RLS Stein-rule in Gretl

Lee C. Adkins

Department of Economics
Oklahoma State University

lee.adkins@okstate.edu

20 June 2013

Motivation

- Stein-rules dominate (or nearly so) the MLE of many models under arbitrary quadratic loss.
- Despite this, they are seldom used. I offer 3 reasons:
  1. They are known to be biased. Oddly, some find this bothersome.
  2. Their complex sampling distributions complicate testing and the estimation of confidence intervals based on them.
  3. There is no Stein-Rule ‘button to push’ in software (we can argue whether this is good or bad).
- My RLS-Stein rule **hansl** package solves (2) and (3). I’d argue that (1) isn’t a problem anyway.

What is a Stein-rule?

1. The original Stein-rule (1956) was proposed as an estimator of a multivariate (K) mean.
2. Stein showed, counterintuitively, that combining the estimation of means via "shrinkage" toward the origin could make least squares (the MVUE) inadmissible!
3. Was it just a math trick? Some thought so. But the insight into the K-means problem turned into a principle that works (at least approximately) in a wide variety of circumstances.

D. V. Lindley, (1961)

"When I first heard of this suggestion several years ago I must admit that I dismissed it as the work of one of these mathematical statisticians who are so entranced by the symbols that they lose touch with reality."

Evolution

- Efron and Morris (1973) propose a positive-part variant with a Bayesian justification
- Judge and Bock (1978) discuss it in a regression context. RLS Stein-rule
- Adkins and Hill (1990) prove that the positive part RLS Stein-rule dominates MLE
- Adkins (1988-1992) explores use of bootstrap to estimate standard errors, confidence intervals and confidence ellipsoids.

GRETL package: Design Goals

My goal is to create a software package that makes using Stein-rule as easy as estimating a linearly restricted regression model. At a minimum, it has to yield point estimates and standard errors.

Statistical Model

The classical normal linear regression model (CNLRM) is:

$$y = X\beta + e \quad e \sim N(0, \sigma^2 I_T) \tag{1}$$

- y $T \times 1$ vector of observable random variables
- X nonstochastic $T \times K$ matrix of rank K
- β is $K \times 1$ unknown parameters
- e is $T \times 1$ normally distributed random errors

Quadratic Loss and Risk

The quadratic loss associated with using an estimator $\hat{\beta}$ to estimate a vector β with weight matrix W is:

$$L(\hat{\beta}, \beta, W) = (\hat{\beta} - \beta)' W (\hat{\beta} - \beta) \tag{2}$$

For squared error loss $W = I_K$ and for mean square error of prediction loss $W = X'X$. Risk is $E[L(\hat{\beta}, \beta, W)]$

MLE/OLS

The ordinary least squares (OLS) and maximum likelihood estimator of β is:

$$b = (X'X)^{-1}X'y \sim N(\beta, \sigma^2 S^{-1}) \tag{3}$$

with $S = X'X$ and the minimum variance unbiased estimator of σ^2 is

$$\hat{\sigma}^2 = (y - Xb)'(y - Xb)/(T - K) \tag{4}$$

with $(T - K)\hat{\sigma}^2/\sigma^2 \sim \chi^2_{T-K}$ and independent of b.

Restricted Least Squares Stein-rule

RLS Stein-rule, a convex combination of OLS and RLS estimators:

$$\delta(b) = (1 - c/u)b + (c/u)b* \tag{5}$$

- b is OLS
- $b*$ is RLS: $= b - S^{-1}R'(RS^{-1}R')^{-1}(Rb - r)$
- $u = (Rb - r)'(RS^{-1}R')^{-1}(Rb - r)/J\hat{\sigma}^2 \sim F_{J,T-K,\lambda}$ is the Wald statistic for testing the J linear hypotheses $R\beta = r$
- $\lambda = (R\beta - r)'(RS^{-1}R')^{-1}(R\beta - r)/2\sigma^2$ is the noncentrality parameter

minimaxity

- $c = a(T - K)/J$.
- a is a shrinkage constant chosen by the user.

