S-values: Conventional measures of the sturdiness of the signs regression coefficients

Ed Leamer
December 2015
BITSS: Berkeley

Related Paper
S-values and Bayesian Weighted All-Subsets Regressions
Introductory Comments

- How I started on this journey and where I am now.
- Specification Searches
  - Hypothesis Testing Searches
  - Interpretative Searches
  - Proxy Variable Searches
  - Simplification Searches
  - Data Selection Searches
  - Data-instigated Models
- Transparency alone is not enough. We need accountability.
  - If you tried only one regression and reported it, but I found the very same regression after 1000 trials and reported it and also reported the 999 others, is there a difference in the reliability of my final equation and your one-and-only-one equation?? It’s the same regression after all.
Two Sources of Model Error

- t-values measure the impact of *statistical uncertainty* on the signs of coefficients.
- s-values measure the impact of *model ambiguity* on the signs of coefficients. (s stands for “sturdy.”)
**Theorem:** When a variable is omitted from a linear regression, an estimated coefficient cannot change its sign if the coefficient of the retained variable has a t-value that exceeds in absolute value the t-value of the coefficient of the omitted variable.


**Operational Significance:**
If you want to change the sign of a coefficient, omit variables with bigger t-values!
Big t good, Small t bad.
Is That All That Can be Said??

- **Statistical Uncertainty:** Compare the absolute t-value with the number 2.
- **Model Ambiguity:** Compare the absolute t-value with the absolute t-values of other coefficients.
Two Variable Case: Is the Sign of a Regression Coefficient Known?

Statistical Uncertainty: How much does knowledge about the sign depend on the limited sample size and the degree of collinearity.

Model Ambiguity: How much does knowledge about the sign depend on the choice of explanatory variables?
Notation

- Standardize the data to have zero means and unit variances.
- The data consists of a sample size and three correlations
  \[ r_1 = \text{correlation}(y,x_1), \ r_2 = \text{correlation}(y,x_2) \]
  \[ \rho = \text{correlation}(x_1,x_2) \]
- OLS estimates

\[
\begin{bmatrix}
\beta_1 \\
\beta_2
\end{bmatrix} = \left( \begin{bmatrix} 1 & \rho \ \\ \rho & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix}
r_{1y} \\
r_{2y}
\end{bmatrix} = \left( \begin{bmatrix} 1 & \rho \\
1 - \rho^2 & -\rho 
\end{bmatrix} \right) / (1 - \rho^2)
\]
2-Variable Sturdiness Theorem

- **Theorem**
  - A necessary and sufficient condition for \( b_1 \) and \( r_1 \) to have the same sign is \( r_{↓2} / r_{↓1} \rho < 1 \).
  - A necessary and sufficient condition for \( b_2 \) and \( r_2 \) to have the same sign is \( r_{↓1} / r_{↓2} \rho < 1 \).
  - A necessary and sufficient condition for the signs of both coefficients to be unchanged when a variable is omitted is \( \max(r_{↓1} / r_{↓2}, r_{↓2} / r_{↓1}) \rho < 1 \).
  - Note \( r_{↓2} / r_{↓1} = t_{↓2} / t_{↓1} \).

- **Implications**
  - A value of \( \rho \) equal to zero makes sturdy estimates.
  - The sign of \( \rho \) opposite \( r_{↓2} / r_{↓1} \) makes sturdy estimates.
  - Equal values of \( r_1 \) and \( r_2 \) make a sturdy estimates.
  - t-values rank the sturdiness of coefficients.
No sign change: The sign of $\rho$ opposite $r_{12} / r_{11}$.
Likelihood ellipse points away from the origin.
Variables are substitutes: sum of the coefficients is well estimated.
Shrink one but enlarge the other.
No sign change: The sign of $\rho$ same as $r_{12}/r_{11}$ but $r_1=r_2$

Likelihood ellipse points toward the origin: Shrink together.
Variables are complements: difference of the coefficients is well estimated.
Sign Change: The sign of $\rho$ same as $r_{12}/r_{11}$ and $|r_1|>|r_2|$

Likelihood ellipse points to the side of the origin favoring one coefficient over the other.
Shrink one coefficient more than the other
An s-value: measures model ambiguity the way a t-value measures statistical uncertainty

- **t-value** = estimate / half the length of the confidence interval

- Center and spread of ambiguity interval
  - Center is average of \((r_{11} - \rho r_{22})/(1 - \rho^2)\) and \(r_{11} : [r_{11} - \rho r_{22} + (1 - \rho^2) r_{11}] / 2(1 - \rho^2)\)
  - Distance between the two estimates is absolute value of \((r_{11} - \rho r_{22})/(1 - \rho^2) - r_{11} = [r_{11} - \rho r_{22} - (1 - \rho^2) r_{11}] / (1 - \rho^2)\)

