DRAFT

Tangential Study – ROM Variance Evaluation

Ratio of Means TABLING Addition

Larry Bednar

January 3, 2002

Purpose

This document examines whether ratio of means estimates will predict population ratios with a reduced estimated variance compared to a "standard" estimator that evaluates the mean value of a sample unit-level ratio taken over an entire population.

Report

Background

The existing PNW FIA TABLING application can be used to estimate ratio parameters of the population. However, the application can only do this when the user request a population sample unit mean of a sample unit level ratio statistic.

Difficulties arise under two causes:

1. When the denominator of the statistic takes a zero value on a sample unit, as the ratio is then mathematically "undefined". At present, the program assumes such a case to represent a "zero" value of the statistic for that sample unit. As a result, the generated population estimate generated by the application makes use of a large number of zero values. The associated computation of estimation error reflects the inclusion of those zero values, which results in an inflated estimate of variance.
2. The application's use of row and column categories is such that zero values are assumed for any sample unit where the characteristic described by the row and column categories of an estimate are not found. This results in an inflated estimate of error under the same mechanism (inclusion of many zero values) as in point 1 above.

This motivation is partly motivated by the need for an estimator that avoids the problems outlined above. Before work on modifcation of the TABLING application begins in earnest, it seems prudent to perform a high-level examination of the behavior of ratio statistics to determine whether the use of ratio-of-means estimators will provide that benefit.

Theoretical Analysis

Start with the form of estimator for variance of a ratio estimator provided by Cochran in equation 6.12.

$$V(\hat{Y}_R) = \frac{(1-f)}{n\bar{X}^2}\left(s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R}s_{yx}\right) \tag{1}$$

Where s_{yx} represents the covariance of the two statistics used to construct the numerator and denominator of the ratio statistic. In cases such as those outlined above, the covariance will assume a large value, since x and y will simultaneously assume zero values on many sample units. Further examination of the formula for covariance is helpful:

$$Cov(y,x) = \frac{\sum (y_i - \bar{Y})(x_i - \bar{X})}{N-1} \quad (2)$$

When both y and x deviate in the same direction from their sample means, the contribution to the covariance will be positive. For a nonnegative statistic, the case where both x and y are zero will then contribute a positive value to the covariance calculation. The rightmost term of equation 1 will then result in a reduction of the estimated variance. This loosely indicates that the situations described in the “background” section above will be accounted for, and the variance estimate adjusted downward reflect the influence of those circumstances.

Simulation Results

A simple simulation was generated in SAS V8 (program “rom_var.sas”) to verify the analysis presented above. A random sample of 100 was generated with two random variables Y (the numerator) and X (the denominator).

X was constructed as normal random variate located close enough to zero that about 40% of the randomly generated values fell below zero. Negative values were converted to zero.

For each instance in which X was greater than zero, a value of Y was generated, where

y = x + 0.3+ normal(0).

Y was constructed so that a noticable proportion of it’s values would be below zero when the value of x was zero. These negative values were then converted to zero.

An estimate of a population parameter Y/X was produced under two techniques:

1. A “sample unit” estimate of Y/X was produced for each observation. If X was zero, the sample unit vale was taken to be zero.
2. A ratio of means estimate of Y/X was produced and the variance computed using Cochran’s equation 6.12.

The results of 100 simulations showed that the ratio-of-means estimate provided an average 50% reduction in the coefficient of variance compared to estimates developed from sample unit means. Furthermore the values based on ratio-of-means estimators were much more reliable – estimates based on sample unit means were much more unstable and in some cases represented extreme differences from the ratio of means estimates.

References

Cochran, William G. Sampling Techniques, 3d ed. 1977 Wiley, NY.