1}R']^{-1}RS^{-1}WS^{-1}R'\). The value of the constant \( a \) which minimizes quadratic risk is the interval's midpoint.

Lee C. Adkins is an assistant professor, Department of Economics, Oklahoma State University, and R. Carter Hill is a professor, Department of Economics, Louisiana State University. The authors would like to thank Ron Mittelhammer for his careful reading of our original manuscript and for kindly suggesting a more elegant transformation which we used in the proof.

Copyright 1990 American Agricultural Economics Association
In many circumstances, the usual Stein estimator is dominated by a simple modification called the positive-part rule. The positive-part rule associated with (2) is denoted

\[
\delta^+(b) = \begin{cases} 
b^* & \text{if } c > u \\
\delta(b) & \text{if } c \leq u.
\end{cases}
\]  

Both Hill, Ziemer, and White; and Mittelhammer and Young correctly claim that the usual Stein rule is dominated by its positive part. Below, we will prove that their conjecture is true by showing that (4) dominates (2) under arbitrary quadratic loss using a simple analytical device which can be easily applied in many similar circumstances.

The Canonical Form

To show dominance of (2) by (4), the model and the hypothesis restrictions are written in their canonical forms. Let

\[
y = X\beta + \epsilon = XPP^{-1}\beta = Z\theta + \epsilon,
\]
where \( Z = XP, \theta = P^{-1}\beta, \) and \( P \) is such that \( P'X'XP = I. \) Under the same transformation the restrictions \( R\beta = r \) become \( RPP^{-1}\beta = H_0\theta = r. \) The rank of \( H_1 \) is \( J; \) therefore, there exists a nonsingular matrix, \( C, \) such that \( CH_1\theta = Cr = H\theta = h, \) where \( H = CH_1\theta \) and \( h = Cr. \) The model is now \( y = Z\theta + \epsilon \) subject to the simple equality restrictions \( H\theta = [I, 0]\theta = \theta_j = h. \) The least squares estimator, restricted least squares estimator, and test statistic are \( \hat{\theta} = Z'y \sim N(\theta^*, \sigma^2I_k), \) \( \theta^* = (h', \theta_{kJ}), \) and \( u = (\hat{\theta}_j - h)'(\hat{\theta}_j - h)/\sigma^2, \) respectively, where \( \theta_{kJ} = [0, I_{KJ}]\theta \) and \( \hat{\theta}_j = [I, 0]\hat{\theta}. \)

For the problem at hand, the analytical advantages of this canonical form are substantial. The regression model subject to linear constraints has been transformed into the problem of estimating the mean of a multivariate normal distribution subject to simple equality restrictions \( \theta_j = h. \) It is a simple matter to derive the desired result in this context.

Result

The goal is to prove that (2) is dominated by the positive-part Stein rule given in (4). To obtain this result, we use the transformation described in the previous section along with a known result from the literature on Stein estimation. Judge and Bock (pp. 245–48) prove that \( b^+ \) dominates \( b_+ \) under arbitrary quadratic loss functions where

\[
(6) \quad b_+ = [1 - c/u]b + (c/u)b_0,
\]

\[
(7) \quad b^+ = \begin{cases} 
b_0 & \text{if } c > u \\
b^+ & \text{if } c \leq u,
\end{cases}
\]

\( J = K, \) and \( b_0 \) is a \( K \times 1 \) vector of known constants.

The Stein estimator (2) defined in terms of the transformed parameter space is

\[
(8) \quad \delta(\hat{\theta}) = [1 - (c/u)]\hat{\theta} + (c/u)\theta^*, \text{ or}
\]

\[
(9) \quad \delta(\hat{\theta}) = \begin{bmatrix} 1 - (c/u)\theta_j + (c/u)h \\ \theta_{kJ} \end{bmatrix} = \begin{bmatrix} \delta(\hat{\theta}_j) \\ \theta_{kJ} \end{bmatrix}.
\]

Because \( \hat{\theta}_j \) is statistically independent of \( \theta_{kJ} \) and \( E[\theta_{kJ}] = \theta_{kJ}, \) the quadratic risk of using \( \delta(\hat{\theta}) \) to estimate \( \theta \) is

\[
(10) \quad R(\delta, \theta) = R(\delta(\hat{\theta}_j), \theta_j) + R(\hat{\theta}_{kJ}, \theta_{kJ}).
\]

The risk of using least squares, \( \hat{\theta}, \) to estimate \( \theta \) is

\[
R(\hat{\theta}, \theta) = R(\hat{\theta}_j, \theta_j) + R(\hat{\theta}_{kJ}, \theta_{kJ}).
\]

Thus, \( \delta(\hat{\theta}) \) dominates \( \theta \) if

\[
R(\delta(\hat{\theta}_j), \theta_j) \leq R(\hat{\theta}_j, \theta_j).
\]

Similarly, the positive-part rule, \( \delta^+(\hat{\theta}) \), which is equal to \( \theta^* \) if \( (c/u) > 1 \) and \( \delta(\hat{\theta}) \) otherwise, will dominate \( \delta(\hat{\theta}) \) if

\[
R(\delta^+(\hat{\theta}_j), \theta_j) \leq R(\delta(\hat{\theta}_j), \theta_j).
\]

As given, \( \delta(\hat{\theta}_j) \) is equivalent to (6) with \( K = J, \) and the Judge and Bock result can be applied to this expression directly.

