Large-Sample Robust and Non-linear Inference

Walter Sosa-Escudero

May 14, 2017

Motivation

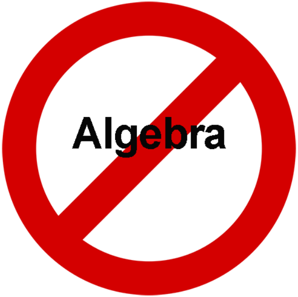
```
. reg ltc lq lpl lpf lpk
------------------------------------------------------------------------------
         ltc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          lq |   .7209135   .0174337    41.35   0.000     .6864462    .7553808
         lpl |   .4559645    .299802     1.52   0.131    -.1367602    1.048689
         lpf |   .4258137   .1003218     4.24   0.000     .2274721    .6241554
         lpk |  -.2151476   .3398295    -0.63   0.528    -.8870089    .4567136
       _cons |  -3.566513   1.779383    -2.00   0.047    -7.084448   -.0485779
------------------------------------------------------------------------------
```

Standard practice under heteroskedasticity

```
. reg ltc lq lpl lpf lpk, robust
------------------------------------------------------------------------------
             |               Robust
         ltc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          lq |   .7209135   .0325376    22.16   0.000      .656585     .785242
         lpl |   .4559645    .260326     1.75   0.082    -.0587139    .9706429
         lpf |   .4258137   .0740741     5.75   0.000     .2793653    .5722622
         lpk |  -.2151476   .3233711    -0.67   0.507    -.8544698    .4241745
       _cons |  -3.566513   1.718304    -2.08   0.040    -6.963693   -.1693331
------------------------------------------------------------------------------
```

Where does 'robust' come from? Does it work?

Ladies' and gentlemen's agreement:



Stupid algebraic steps will be left for homework. Sign here: ....

Cauchy-Schwartz inequality

$$E(X,Y)^2 \leq E(X^2)\ E(Y^2)$$

Recall (asympototic normality of linear model):

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) \xrightarrow{p} N(0, \Sigma_x^{-1} S \Sigma_x^{-1})$$

where $\Sigma_x = E(x_i x_i')$ and $S = V(x_i u_i)$.

Notation: $AV(\hat{\beta}_n) \equiv \Sigma_x^{-1} S\, \Sigma_x^{-1}$. We need a consistent estimate of $AV(\hat{\beta}_n)$

- Note that $n^{-1} X_i' X_i$ is a consistent estimator for Σ_x.
- Recall we are allowing for *heteroskedasticity*. Alternative consistent estimators for S depend on what we are willing to assume on this.

Variance estimation under homoskedasticity

Under homoskedasticity $E(u_i^2|x_i) = \sigma_0^2$. Using using LIE:

$$S = E(u_i^2 x_i x_i') = \sigma_0^2\ E(x_i x_i') = \sigma_0^2 \Sigma_x$$

So

$$AV(\hat{\beta}_n) = \Sigma_x^{-1} S \Sigma_x^{-1} = \sigma_0^2\ \Sigma_x^{-1} \Sigma_x \Sigma_x^{-1} = \sigma_0^2\ \Sigma_x^{-1}$$

Hence a consistent estimator for $AV(\hat{\beta}_n)$ can be

$$\widehat{\mathsf{AV}}_h = n\ s^2 (X'X)^{-1}$$

n times the classical estimator.

Heteroskedsaticity robust variance estimation

Can we estimate $AV(\hat{\beta}_n) = \Sigma_x^{-1} S \, \Sigma_x^{-1}$ without assuming homoskedasticity?

Recall $S = E(x_i u_i u_i x_i') = E(u_i^2 x_i x_i')$. We will need an additional assumption

Assumption (fourth moments): $E\left[(x_{ik} x_{ij})^2\right]$ exists and is finite for all $k, j = 1, 2, \ldots, K$.

Result

$$\hat{S}_w \equiv \frac{1}{n} \sum_{i=1}^{n} e_i^2 x_i x_i' \;\; \xrightarrow{p} \;\; S$$

where e_i's are OLS residuals.

