**STUDENT QUERY ABOUT COVARIANCE MATRIX FOR LOGIT AND PROBIT PARAMETERS**

A student as has asked

> How do you find the standard error of a regression coefficient in probit or logit regression? In the linear probability model, S&W state that the coefficient standard errors could be calculated using the same heteroskedasticity-robust OLS methods, but they do not state that explicitly for logit and probit regression. Is it implied? Or do we need new methods for logit and probit?

To calculate and justify a covariance matrix for logit and probit parameters does require new methods, which we have not discussed. But as with linear regression or IV, in large samples it is justified to treat the estimators (in this case maximum likelihood or Bayesian posterior means) as approximately normal, and thus to treat the posterior distribution of the parameters as approximately normal and centered at the estimator. So it is approximately correct to use the coefficient estimates, standard errors, and covariance matrix produced by standard statistical program output the same way you would treat coefficient estimates, standard errors, and covariance matrix that come out of a linear regression estimate.

If you want to do better than this approximate large-sample theory, you need to recognize that the likelihood function (or prior times likelihood function if the sample size is small enough that the prior matters) will not have exactly the shape of a normal distribution. The usual ways to use coefficient estimates and covariance matrices rely on treating the likelihood shape as approximately normal (i.e., $e$ to a quadratic function in the parameters). But if the number of parameters is small (2 or 3) you can plot the level curves of the likelihood and get an idea of its actual shape, rather than rely on the normal approximation.

You don't need to know this for the exam, but the large-sample approximation to the shape of the likelihood is based on a second order Taylor expansion of the log likelihood around the MLE. If the parameter vector is $\theta$ and $\hat{\theta}$ is the MLE and the likelihood (or likelihood times prior) is $p(Y,\theta)$, the approximation is

$$p(Y,\theta) \simeq p(Y,\hat{\theta})e^{-.5(\theta-\hat{\theta})'\Sigma^{-1}(\theta-\hat{\theta})} \propto N(\hat{\theta},\Sigma) \text{ pdf.}$$

The $\Sigma$ matrix in this formula is the inverse of

$$-\frac{\partial^2 \log\big(p(Y,\hat{\theta})\big)}{\partial\theta\partial\theta'} .$$