In the context of the transformed model the quadratic loss function associated with use of an arbitrary estimator of \( \theta, \) call it \( \hat{\theta}, \) is

\[
L(\hat{\theta}, \theta, W^*) = (\hat{\theta} - \theta)'W^*(\hat{\theta} - \theta) = (\beta - \beta)'P^{-1}W^*P^{-1}(\beta - \beta).
\]

Squared error loss in the \( \theta \)-space is equivalent to mean square error of prediction loss in the original \( \beta \) parameter space, i.e., \( L(\theta, \theta, I_j) = L(\beta, \beta, X'X). \) For squared error loss in the \( \beta \)-space, let \( W^* = P'P. \) The constant \( a_{\max} \) in the transformed model is obtained by substituting \( R = [I, 0], S = I, \) and \( W = W^* \) into (3) and the Stein-rule or positive-part Stein-rule estimator is minimax if the number of restrictions \( (J) \) is strictly greater than two.
The mean, covariance, and risk of the RLS Stein rule and its positive part may be obtained similarly using existing results for (6) and (7) in conjunction with the canonical form. The usefulness of these results is limited in practice because they depend on the unknown parameters, \( \beta \) and \( \sigma^2 \). Replacing the unknown parameters in the moment expressions with consistent estimators like the unrestricted MLEs or Stein rules yields statistics and interval estimators with unknown sampling distributions. Though biased, the Stein rule (2) has lower risk than the unrestricted MLE, and the positive-part rule (4) has lower risk than the usual Stein rule; these are sufficient reasons to use the positive-part rule for point estimation in applications. Work is under way on procedures to evaluate the small-sample precision of (2) and (4) and on interval estimation techniques based on the Stein rules. See, for example, Adkins; Adkins and Hill; and Ullah, Carter, and Srivastava.

**Discussion**

The positive-part rule takes a convex combination of the unrestricted and restricted MLEs; geometrically, this means that (4) lies between \( b \) and \( b^* \). The positive-part rule is intuitively appealing because it draws, or "shrinks," the unrestricted estimates toward, but not past, the restricted estimates. The usual Stein estimator (2) is less attractive because it changes the sign of the unrestricted MLEs whenever the value of the test statistic, \( u \), is smaller than the constant, \( c \).

The positive-part Stein rule can be used whenever one has uncertain nonsample information involving three or more linear parameter restrictions. These situations are common in agricultural economics and a wide range of potential applications can be found. Consider the following. First, one may have panel data consisting of multiple observations on several cross-sectional units and suspect that the dependent variable responds similarly to changes in the exogenous variables, \( X \), across units. The (uncertain) equality of coefficients across units can be represented by a set of linear parameter restrictions. Second, Almon polynomial distributed lag models require one to choose lag length and polynomial degree which impose linear restrictions on coefficients of the distributed lag. These exact restrictions on the parameters are only approximately true at best. Third, principal components regression is sometimes recommended for models estimated using poorly conditioned data (i.e., in the context of near multicollinearity). The deletion of principal components from the model imposes exact, sample specific, linear constraints on the model's parameters which are often difficult to interpret and evaluate. Once again, these restrictions are only approximately true and the positive-part RLS Stein rule can be used to obtain a risk improvement.

Additional examples can be found in production theory in which under certain circumstances production functions are homothetic in the inputs. Depending on the functional form used, homotheticity often implies restrictions of the form \( R\beta = r \). Similarly, in demand equations Engle aggregation or homotheticity restrictions can often be expressed in the \( R\beta = r \) form and used with the linearized equation.

Another possibility to consider when faced with uncertain linear parameter restrictions is to assume that the information they embody is correct and to impose the restrictions on the model's parameters using the RLS estimator. Unfortunately, the risk of the RLS estimator is unbounded and increases with the value of the noncentrality parameter, \( \lambda \), which measures the specification error implicit in the restrictions. A common econometric practice is to treat the restrictions as a hypothesis and employ them if they cannot be rejected by conventional tests. The resulting pretest estimator, which selects either the MLE or RLS estimators based on the outcome of the hypothesis test of the restrictions \( R\beta = r \), has greater risk than the MLE (or the Stein rules) over large regions of the parameter space. The positive-part Stein rule, which combines the MLE and RMLE, has uniformly lower risk than the MLE under arbitrary quadratic loss and may be used in any of these situations.

In the presence of uncertain nonsample information an appealing alternative to pretesting is to adopt a Bayesian procedure which uses a prior distribution on the parameters (or the sampling "equivalent," mixed estimation). The Stein rules have been shown to be empirical Bayes rules (Judge, Hill, and Bock); consequently, they are formal as well as intuitive alternatives to pretesting which have desirable statistical properties.

**Concluding Comments**

In general, most researchers should use positive-part Stein rules rather than the usual ones.
because they are known to have smaller quadratic risk under general circumstances. In this article, we have shown that a general family of Stein rules mentioned in Judge and Bock, extended by Mittelhammer and Young, and generalized by Mittelhammer (1984, 1985) is in fact dominated by the positive-part rule which sets the Stein estimator equal to the RLS estimator whenever the shrinkage factor $(c/u) > 1$. In addition, some possible uses for the positive-part rule in economic and agricultural economic research have been discussed.

[Received June 1989; final revision received September 1989.]

References


