



ELSEVIER

Economics Letters 70 (2001) 15–19

**economics
letters**

www.elsevier.com/locate/econbase

An almost unbiased estimator of the coefficient of variation

Robert Breunig*

Australian National University, Department of Statistics and Econometrics, Canberra, ACT 0200 Australia

Received 27 September 1999; accepted 15 July 2000

Abstract

A bias correction method for inequality measures is proposed. The coefficient of variation squared (CV^2) is used as an example and its sampling properties, bias, and mean squared error are provided. CV^2 is shown to be downward biased for positively skewed distributions. A bias corrected estimator is provided. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Coefficient of variation; Small sample bias correction; Inequality measurement

JEL classification: C13; D31

1. Introduction

The objective of this paper is to show how the large- n expansion can be used to derive the approximate bias and mean squared error (MSE) of the coefficient of variation squared (CV^2), an inequality index. The approach is to study the sample properties of CV^2 without imposing any restrictions on the form of the distribution of the economic variable under consideration. I use a large- n approximation to derive the approximate bias and MSE of CV^2 . This technique is straightforward to employ and provides simple expressions.

Inequality indices are popular tools for comparing inequality across regions and over time in the evaluation of policy. CV^2 (along with most commonly used inequality measures such as Theil's Entropy indices and Atkinson's measure, see Sen, 1973, and Ravallion, 1994 for reviews) is a ratio of random variables and is biased in small samples. The expected value of such inequality measures will be of the form

$$E\hat{I} = I_0 + O\left(\frac{1}{n}\right) \quad (1)$$

*Corresponding author. Tel.: +61-2-6249-4991; fax: +61-2-6249-0087.

E-mail address: Robert.Breunig@anu.edu.au (R. Breunig).

where I_0 is the true population value of the inequality measure and the other terms approach zero as the sample size, n , becomes very large. This bias problem is generally ignored in inequality estimation.¹

The squared coefficient of variation (CV^2) has been used as a measure of inequality (see, for example, Lehrer and Nerlove, 1981; Blackburn and Bloom, 1990, and Cancian et al., 1992) and this interest in CV^2 has been justified recently in the *Letters* by Formby et al. (1999) who have provided the relationship between CV^2 and the rankings from second-order stochastic dominance ordering of normalized distributions. Despite the use of CV^2 , not much is known about its sampling properties, especially under non-normality

This note shows that the sample CV^2 is biased – that is it under/over-estimates the true CV^2 – and a neat analytical condition under which the bias will be positive or negative is provided. Such an analytical condition, as far as the author is aware, is not available in the literature and it indicates that the sample CV^2 will generally be an underestimate of the true CV^2 for positively-skewed distributions. Since income distributions are almost always positively-skewed, income inequality estimates based on CV^2 will underestimate the true degree of inequality, a problem that will be particularly severe in small samples. An almost unbiased estimator is proposed, using the expression for bias.

2. Bias in sample CV^2

Let us consider the population mean model as

$$y_i = \mu + u_i \quad (2)$$

where y_i is the i th observation on the study variable, μ and σ^2 are the unknown population mean and variance, respectively, and u_i is an unobserved error variable. We assume that the elements u_i are independently and identically distributed such that

$$\begin{aligned} Eu_i &= 0, Eu_i^2 = \sigma^2, Eu_i^3 = \gamma_1 \sigma^3, Eu_i^4 = (\gamma_2 + 3)\sigma^4 \\ Eu_i^5 &= (\gamma_3 + 10\gamma_1)\sigma^5, Eu_i^6 = (\gamma_4 + 10\gamma_1^2 + 15\gamma_2 + 15)\sigma^6 \end{aligned} \quad (3)$$

where γ_1 and γ_2 are Pearson's measures of skewness and kurtosis of the distribution. Likewise the quantities γ_3 and γ_4 can also be regarded as measures of deviation from normality. See Kendall and Stuart (1977, p.72) for expression of (3) in terms of cumulants. For normal distribution, γ_1 , γ_2 , γ_3 , and γ_4 are zero while for a symmetric, non-normal distribution only γ_1 and γ_3 are zero. Thus non-zero values of γ_1 to γ_4 indicate a departure from normality. We also assume that $\mu \neq 0$.

Population CV^2 is given by $\theta \equiv \sigma^2/\mu^2$ and the sample CV^2 by $\hat{\theta} \equiv s^2/\bar{y}^2$, where \bar{y} and s^2 are the sample mean and variance, respectively, written for a sample of size n as

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i, s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2. \quad (4)$$

¹Maasoumi and Theil (1979) consider the effect of the *population* skewness and kurtosis on the *population* value of Theil's measures of inequality. However, they do not consider the problem of sample bias of the measures.