Proof:

$$e_i = y_i - x_i'\hat{\beta} = y_i - x_i'\beta - x_i'(\hat{\beta}-\beta) = u_i - x_i'(\hat{\beta}-\beta)$$
$$e_i^2 = u_i^2 - 2u_i x_i'(\hat{\beta}-\beta) + (\hat{\beta}-\beta)' x_i x_i'(\hat{\beta}-\beta)$$

Replacing

$$\frac{1}{n}\sum_{i=1}^{n} e_i^2 x_i x_i' = \frac{1}{n}\sum_{i=1}^{n} u_i^2 x_i x_i' - \frac{2}{n}\sum_{i=1}^{n} u_i x_i'(\hat{\beta}-\beta) x_i x_i' + + \frac{1}{n}\sum_{i=1}^{n} (\hat{\beta}-\beta)' x_i x_i' (\hat{\beta}-\beta) x_i x_i'$$

First note that

$$\frac{1}{n}\sum_{i=1}^{n} u_i^2 x_i x_i' \xrightarrow{p} E(u_i^2 x_i x_i')$$

by Kolmogorov's LLN, since we assumed finite second moments (expectaction exists) and iid.

We will show the other two terms converge to zero

I) $\mathbf{A} = \frac{2}{n}\sum_{i=1}^{n} u_i \; x_i'(\hat{\beta}-\beta)\; x_i x_i'$

$$\begin{aligned} \mathbf{A} &= \frac{2}{n}\sum_{i=1}^{n} u_i \left[\sum_{k=1}^{K} x_{ik}(\hat{\beta}_k-\beta_k)\right] x_i x_i' \\ &= 2\sum_{k=1}^{K}(\hat{\beta}_k-\beta_k)\left[\frac{\sum_{i=1}^{n} u_i x_{ik}\; x_i x_i'}{n}\right] \end{aligned}$$

Note $\hat{\beta}_k-\beta_k \xrightarrow{p} 0$, by consistency. So if we can show $\left[\;\right] \xrightarrow{p} < \infty$, we are done.

$\left[\frac{1}{n}\sum_{i=1}^{n} u_i x_{ik}\ x_i x_i'\right]$ is a $K \times K$ matrix with typical (h, j) element:

$$\frac{\sum_{i=1}^{n} u_i x_{ik} x_{ih} x_{ij}}{n}$$

By the Cauchy-Schwartz inequality:

$$E|x_{ik}x_{ih}x_{ij}u_i| \leq E\left[|x_{ik}x_{ih}|^2\right]^{1/2} E\left[|x_{ij}u_i|^2\right]^{1/2}$$

Both factors in the RHS are $< \infty$, by our fourth moments assumption 5 and by assumption 3. Hence, we can use the LLN:

$$\frac{1}{n}\sum_{i=1}^{n} u_i x_{ik}\ x_i x_i' \xrightarrow{p} E(u_i x_{ik} x_i x_i') < \infty,$$

so by the product rule and continuity, $\mathbf{A} \xrightarrow{p} 0$.

II) $\mathbf{B} = \frac{1}{n}\sum_{i=1}^{n}(\hat{\beta}-\beta)'x_i\, x_i'(\hat{\beta}-\beta)\, x_i x_i'$

Using the same trick as before:

$$\mathbf{B} = \frac{1}{n}\sum_{i=1}^{n}\left[\sum_{k=1}^{K} x_{ik}(\hat{\beta}_k-\beta_k)\right]\left[\sum_{k'=1}^{K} x_{ik'}(\hat{\beta}_{k'}-\beta_{k'})\right] x_i x_i'$$

Now we have a sum of K^2 matrices. The (h,j) element of the k,k' summand will be

$$(\hat{\beta}_k-\beta_k)(\hat{\beta}_{k'}-\beta_{k'})\;\frac{1}{n}\sum_{i=1}^{n} x_{ik}x_{ik'}x_{ih}x_{ij}$$