- **s-value** is the average divided by half the distance:
  \[ s_1 = [(2 - \rho^2)r_{11} - \rho r_{22}] / \text{abs}[\rho^2 r_{11} - \rho r_{22}] \]
Conjecture: The alternative signs of the regression coefficients in the $2^k$ regressions formed by omitting subsets of the $k$ explanatory variables can be found by computing the unconstrained regression and the $k$ regressions with one-at-a-time omissions of the $k$ explanatory variables.
Multivariate Regression With “Soft” Constraints
Regression Setting

- The nx1 vector $y$ is normally distributed with mean $X\beta$ and covariance matrix $\sigma^2I$, where $X$ is an nxk matrix of explanatory variables, $\beta$ is a kx1 vector of unknown coefficients and $\sigma^2$ is a scalar variance applicable to all n observations.
- The OLS estimate of $\beta$ is $b = (X'X)^{-1} X'y$.
- The likelihood function is constant on the ellipsoids $c = (\beta - b)'X'X(\beta - b)$.
Pooling Two Data Sets: Matrix Weighted Averages

- If there are two sets of observations \((y_1, X_1, y_2, X_2)\) with different residual variances, \(\sigma_1^2\) and \(\sigma_2^2\), the data can be transformed into homoscedastic form by dividing by the standard deviations, \(\sigma_1\) and \(\sigma_2\) to produce the pooled estimate:

\[
b = (X'X)^{-1} X'y = \left[\frac{X_1'X_1}{\sigma_1^2} + \frac{X_2'X_2}{\sigma_2^2}\right]^{-1} \left[\frac{X_1'y_1}{\sigma_1^2} + \frac{X_2'y_2}{\sigma_2^2}\right]
\]

\[
= \left[H_1 + H_2\right]^{-1} \left[H_1 b_1 + H_2 b_2\right]
\]

- \(H_i = \frac{X_i'X_i}{\sigma_i^2}, H_i b_i = \frac{X_i'y_i}{\sigma_i^2}\)

Figure 1: OLS and Three Estimates Closer to Zero
Feasible Ellipse
Intervals of Prior Covariance Matrices Map Into Ellipsoids of Estimates

Figure 2  Four Ellipses of Estimates

$V \leq V^*$

$V^* \leq V$

$0 \leq V$
The Orthant of Simple Correlations Is Special

Figure 3  The Feasible Ellipse and the Impossible Orthant

Feasible Directions in Quadrant of $b$

Quadrant of Simple Correlations

$\hat{\beta}'X'y = 0$
The Orthant of Simple Correlations Is Special When the Prior is Strong and the Prior covariance matrix is diagonal

Theorem: If the prior precision matrix is diagonal, then estimates may not lie in opposite orthants.
ALERT:
s-values can be proportional to t-values
Standardized Variables Are Amenable to Context-Independent Conventional Prior Covariance Matrices

\[ y_t = x_t' \beta + \varepsilon_t \]
\[ \sigma_y^2 = \beta' \Sigma_{xx} \beta + \sigma_z^2 \]
\[ R^2 = \beta' \Sigma_{xx} \beta \]
\[ \text{Var}(\beta) = \nu^2 I, \]
\[ \text{trace}(\beta' \Sigma_{xx} \beta) = \text{trace}(\Sigma_{xx} \beta \beta') \]
\[ E(R^2) = \text{trace}(\nu^2 \Sigma_{xx}) = \nu^2 k \]
\[ \nu_L^2 = \frac{\min E(R^2)}{k} \leq \nu^2 \leq \frac{\max E(R^2)}{k} = \nu_U^2 \]
Regression Explaining Growth Rate of 87 Countries: 1960 to 1996
The Proposed Reporting Style
Apparent Sign Conflicts In Bold; Significant Results Shaded

Models explaining growth rates of real per capita income from 1960 to 1996: 87 countries
Data Source: Sala-i-Martin, Doppelhofer, and Miller (2004)
Standardized variables (unit variance and zero mean)
All variables treated the same
Sorted first by category and then by p-value