The estimator is minimax if the scalar a is chosen to lie within the interval $[0, a_{max}]$, where

$$a_{max} = [2/(T - K + 2)]\{\lambda_L^{-1} tr[(RS^{-1}R')^{-1}RS^{-1}WS^{-1}R'] - 2\}, \tag{6}$$

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

Positive-Part

the usual Stein estimator is dominated by a simple modification called the positive-part rule. The positive-part rule associated with (5) is denoted

$$\delta(b)^{+} = \begin{cases} b*, & \text{if } c > u \\ \delta(b), & c \leq u. \end{cases} \tag{7}$$

The positive-part rule keeps one from over-shrinking when u is very small.

Bootstrap Standard Errors

- Semiparametric bootstrap standard errors can be based on the empirical distribution of the least squares residuals.
- LS residuals are rescaled $\hat{e}_t^* = (T/(T-K))^{1/2}\hat{e}_t,\ t = I, \ldots, T$
- A bootstrap sample of size T is drawn randomly and with replacement from $\hat{e}$, denoted e^*
- The sample covariance of the sequence of RLS Stein-rule estimates is used to estimate the precision of the RLS Stein-rule.

Bootstrap Standard Errors

- Adkins (1992) shows that resampling from $\hat{e}$ tends to overstate the standard errors of the Stein-rule, especially for small amounts of noncentrality.
- He suggests resampling randomly from the RLS Stein-rule residuals $e_\delta = y - X\delta$ and generate bootstrap samples using $y^* = X\delta + e_\delta^*$ where e_δ^* represents a random resample from the (possibly rescaled) RLS Stein-rule residuals e_δ. This is similar to Brownstone (1990).

Main Function

```
function bundle RLSStein (series y "Dependent Variable",
                         list EXOG "Regressors",
                         matrix R "R for linear hypotheses RB=r",
                         matrix r "r for linear hypotheses",
                         int Loss[0:1:1] "Loss function" {"SEL", "MSEP"} ,
                         int verb[0:1:1] "Verbosity" {"no print","print"} ,
                         int B[100] "Bootstrap Replications")
    # first thing, drop all obs with missing values anywhere
    list EVERYTHING = y || EXOG
    smpl EVERYTHING --no-missing
```

```
    bundle rr = Stein_setup(y, EXOG, R, r, Loss, verb, B)
    scalar err = aw(&rr)
    scalar err = Stein_estimate(&rr)
    scalar err = bootStein(&rr)
    if verb == 1
        Stein_printout(&rr)
    endif
    return rr
end function
```

Helper Functions

Stein_setup Initiates the bundle. Computes OLS, RLS, and the test statistic, u based on $R\beta = r$.

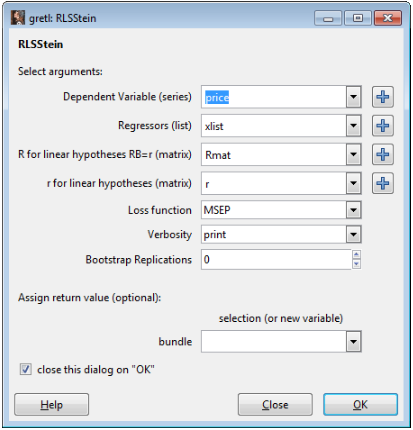
aw Wrapper for two other functions, `Wmat` and `amax`. `Wmat` produces the weight matrix for the quadratic loss function. `amax` determines the maximum shrinkage constant a_{max}.

Stein_estimate Computes the RLS Stein-rule based on W and a_{max}.

bootStein Computes the bootstrap standard errors using number of samples chosen.

Stein_printout Handles printing based on verbosity level.