In what follows, I present large- n approximations for the bias and MSE of $\hat{\theta}$. Exact analytical results under non-normal distributions are difficult to obtain and depend on the unknown distribution of u . The results here are for any non-normal distribution having finite moments of at least six.

Proposition 1. *If the errors follow (3), the bias of $\hat{\theta}$, up to $O(n^{-1})$, and the MSE of $\hat{\theta}$, up to $O(n^{-2})$, are respectively given as:*

$$\text{Bias}(\hat{\theta}) = \frac{\theta^{3/2}}{n} [3\theta^{1/2} - 2\gamma_1] \tag{5}$$

and

$$\begin{aligned} \text{MSE}(\hat{\theta}) = & \frac{\theta^2}{n} \left[\gamma_2 + 4\theta - 4\gamma_1\theta^{1/2} + 2\frac{n}{n-1} \right] \\ & + \frac{\theta^2}{n^2} \left[\theta(24\gamma_2\frac{n}{n-1} + 20\gamma_1^2\left(\frac{n+1}{n-1}\right) + 20\frac{n}{n-1}) \right. \\ & \left. - 4\theta^{1/2}\left(\gamma_3 + 4\gamma_1\frac{n}{n-1}\right) + 75\theta^2 - 108\theta^{3/2}\gamma_1 \right] \end{aligned} \tag{6}$$

The approximate results in the above proposition are for normal ($\gamma_1 = \gamma_2 = 0$) as well as non-normal errors.²

From (5) it is clear that up to the order of approximation considered, *Bias* ($\hat{\theta}$) is positive for negatively-skewed distributions ($\gamma_1 < 0$), and it is negative for positively-skewed ($\gamma_1 > 0$) distributions provided

$$\gamma_1 > \frac{3}{2} (CV) \tag{7}$$

where $CV = \theta^{1/2}$. That is $\hat{\theta}$ provides an over-estimation of θ for negatively-skewed distribution, and an underestimation for positively-skewed distributions provided (7) holds. When $\gamma_1 = 3/2CV$, the bias in $\hat{\theta}$ vanishes. Since income distributions are usually positively skewed, the use of $\hat{\theta}$ to measure income inequality will underestimate the extent of income inequality. For example, when u is lognormally distributed, $\gamma_1 = (e^{\sigma^2} - 1)^{1/2}(e^{\sigma^2} + 2)$ and $CV = (e^{\sigma^2} - 1)^{1/2}$. Thus (7) is satisfied and $\hat{\theta}$ will underestimate θ . Similarly if u were exponentially distributed, $\gamma_1 = 2$ and $CV = 1$ and hence $\gamma_1 > 3/2(CV)$.

Since the estimated $\hat{\theta}$ may be used in evaluation of policy or in development of rankings of income distributions, it is useful to develop a bias-corrected estimator. This is given by

$$\tilde{\theta} = \hat{\theta} - \widehat{\text{Bias}}(\hat{\theta}) \tag{8}$$

where $\widehat{\text{Bias}}(\hat{\theta})$ is the *Bias* ($\hat{\theta}$) in (5) with θ replaced by $\hat{\theta}$ and γ_1 replaced by $\hat{\gamma}_1 = \frac{1}{n} \sum^n (y_i - \bar{y})^3 / s^3$.

The estimator $\tilde{\theta}$ is an almost unbiased estimator of θ in the sense that bias of $\tilde{\theta}$ is zero up to $O(n^{-1})$.

²Proofs of Propositions 1 and 3 are available from the author upon request.

Proposition 2. *If the errors follow (3), the bias of $\tilde{\theta}$, up to $O(n^{-3/2})$, is zero. That is*

$$\text{Bias}(\tilde{\theta}) = 0 \quad (9)$$

The proof of Proposition 2 follows by first noting that, under (3), $\hat{\theta}^r = \theta^r + O_p(1/\sqrt{n})$ for a constant $r > 0$, and $\hat{\gamma}_1 = \gamma_1 + O_p(1/\sqrt{n})$. Substituting these in $\text{Bias}(\hat{\theta})$ gives $\text{Bias}(\tilde{\theta}) = \text{Bias}(\hat{\theta}) + O(1/n^2)$, and hence from (8) $\text{Bias}(\tilde{\theta}) = O(1/n^2)$ which proves the result in Proposition 2.