Using again the Cauchy Schwartz inequality and the finite fourth moments assumption: $E|x_{ik}x_{ik'}x_{ih}x_{ij}| < \infty$

And again, by consistency and LLN, $\mathbf{B} \xrightarrow{p} 0$. *q.e.d.*

Then, using $\hat{S}_w$ as an estimator for S and noting

$$\hat{S}_w = \frac{1}{n}\sum_{i=1}^{n} e_i^2 x_i x_i^2 = \frac{1}{n}(X'BX)$$

with $B \equiv \text{diag}(e_1^2, \ldots, e_n^2)$,

$$\begin{aligned} \widehat{\text{AV}}_w(\hat{\beta}_n) &= \hat{\Sigma}_x^{-1}\hat{S}_w\hat{\Sigma}_x^{-1} \\ &= n(X'X)^{-1}n^{-1}(X'BX)n(X'X)^{-1} \\ &= n(X'X)^{-1}(X'BX)(X'X)^{-1} \end{aligned}$$

This is White's heteroskedasticity consistent estimator for the asymptotic variance of $\hat{\beta}_n$. Remember that in the derivation of all result we never ruled out the possibility of conditional heteroskedasticity, then its consistency *does not* depend on it.

Returns-to-scale:

```
. reg ltc lq lpl lpf lpk
------------------------------------------------------------------------------
         ltc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          lq |   .7209135   .0174337    41.35   0.000     .6864462    .7553808
         lpl |   .4559645    .299802     1.52   0.131    -.1367602    1.048689
         lpf |   .4258137   .1003218     4.24   0.000     .2274721    .6241554
         lpk |  -.2151476   .3398295    -0.63   0.528    -.8870089    .4567136
       _cons |  -3.566513   1.779383    -2.00   0.047    -7.084448   -.0485779
------------------------------------------------------------------------------

. reg ltc lq lpl lpf lpk, robust
------------------------------------------------------------------------------
             |               Robust
         ltc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          lq |   .7209135   .0325376    22.16   0.000      .656585     .785242
         lpl |   .4559645    .260326     1.75   0.082    -.0587139    .9706429
         lpf |   .4258137   .0740741     5.75   0.000     .2793653    .5722622
         lpk |  -.2151476   .3233711    -0.67   0.507    -.8544698    .4241745
       _cons |  -3.566513   1.718304    -2.08   0.040    -6.963693   -.1693331
------------------------------------------------------------------------------
```

The Delta Method

Suppose we want to perform inference about a non-linear function of β, say $a(\beta)$.

Example 1

$$y_i = \beta_0 + \beta_1 \text{south} \; + z_i' \beta_2 + u_i$$

y_i is log-wages, *south* is a dummy indicating if the person lives in the southern region z_i is a vector of control variables.
The percent difference between south/not south is given by

$$\gamma \equiv e^{\beta_1} - 1$$

and for small values of β_1 is very similar to β_1. Suppose we are interested γ *exactly*. A consistent estimator is $e^{\hat{\beta}_1} - 1$. A natural problem is how to construct a confidence interval for γ.

Example 2: consider now

$$y_i = \beta_1 \text{ exper } + \beta_2 \text{ exper}^2 \, + z_i' \beta + u_i$$

where *exper* is work experience in years. The level of experience that maximizes expected wages is:

$$\gamma \equiv -\frac{\beta_1}{2\beta_2}$$

and a consistent estimate is provided by

$$\hat{\gamma} = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}$$

How can we construct an estimate for the standard deviation of a confidence interval for γ?

