ChrisAuld.com
Economics, econometrics, etc.

| About | Blog archive | Academic papers | Comments policy |

Search…

Recent Posts

Steve Keen still butchering basic microeconomics

More bizarre anti-economics from the Globe and Mail

The intuition of robust standard errors

Debunking Debunking Economics

Behavioral hazard in health insurance

What do we know about the effect of income inequality on health?

The World's Colonisation and Trade Routes Formation as Imitated by Slime Mould

Externality [ek-ster-nal-i-tee] [noun]: "We don't give a shit."

Blogroll

Bad Science

Becker-Posner Blog

Causal analysis in theory and practice

Consider the evidence

Econ Academics

EconJeff

Econlog

Economic Logic

Economix

Economy Lab

Five Thirty Eight

Freakonomics

Free Exchange

Marginal Revolution

Miles Corak

Modeled Behavior

Offsetting Behavior

Oikos

Overcoming Bias

Popehat

Rescuing the Frog

Social Science Statistics Blog

Statistical modeling, causal inference, and social science (Gelman)

The Agitator

The Incidental Economist

The Reality-Based Community

Worthwhile Canadian Initiative

Follow Blog via Email

Enter your email address to follow this blog and receive notifications of new posts by email.

> [   ]

Sign me up!

Twitter

« Debunking Debunking Economics | More bizarre anti-economics from the Globe and Mail »

The intuition of robust standard errors

Commonly econometricians conduct inference based on covariance matrix estimates which are consistent in the presence of arbitrary forms of heteroskedasticity; the associated standard errors are referred to as "robust" (also, confusingly, White, or Huber-White, or Eicker-Huber-White) standard errors. These are easily requested in Stata with the "robust" option, as in the ubiquitous

reg y x, robust.

Everyone knows that the usual OLS standard errors are generally "wrong," that robust standard errors are "usually" bigger than OLS standard errors, and it often "doesn't matter much" whether one uses robust standard errors. It is whispered that there may be mysterious circumstances in which robust standard errors are smaller than OLS standard errors. Textbook discussions typically present the nasty matrix expressions for the robust covariance matrix estimate, but do not discuss in detail when robust standard errors matter or in what circumstances robust standard errors will be smaller than OLS standard errors. This post attempts a simple explanation of robust standard errors and circumstances in which they will tend to be much bigger or smaller than OLS standard errors.

Expressions for OLS and robust standard errors.

Consider the univariate linear model

$$(y_i - \bar{y}) = \beta(x_i - \bar{x}) + u_i,$$

where y is the dependent variable, x is a covariate, u is the error term, and β is the parameter over which we would like to make inferences. I've omitted a constant by expressing the model in deviations from sample means, denoted with overbars. Assume u is mean independent of x and serially uncorrelated, but allow heteroskedasticity, $V(u_i) = \sigma_i^2$. Let $\hat{\beta}$ denote the OLS estimate of β.

If we erroneously assume the error is homoskedastic, we estimate the variance of $\hat{\beta}$ with

$$\hat{V}^{OLS}(\hat{\beta}) = \frac{s^2}{\sum_i (x_i - \bar{x})^2} \approx \frac{\bar{\sigma}^2}{\sum_i (x_i - \bar{x})^2},$$

where $s^2 = (n-2)^{-1}(SSR)$. I will refer to the square root of this estimate throughout as the "OLS

standard error." When the errors are heteroskedastic, s^2 converges to the mean of σ_i^2, denote that $\bar{\sigma}^2$. However, the true sampling variance of $\hat{\beta}$ can easily be shown to be

$$V(\hat{\beta}) = \left(\frac{1}{\sum_i (x_i - \bar{x})^2}\right)^2 \sum_i \sigma_i^2 (x_i - \bar{x})^2.$$

Robust standard errors are based on estimates of this expression in which the σ_i^2 are replaced with squared OLS residuals, or sometimes slightly more complicated expressions designed to perform better in small samples, see for example Imbens and Kolsar (2012).

When do robust standard errors differ from OLS standard errors?