Proposition 3. *If the errors follow (3), the mean squared error of $\tilde{\theta}$, up to $O(n^{-2})$, is*

$$\begin{aligned} \text{MSE}(\tilde{\theta}) = & \frac{\theta^2}{n} \left[\gamma_2 + 4\theta - 4\gamma_1\theta^{1/2} + 2\frac{n}{n-1} \right] \\ & + \frac{\theta^{5/2}}{n^2} \left\{ 8\theta^{1/2}\gamma_2 \left(1 + \frac{3n}{n-1} \right) - 8\gamma_3 + 6\gamma_1\gamma_2 - 4\theta^{1/2}\gamma_1^2 \left(2 - 5\frac{n}{n-1} \right) \right. \\ & - 4\gamma_1 \left(6 + \frac{n}{n-1} \right) + 48\theta^2\gamma_1 - 108\theta\gamma_1 + 8\theta^{1/2}\gamma_1 - 12\theta^{3/2}\gamma_2 \\ & \left. + \theta^{1/2} \left(\frac{20n}{n-1} - 12\theta^{1/2} + \theta(66 - \frac{24n}{n-1}) - 48\theta^2 \right) \right\} \end{aligned} \quad (10)$$

When the distribution is normal, $\gamma_1 = \gamma_2 = \gamma_3 = 0$, and,

$$\text{MSE}(\tilde{\theta}) - \text{MSE}(\hat{\theta}) = -\frac{\theta^{7/2}}{n^2} \left[12 + \theta^{1/2} \left(9 + \frac{24n}{n-1} \right) + 48\theta^{3/2} \right] \quad (11)$$

which is clearly negative. Thus even under normality bias will be present and the suggested adjusted estimator will reduce bias and will have a lower mean squared error. As discussed above, however, the bias problem is most severe for skewed distributions. For non-normal distributions, it will not in general be possible to show that $\text{MSE}(\tilde{\theta}) - \text{MSE}(\hat{\theta})$ is negative. It will in fact depend upon the shape of the distribution.

Finally, I note from Propositions 1, 2, and 3 that for large samples ($n \rightarrow \infty$), $\text{Bias}(\hat{\theta}) \rightarrow 0$ and $\text{Bias}(\tilde{\theta}) \rightarrow 0$ and

$$\lim_{n \rightarrow \infty} nV(\hat{\theta}) = \lim_{n \rightarrow \infty} nV(\tilde{\theta}) = \theta^2 [\gamma_2 + 4\theta - 4\gamma_1\theta^{1/2} + 2] \quad (12)$$

which is the asymptotic variance of $\hat{\theta}$ and $\tilde{\theta}$. A consistent estimator of the asymptotic variance can be written by substituting $\hat{\theta}$, $\hat{\gamma}_1$ and $\hat{\gamma}_2$ for θ , γ_1 and γ_2 , respectively, in (12), where $\hat{\gamma}_2$, is given by

$$\hat{\gamma}_2 = (n-1)^{-1} \sum \left(\frac{(y_i - \bar{y})^4}{s^4} \right) - 3.$$

3. Conclusions

The bias corrected estimator (8) can make a substantial difference in the estimated CV^2 even for relatively large samples. (See Breunig, 1998, for an example using a sample of 2400 households from Kenya where the correction provides a change of over 5% in the inequality measure.) Frequently

inequality estimates are calculated using much smaller samples and analysts should be aware of the large bias possible from such estimates.

It may be possible to improve upon the suggested bias corrected estimator through a simple iterative process. By calculating the bias correction, then replacing $\hat{\theta}$ with $\tilde{\theta}$ in (8) and recalculating the bias-corrected estimator and repeating this process until $\hat{\theta}$ and $\tilde{\theta}$ converge, we may get an improved bias-corrected estimator. I propose to explore this question in future research. Another area for future research is the application of this technique to other inequality measures.

Acknowledgements

I would like to thank Aman Ullah, Jang-Ting Guo, and an anonymous referee for comments on an earlier draft of this paper. I am also grateful to an anonymous referee for helpful comments. All remaining errors are my own.

References

- Blackburn, M., Bloom, D. 1990. Changes in the structure of family income inequality in the US and other industrialized nations during the 1980s. Mimeo.
- Breunig, R., 1998. Almost unbiased estimator of an inequality measure. Mimeo, Australian National University.
- Cancian, M., Danziger, S., Gottschalk, P., 1992. Working wives and family income inequality among married couples. In: Danziger, S., Gottschalk, P. (Eds.), *Uneven Tides: Rising Inequality in America*. Russell Sage Foundation, New York.
- Formby, J., Smith, W., Zheng, B., 1999. The coefficient of variation, stochastic dominance and inequality: a new interpretation. *Economics Letters* 62, 319–323.
- Kendall, M., Stuart, L. (Eds.), 1977. 4th Edition. *The Advanced Theory of Statistics*, Vol. 1. Charles Griffin, London.
- Lehrer, E., Nerlove, M., 1981. The impact of female work on family income distribution in the United States: black–white differentials. *Review of Income and Wealth* 27, 423–431.
- Maasoumi, E., Theil, H., 1979. The effect of the shape of the income distribution on two inequality measures. *Economics Letters* 4, 289–291.
- Ravallion, M., 1994. *Poverty Comparisons*. Harwood Academic, Langhorne, PA.
- Sen, A., 1973. *On Economic Inequality*. Oxford University Press, Dehli.