AAE 637 Lab 2: Properties of estimators, with simulation*

2/4/2015

Small sample properties

Bias

Let $\hat{\beta}(x)$ be an estimator for the underlying true parameter β. Then $\hat{\beta}(x)$ is a biased estimator if:

$$E\left[\hat{\beta}(x)\right] - \beta \neq 0$$

In general, nonlinear least squares is a biased estimator. Consider the following model:

$$y = \beta_0 + (\beta_1 \cdot x)^{\beta_0} + \epsilon$$

[R simulation]

Obtaining unbiased estimates in nonlinear models is hard or sometimes impossible. One source of the difficulty is **Jensen's inequality**.

Let X be a random variable and let g:$\mathbb{R} \to \mathbb{R}$ be a convex function. Then:

$$g(E[X]) \leq E[g(X)]$$

Jensen's inequality shows up everywhere. You are familiar with the unbiased estimate of a sample's variance with the $n-1$ bias correction:

$$\hat{s^2}(x) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

A naive estimate of the sample standard deviation is the square root of $\hat{s^2}(x)$.

But the square root is a concave function, so such an estimate is biased. In fact there is no unbiased estimate of a sample's standard deviation when the sample's distribution family is unknown. If the underlying distribution is known, special corrections can eliminate the bias.

Variance

The variance of an estimator $\hat{\beta}(x)$ is:

$$\mathrm{var}\left(\hat{\beta}(x)\right) = E\left[\left(\hat{\beta}(x) - E\left[\hat{\beta}(x)\right]\right)^2\right]$$

Lower variance yields more precise estimates. All else equal, an estimator with a lower variance is preferred over one with higher variance.

Example: **Seemingly unrelated regressions model**.

$$\begin{gathered} y_1 = \mathbf{X}_1 \boldsymbol{\beta}_1 + \epsilon_1 \\ \vdots \\ y_M = \mathbf{X}_M \boldsymbol{\beta}_M + \epsilon_M \end{gathered}$$

*prepared by Travis McArthur, UW-Madison (http://www.aae.wisc.edu/tdmcarthur/teaching.asp)

Used, for example, in estimating demand for various products when our sample consists of consumer purchases. Using OLS on each equation separately yields consistent estimates of the parameters, but if there is correlation among the error terms across the equations, we can achieve lower variance with the seemingly unrelated regressions model by incorporating this information about error correlation.

[R simulation]

Bias-Variance tradeoff

Mean squared error (MSE) can be used to compare biased estimators:

$$\text{MSE}\left(\hat{\beta}\left(x\right)\right) = E\left[\left(\hat{\beta}\left(x\right) - \beta\right)^2\right] = \text{var}\left(\hat{\beta}\left(x\right)\right) + \left(\text{bias}\left(\hat{\beta}\left(x\right)\right)\right)^2$$

Examples of biased estimators that may have lower MSE than unbiased estimators are Bayesian estimators, the James–Stein estimator, and entropy maximization.

Large sample (asymptotic) properties

Consistency

Choose any small distance ϵ. An estimator $\hat{\beta}\left(x\right)$ is consistent for the true value β if the probability that $\hat{\beta}\left(x\right)$ attains a value further away from β than ϵ shrinks to zero as the sample size approaches infinity. In other words, larger and larger samples will place the estimate closer and closer to the true value. An estimator without this property is usually considered unusable since the estimate would converge on the wrong answer as the sample size increases.

OLS consistently estimates parameter values in the presence of heteroskedasticity. However, the probit model fails to be a consistent estimator of the parameters in a binary dependent variable situation with heteroskedasticity. This occurs where

$$Pr\left(y = 1\right) = \Phi\left(\sum_j x_j \beta_j\right)$$

but the variance of Φ is not constant. A particular case of this is:

$$\sigma^2 = \left(\exp\left\{\sum_k z_k \gamma_k\right\}\right)^2$$

where the z's may be a subset of the x's.

[R simulation]

Asymptotic normality

We must estimate the uncertainty of the parameter point estimate when we want to conduct hypothesis tests. The sampling distribution of most estimators approaches normality with an easily-calculated standard error as the sample size becomes large. In small sample sizes the normality approximation may be quite bad, leading to wrong conclusions from hypothesis tests. Bootstrapping, for example, can deal with this problem.

[R simulation]

Tradeoff between precision and robustness

When you incorporate more information into estimation, you can get more precise estimates. If the "information" is assumptions and the information happens to be false, then your estimate may be wrong. Maximum likelihood estimation makes assumptions about the whole distribution of the error term, so its estimates can be precise. Assuming the wrong distribution can lead to inconsistency.

[R simulation]