LIMITED MOBILITY BIAS CORRECTION

SIMEN GAURE

Abstract. A description of what limited bias correction is. And how to compensate for it.

1. Introduction

We assume we have an OLS model in matrix form

$$Y = X\beta + D\alpha + F\gamma + \epsilon \tag{1}$$

where X is a $(n \times k)$-matrix, D is a $(n \times g_1)$-matrix, and F is a $(n \times g_2)$-matrix. It is assumed that D and F are matrices of dummy variables and that both g_1 and g_2 are very large (as in $10^5 - 10^7$). The **lfe**-package estimates α and γ as OLS coefficients. Some applications study the variances $\mathrm{var}(D\alpha)$ and $\mathrm{var}(F\gamma)$, as well as the covariance $\mathrm{cov}(D\alpha, F\gamma)$. It was shown in [1] that if one uses the estimates $D\hat{\alpha}$ and $F\hat{\gamma}$ for this purpose, the resulting variances are positively biased, and the covariance is typically negatively biased. The biases can be large enough to change the sign of the covariance. They also provided explicit formulas for the size of the bias, in terms of the trace of some very large matrices.

In short, the bias for $\mathrm{var}(D\hat{\alpha})$ can be computed as

$$\delta_\alpha = \sigma_\epsilon^2 \operatorname{tr}(A), \tag{2}$$

where A is a large matrix depending only on X, D and F. The unbiased variance can then be estimated as

$$\mathrm{var}(D\alpha) = \mathrm{var}(D\hat{\alpha}) - \sigma_\epsilon^2 \operatorname{tr}(A) \tag{3}$$

In some typical instances, the matrix A will be too large to handle directly, being of size g_1. The bias for $\mathrm{var}(F\hat{\gamma})$ and the covariance, are of the same form.

2. How lfe handles this

In [2] a method for estimating δ_α was outlined. It uses the fact that

$$\operatorname{tr}(A) = \mathrm{E}(x^t A x), \tag{4}$$

if x is a vector of independent random variables x_i with $\mathrm{E}(x_i) = 0$ and $\mathrm{var}(x_i) = 1$. In fact, each x_i can optimally be choosen as drawing from $\{-1, 1\}$ with uniform probability. This works because even if A is too large to be handled directly, it is possible to compute Ax for a vector x. Thus, **lfe**'s `fevcov` routine uses a sample mean to estimate $\operatorname{tr}(A)$.

Date: November 12, 2015.

3. Another method

Formula 3 suggests another method for estimating $\mathrm{var}(D\alpha)$. We rewrite it as

$$\mathrm{var}(D\hat{\alpha}) = \mathrm{var}(D\alpha) + \sigma_\epsilon^2 \,\mathrm{tr}(A) \tag{5}$$

Now, $b = \mathrm{var}(D\alpha)$ and $a = \mathrm{tr}(A)$ are constants, so we can write

$$\mathrm{var}(D\hat{\alpha}) = a\sigma_\epsilon^2 + b \tag{6}$$

If we resample and scale the residuals $\hat{\epsilon}$, and do the estimation of $\hat{\alpha}$ over again, we will obtain a series of different observations of formula 6. We may then do an OLS on model 6, and the estimate $\hat{a}$ will be an estimate for $\mathrm{tr}(A)$ which can be used in equation 3. The $\mathrm{var}(D\hat{\alpha})$ is not normally distributed around its mean, this typically skews $\hat{b}$ more than $\hat{a}$, so $\hat{a}$ is a better estimate. **lfe** does not implement this method.

References

1. M. Andrews, L. Gill, T. Schank, and R. Upward, *High wage workers and low wage firms: negative assortative matching or limited mobility bias?*, Journal of the Royal Statistical Society(A) **171(3)** (2008), 673–697.
2. S. Gaure, *Correlation bias correction in two-way fixed-effects linear regression*, Stat **3** (2014), no. 1, 370–390.

Ragnar Frisch Centre for Economic Research, Oslo, Norway