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>&quot;0&quot;</th>
<th>0.1</th>
<th>0.5</th>
<th>1</th>
<th>Infinite</th>
<th>&quot;0&quot;</th>
<th>0.1</th>
<th>0.5</th>
<th>1</th>
<th>Infinite</th>
<th>(0.1, 0.3)</th>
<th>(0.5, 1.0)</th>
</tr>
</thead>
<tbody>
<tr>
<td>catchup</td>
<td>GDP in 1960 (log)</td>
<td>0.318</td>
<td>-0.023</td>
<td>-0.195</td>
<td>-0.283</td>
<td>-0.462</td>
<td>3.09</td>
<td>-0.32</td>
<td>-1.60</td>
<td>-1.99</td>
<td>-2.61</td>
<td>-0.48</td>
<td>-0.47</td>
</tr>
<tr>
<td>culture</td>
<td>Fraction Speaking Foreign Language</td>
<td>0.258</td>
<td>0.092</td>
<td>0.127</td>
<td>0.135</td>
<td>0.145</td>
<td>2.46</td>
<td>1.52</td>
<td>1.54</td>
<td>1.55</td>
<td>1.59</td>
<td>0.67</td>
<td>0.72</td>
</tr>
<tr>
<td>geography</td>
<td>Absolute Latitude</td>
<td>0.394</td>
<td>0.105</td>
<td>0.209</td>
<td>0.249</td>
<td>0.314</td>
<td>3.95</td>
<td>1.58</td>
<td>1.90</td>
<td>2.01</td>
<td>2.18</td>
<td>0.67</td>
<td>0.69</td>
</tr>
<tr>
<td>geography</td>
<td>Fraction of Land Area Near Navigable Water</td>
<td>0.404</td>
<td>0.113</td>
<td>0.145</td>
<td>0.149</td>
<td>0.155</td>
<td>4.07</td>
<td>1.85</td>
<td>1.70</td>
<td>1.64</td>
<td>1.57</td>
<td>0.76</td>
<td>0.32</td>
</tr>
<tr>
<td>Government</td>
<td>British Colony Dummy</td>
<td>0.076</td>
<td>0.066</td>
<td>0.104</td>
<td>0.109</td>
<td>0.107</td>
<td>0.70</td>
<td>1.11</td>
<td>1.30</td>
<td>1.27</td>
<td>1.17</td>
<td>-0.42</td>
<td>-0.45</td>
</tr>
<tr>
<td>Government</td>
<td>Government: GDP Share 1960s</td>
<td>-0.073</td>
<td>-0.048</td>
<td>-0.073</td>
<td>-0.079</td>
<td>-0.088</td>
<td>-0.67</td>
<td>-0.81</td>
<td>-0.94</td>
<td>-0.92</td>
<td>-0.90</td>
<td>-0.24</td>
<td>-0.15</td>
</tr>
<tr>
<td>Government</td>
<td>Government: Fraction Spent in War 1960-90</td>
<td>-0.135</td>
<td>-0.029</td>
<td>-0.038</td>
<td>-0.041</td>
<td>-0.050</td>
<td>-1.25</td>
<td>-0.50</td>
<td>-0.48</td>
<td>-0.51</td>
<td>-0.57</td>
<td>0.58</td>
<td>0.59</td>
</tr>
<tr>
<td>Government</td>
<td>Curried Orientation</td>
<td>0.030</td>
<td>0.009</td>
<td>-0.008</td>
<td>-0.017</td>
<td>-0.022</td>
<td>0.25</td>
<td>0.10</td>
<td>-0.10</td>
<td>-0.20</td>
<td>-0.30</td>
<td>-0.03</td>
<td>-0.32</td>
</tr>
<tr>
<td>resources</td>
<td>Primary Schooling in 1960</td>
<td>0.574</td>
<td>0.201</td>
<td>0.365</td>
<td>0.424</td>
<td>0.525</td>
<td>6.47</td>
<td>3.03</td>
<td>3.61</td>
<td>3.81</td>
<td>4.14</td>
<td>1.34</td>
<td>1.40</td>
</tr>
<tr>
<td>resources</td>
<td>Primary Exports 1970</td>
<td>-0.491</td>
<td>-0.157</td>
<td>-0.271</td>
<td>-0.302</td>
<td>-0.342</td>
<td>-5.19</td>
<td>-2.54</td>
<td>-2.85</td>
<td>-2.69</td>
<td>-2.61</td>
<td>0.17</td>
<td>0.17</td>
</tr>
<tr>
<td>resources</td>
<td>Oil Producing Country Dummy</td>
<td>-0.019</td>
<td>0.017</td>
<td>0.055</td>
<td>0.072</td>
<td>0.102</td>
<td>-0.18</td>
<td>0.28</td>
<td>0.70</td>
<td>0.86</td>
<td>1.14</td>
<td>0.15</td>
<td>0.18</td>
</tr>
<tr>
<td>resources</td>
<td>Higher Education 1960</td>
<td>0.308</td>
<td>0.034</td>
<td>0.035</td>
<td>0.048</td>
<td>0.082</td>
<td>2.98</td>
<td>0.51</td>
<td>0.34</td>
<td>0.40</td>
<td>0.61</td>
<td>-0.17</td>
<td>-0.17</td>
</tr>
<tr>
<td>resources</td>
<td>Fraction Population Less than 15</td>
<td>-0.228</td>
<td>0.024</td>
<td>0.082</td>
<td>0.099</td>
<td>0.122</td>
<td>-2.16</td>
<td>0.34</td>
<td>0.63</td>
<td>0.62</td>
<td>0.53</td>
<td>-0.96</td>
<td>-1.00</td>
</tr>
<tr>
<td>resources</td>
<td>Fraction Population Over 65</td>
<td>0.234</td>
<td>-0.025</td>
<td>-0.085</td>
<td>-0.098</td>
<td>-0.107</td>
<td>2.21</td>
<td>-0.32</td>
<td>-0.65</td>
<td>-0.62</td>
<td>-0.44</td>
<td>0.29</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Mean Squared Coefficient: 0.010 0.028 0.035 0.058
Can you see the conflict between the signs of the simple regressions and the signs of the multiple regression?
t-values and s-values are related