Compare the expressions above to see that OLS and robust standard errors are (asymptotically) identical in the special case in which σ_i^2 and $(x_i - \bar{x})^2$ are uncorrelated, in which case

$$\sum_i \sigma_i^2 (x_i - \bar{x})^2 \to \bar{\sigma}^2 \sum_i (x_i - \bar{x})^2.$$

If, on the other hand, σ_i^2 and $(x_i - \bar{x})^2$ are positively correlated, then OLS standard errors are too small and robust standard errors will tend to be larger than OLS standard errors. And if σ_i^2 and $(x_i - \bar{x})^2$ are negatively correlated, then OLS standard errors are too big and robust standard errors will tend to be smaller than OLS standard errors. These cases are illustrated in the graphs: in the left panel, the variance of the error terms increases with the distance between x_i and its mean $\bar{x}$, whereas in the right panel observations are most dispersed around the regression line when x_i is at its mean.
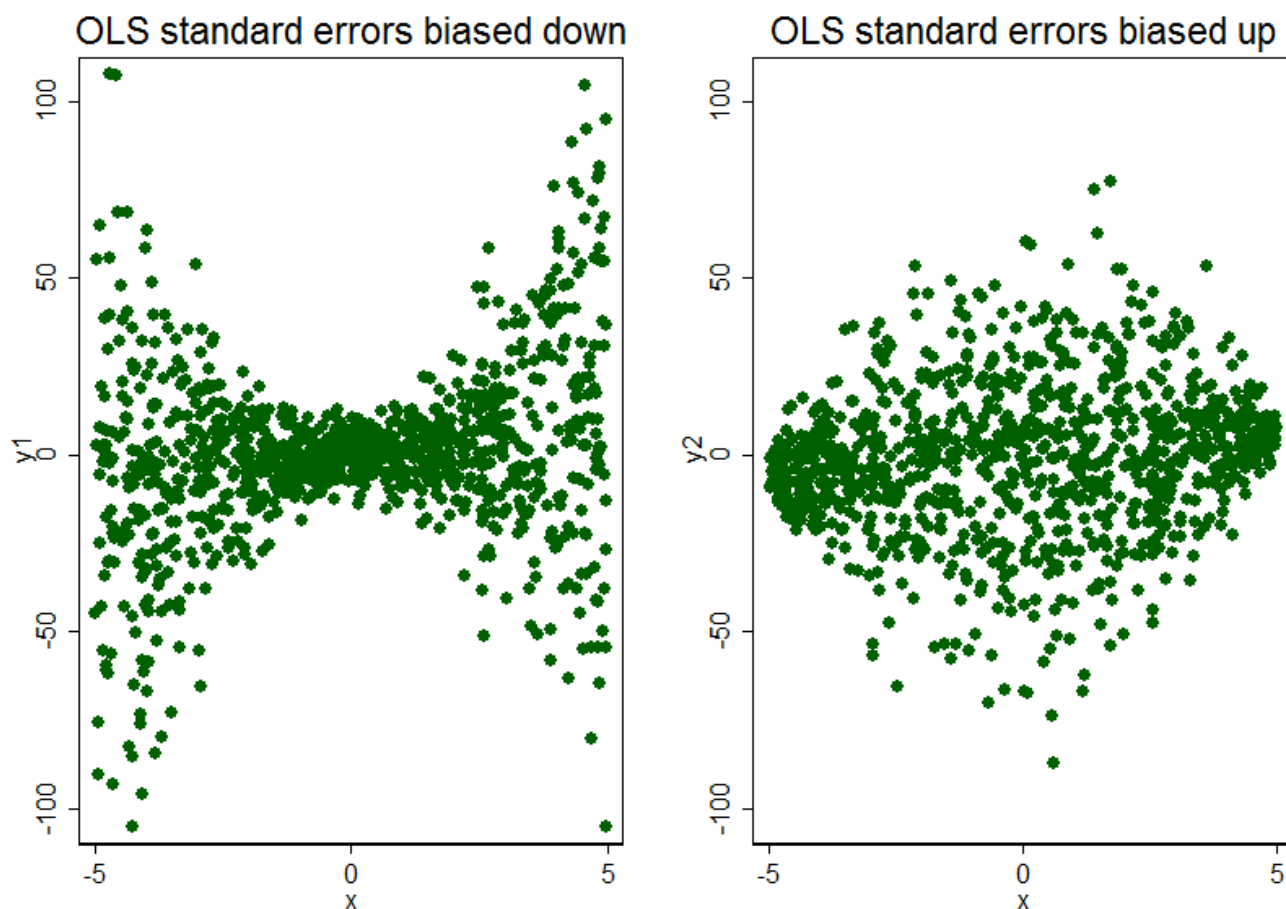
The graphs have been constructed such that the unconditional variance of the errors terms and the variance of x are the same in each graph. But by inspection we can guess that our estimate of the slope is much less precise if the data look like the left panel than the right panel: perform a thought experiment to see that lots of regression lines fit the data in the left panel quite well, but the data in the right panel do a better job pinning down the slope. There is more information about the relationship between y and x in the data in the right panel even though the variance of x and the unconditional variance of the error term are identical.