GUI dialog box



Model

Home prices in San Diego, 1990 (Ramanathan)

$$\text{price} = \beta_1 + \beta_2 \text{sqft} + \beta_3 \text{sqft}^2 + \beta_4 \text{bedrms} + \beta_5 \text{baths} + e \quad (8)$$

```
price = sale price in thousands of dollars (Range 199.9 - 505)
sqft  = square feet of living area (Range 1065 - 3000)
bedrms = number of bedrooms (Range 3 - 4)
baths = number of bathrooms (Range 1.75 - 3)
```

The following restrictions are considered: $\beta_2 = 360$; $\beta_3 = -50$; $\beta_4 = 0$; and, $\beta_5 = 0$.

hansl script

```
open data4-1
series sqft = sqft/1000
square sqft bedrms
list xlist = const sqft sq_sqft bedrms baths
matrix Rmat = zeros(4,1)~I(4)
matrix r = { 350 ; -50 ; 0 ; 0}

bund = RLSStein(price, xlist, Rmat, r, 1, 1, 100)
```

OLS

```
Stein-Rule Estimation of a linear regression
using observations 1-14
Mean Square Error of Prediction Loss
Dependent Variable y

Unrestricted OLS

             coefficient   std. error      z        p-value
  --------------------------------------------------------
  const        -14.8037     138.026     -0.1073    0.9146
  sqft         367.990      163.896      2.245     0.0248   **
  sq_sqft      -51.1936      38.6554    -1.324     0.1854
  bedrms       -43.7401      30.9703    -1.412     0.1579
  baths         -3.71536     42.1948    -0.08805   0.9298
```

RLS

```
Restricted LS

              coefficient   std. error      z        p-value
  -------------------------------------------------------------
  const       -153.252      10.2323     -14.98     1.03e-050 ***
  sqft          350.000      0.00000      NA        NA
  sq_sqft       -50.0000     0.00000      NA        NA
  bedrms          0.00000    0.00000      NA        NA
  baths           0.00000    0.00000      NA        NA
```

Stein-Rule

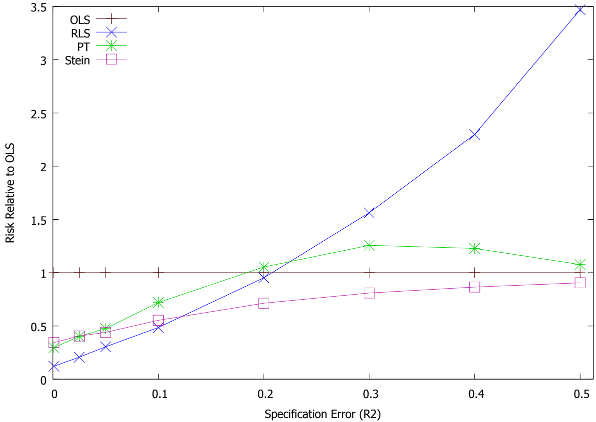
```
Stein-Rule estimates
             Stein        SE
  const   -84.0725  102.2714
   sqft   358.9891  120.2854
sq_sqft   -50.5964   28.5793
 bedrms   -21.8559   20.6569
  baths    -1.8565   30.4958

a=0.3636, c=0.4091, shrinkage is 0.5003 and F=0.8177
```

Monte Carlo Design

- Orthonormal linear regression model
- MSEP Loss
- $\beta = \ell j_K$ where $\ell = R^2 T\sigma^2/((1-R^2)K)^{1/2}$
- R^2= .001, .025, .05, .1, .2, .3, .4, .5
- NMC = 500, T=30, K=8, $\sigma^2 = 1$
- Pretest at 10%
- Restrictions: $\beta_2 = \beta_3 = \ldots \beta_8 = 0$

Stein-Rule Risk



Some Issues

- Additional error checking. The routine checks for $K > 3$ and failure of the minimaxity condition. However when J=K, I can't use gretl's `restrict` successfully.
- Add confidence intervals to output. But, which version to use? What is a suitable minimum number of bootstrap samples to estimate tail behavior?
- A check box for Lindley's version of the James-Stein rule. Shrinks only slopes toward zero, thus $b* = \bar{y}$.
- The pdf Help file not opening from the dialog help.
- Where to put this in the GUI menu?