Result (Delta Method): suppose x_n is a sequence of random vector of dimension K such that

$$x_n \xrightarrow{p} \beta \text{ and } \sqrt{n}(x_n - \beta) \xrightarrow{d} Z$$

and $a(x) : \Re^K \rightarrow \Re^r$ is a function with continuous derivatives

$$A(\beta) \equiv \frac{\partial a(\beta)}{\partial \beta'}$$

(note $A(\beta)$ is an $r \times K$ matrix).
Then:

$$\sqrt{n}\left[a(x_n) - a(\beta)\right] \xrightarrow{d} A(\beta) Z$$

Proof: Take a first-order mean value expansion of $a(x_n)$ around β:

$$a(x_n) = a(\beta) + A(y_n)\,(x_n - \beta)$$

where the 'mean value' y_n is a vector between x_n and β.
From this, get

$$\sqrt{n}\big[a(x_n) - a(\beta)\big] = A(y_n)\,(x_n - \beta)$$

Now $y_n \xrightarrow{p} \beta$ (why?) so $A(y_n) \xrightarrow{p} A(\beta)$ by continuous differentiability.

Then, by the hypothesis of the theorem and Slutzky's Theorem

$$\sqrt{n}\,\left[a(x_n) - a(\beta)\right] \xrightarrow{d} A(\beta) Z$$

As a simple corollary note that if

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{d} N(0, \mathsf{AV}(\hat{\beta}_n))$$

then

$$\sqrt{n}\ \left[a(\hat{\beta}_n) - a(\beta)\right] \xrightarrow{d} N\Big(0, A(\beta)\mathsf{AV}(\hat{\beta}_n)A(\beta)'\Big)$$

Example 1 (Blackburn and Neumark, 1992, also in Wooldridge, 2002)

$$y_i = \beta_0 + \beta_1 \text{ south } + z_i' \beta_2 + u_i$$

$$a(\beta_1) = e^{\beta_1} - 1$$

with

$$A(\beta_1) = e^{\beta_1}$$

So, according to the delta-method

$$\widehat{\text{AV}} \left(e^{\hat{\beta}_1} - 1 \right) = \left[e^{\hat{\beta}_1} \right]^2 \text{AV}(\hat{\beta}_1)$$

```
. reg lwage exper tenure married black south urban educ

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       exper |    .014043   .0031852     4.41   0.000      .007792     .020294
      tenure |   .0117473    .002453     4.79   0.000     .0069333    .0165613
     married |   .1994171   .0390502     5.11   0.000     .1227801     .276054
       black |  -.1883499   .0376666    -5.00   0.000    -.2622717   -.1144281
       south |  -.0909036   .0262485    -3.46   0.001     -.142417   -.0393903
       urban |   .1839121   .0269583     6.82   0.000     .1310056    .2368185
        educ |   .0654307   .0062504    10.47   0.000     .0531642    .0776973
       _cons |   5.395497    .113225    47.65   0.000      5.17329    5.617704
------------------------------------------------------------------------------

. nlcom exp(_b[south])-1

       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |  -.0868943   .0239677    -3.63   0.000    -.1339315   -.0398571
```

Example 2:

$$y_i = \beta_1 \text{ exper } + \beta_2 \text{ exper}^2 + z_i'\beta + u_i, \quad a(\beta_1, \beta_2) = -\beta_1/(2\beta_2)$$

```
. regress lwage edup edusi edus eduui eduu exper exper2 if muest==1
------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        edup |   .2104513   .0629835     3.34   0.001     .0869123    .3339903
       edusi |   .4148728   .0678469     6.11   0.000     .2817946     .547951
        edus |   .7587112   .0695764    10.90   0.000     .6222406    .8951817
       eduui |   1.018209    .077569    13.13   0.000      .866061    1.170356
        eduu |   1.560496   .0769774    20.27   0.000     1.409509    1.711483
       exper |   .0283668   .0071065     3.99   0.000     .0144279    .0423058
      exper2 |  -.0002502   .0001509    -1.66   0.098    -.0005462    .0000458
       _cons |   .2130178   .0934142     2.28   0.023     .0297906    .3962449
------------------------------------------------------------------------------

. nlcom -_b[exper]/(2*_b[exper2])

       _nl_1:  -_b[exper]/(2*_b[exper2])

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _nl_1 |    56.6962    20.7191     2.74   0.006     16.05673    97.33567
------------------------------------------------------------------------------
```