We see that heteroskedasticity doesn't matter *per se*, what matters is the relationship between the variance of the error term and the covariates—if the errors are heteroskedastic but uncorrelated with $(x_i - \bar{x})^2$, we can safely ignore the heteroskedasticity. To see why this is so, recall that in the homoskedastic case the variance of $\hat{\beta}$ is inversely proportional to $\sum_i (x_i - \bar{x})^2$. If we add one more observation for which x_i happens to equal $\bar{x}$, the variance of our estimate doesn't change—there is no information in that observation about the relationship between y and x. As the draw of x_i moves farther from its mean, the variance of $\hat{\beta}$ falls more and more, because such draws, in the homoskedastic case, are more and more informative.

Now consider the case in which the variance of u_i increases with $(x_i - \bar{x})^2$, as in the left panel of the graph above. When we get one more observation, the amount of information it contains increases with $(x_i - \bar{x})^2$ for the same reasons as the homoskedastic case, but this effect is blunted by the higher

variance of u_i. The amount of information contained in a draw in which x_i is far from its mean is lower than the OLS variance estimate "thinks" there is, so to speak, because the OLS variance estimate ignores the fact that such draws are more highly dispersed around the regression line. The OLS standard errors in this case are too small.

If on the other hand the variance of u_i decreases with $\left(x_i - \bar{x} \right)^2$, then observations of x_i far from its mean both contain more information for the usual reason in the homoskedastic case *and* are less dispersed around the regression line, as in the right panel of the graph above. These observations are even more highly informative than the OLS variance estimate "thinks" they are, and the OLS standard errors will tend to be too *large*. In this case, robust standard errors will tend to be *smaller* than OLS standard errors.

Summarizing.

The upshot is this: if you have heteroskedasticity but the variance of your errors is independent of the covariates, you can safely ignore it, but if you calculate robust standard errors anyways they will be very similar to OLS standard errors. However, if the variance of your error terms tends to be higher when x is far from its mean, OLS standard errors will tend to be biased down, and robust standard errors will tend to be larger than OLS standard errors. In the opposite case in which the variance of the error terms tends to be lower when x is far from its mean, OLS standard errors will tend to be too large, and robust standard errors will tend to be smaller than OLS standard errors. With real data it's commonly but not always going to be the case that the variance of the error will be higher when x is far from its mean, explaining the result that robust standard errors are typically larger than OLS standard errors in economic applications.

Posted on October 31, 2012 at 2:55 pm in Econometrics | RSS feed | Respond | Trackback URL

Tags: econometrics, robust standard error, statistics

3 Comments to "The intuition of robust standard errors"

Joel W
November 8, 2012 at 3:05 pm

Hi Chris. Great to see you blogging again. Thanks for this post; I am sure many applied researchers will find it a very worthwhile read.

cheers,

Joel

EZra
January 10, 2013 at 3:30 am

thanks a lot for your insight!

Kristina
February 16, 2013 at 3:12 am

Thank you so much!!

It helped a lot with my assignment!

Leave a Reply

Enter your comment here...

Categories —   Archives —   RSS links —

Select Category ▼   Select Month ▼

RSS - Posts

RSS - Comments

Search...

Blog at WordPress.com. | The Under the Influence Theme.