E c o n o m i c s B u l l e t i n

Inequality and the accounting period

Quentin Wodon
World Bank

Shlomo Yitzhaki
Hebrew University

Abstract

Income inequality typically declines with the length of time taken into account for measurement. This note derives an exact analytical relationship between the accounting period and inequality as measured by the Gini index. The formal relationship is similar to the decomposition of the coefficient of variation. The methodology is illustrated with panel data on urban wages from Mexico. It is found that the effect of the accounting period on inequality is sensitive to the properties of the Gini correlations between the periodical incomes. Reporting this type of correlation enables the evaluation of the impact of the length of the accounting period on inequality.

This paper was funded by a research grant from the World Bank. The views expressed here are those of the authors and need not reflect those of the World Bank.
Citation: Wodon, Quentin and Shlomo Yitzhaki, (2003) "Inequality and the accounting period." *Economics Bulletin,* Vol. 4, No. 36 pp. 1−8
Submitted: September 12, 2003. **Accepted:** December 6, 2003.
URL: http://www.economicsbulletin.com/2003/volume4/EB−03D30005A.pdf

1. Introduction

It is well known that income inequality typically declines with the length of time taken into account for measurement (e.g., Creedy 1979, 1991). Recent evidence from Gibson, Huang and Rozelle (2001) demonstrates that inequality in urban China is low relative to other countries in large part because the accounting period is based on a full year of income, while in many other countries, surveys record household income for a month or less. They estimate that if the accounting period in China would be one month, then inequality would be between 17 and 69 percent higher. Similar findings have been observed for the United States and Germany by Burkhauser and Poupore (1997). Other statistics may also be sensitive to the accounting period. For example, Behrman and Taubman (1989) find that the estimated inter-generational correlation of parental income and off-springs is 0.58 when ten years of earnings are used, compared to 0.37 for a single year (see also Bowles and Gintis 2002, for additional evidence on the impact of the accounting period.)

In this note, using information on the Gini correlation of income between different time periods, we derive an exact analytical relationship between inequality and the length of the accounting period for the Gini index of inequality. The methodology is illustrated with panel data on urban wages from Mexico. While we focus our discussion on the accounting period, our decomposition could have many other useful applications since it is very similar to the decomposition of the widely used coefficient of variation, and thereby simple to implement.

2. Methodology

This section derives the relationship between the Gini index of inequality for the weighted average of several time periods and the Gini inequality indices for each period taken separately. To simplify notation, we restrict the proof to two periods. The extension to many periods is immediate. Let (Y_1,Y_2) be drawn from a bi-variate continuous distribution, where Y_i is the income distribution in period i. Let $Y_0 = b_1Y_1 + b_2 Y_2$, where $b_i > 0$ (i=1,2) is a constant. The value of the constant determines whether we are dealing with the sum of the incomes, or the average income, possibly weighted. For example, if $b_1= b_2= 0.5$ then Y_0 represents the straight average income measured over two periods. Denoting by $F(Y_i)$ the cumulative distribution and μ_i the expected income, the Gini coefficient (Lerman and Yitzhaki 1984) is:

$$G_i = 2 \text{ cov}(Y_i,F(Y_i))/ \mu_i \ . \qquad (1)$$

Denote by $\Gamma_{ij} = \dfrac{\text{COV}(Y_i, F(Y_j))}{\text{COV}(Y_i, F(Y_i))}$, i, j = 0, 1, 2, the Gini correlation between incomes from periods Y_i and Y_j, or by extension between income from one period and average income. As discussed in Schechtman and Yitzhaki (1987, 1999), the properties of the Gini correlations are a mixture of Pearson's and Spearman's correlation coefficients. In particular, Γ_{ij} is bounded by minus one and one, but Γ_{ij} is not necessarily equal to Γ_{ji}. Define also $D_{i0} = \Gamma_{i0} - \Gamma_{0i,}$ for i=1,2 (here, the Gini correlations are taken between the income in each of the two periods and the average income over time), and $a_i = b_i (\mu_i/\mu_0)$, where $\mu_i > 0$.

Proposition:

(a) The following identity holds:

$$G_0^2 - [a_1 D_{10} G_1 + a_2 D_{20} G_2] G_0 = a_1^2 G_1^2 + a_2^2 G_2^2 + a_1 a_2 G_1 G_2 (\Gamma_{12} + \Gamma_{21}). \quad (2)$$

(b) Provided that $D_{i0}=0$, for i=1,2, and $\Gamma_{12} = \Gamma_{21} = \Gamma$, then:

$$G^2_0 = a_1{}^2 G_1{}^2 + a_2{}^2 G_2{}^2 + 2a_1 a_2 G_1 G_2 \Gamma. \quad (3)$$

Proof: The proof is in the appendix.

The extension of equations (2) and (3) to k periods is trivial. Let $Y_0 = \sum_{i=1}^{k} b_i Y_i$, and $a_i = b_i \mu_i / \mu_0$, then

$$G_0^2 - G_0 \sum_{i=1}^{k} a_i D_{i0} G_i = \sum_{i=1}^{k} a_i^2 G_i^2 + \sum_{i=1} \sum_{i \neq j} a_i a_j G_i G_j \Gamma_{ij} \quad . \quad (4)$$

If $D_{i0}=0$, for i=1, …, k and $\Gamma_{ij} = \Gamma_{ji}$, then:

$$G_0^2 = \sum_{i=1}^{k} a_i^2 G_i^2 + 2\sum_{i=1}^{k} \sum_{i<j} a_i a_j G_i G_j \Gamma_{ij} \quad . \quad (5)$$

Equation (5) is identical in its structure to the decomposition of the variance and the coefficient of variation. For it to hold, the Gini correlations between each pair of variables Y_0, Y_1, and Y_2 must be equal. Schechtman and Yitzhaki (1987) show that a sufficient condition for $\Gamma_{ij} = \Gamma_{ji}$ is that the variables are exchangeable up to a linear transformation. Examples of such distributions are the multinormal and the multivariate lognormal, provided that $\sigma_i = \sigma_j$, where σ is the logarithmic standard deviation. If the Gini correlations between pairs of variables are not equal, we need to use equation (4), where each "violation" of the equality of the Gini correlations is captured by an additional term in the decomposition (hence, we can treat each violation separately and evaluate its effect on the decomposition; in particular we can see whether the violation tends to increase or decrease overall inequality). Since $a_i < 1$ for all i, it is easy to see that the shorter the accounting period is, the higher the inequality will be.

Note that lower values for Γ_{ij} and Γ_{ji} will yield a larger decrease in the Gini index of inequality over several periods of time, as compared to the average level of inequality for the various periods taken separately. The magnitude of the (Gini) correlations between incomes in different time periods is thus a key factor in determining the impact of the length of the accounting period on measured inequality.

Note also that it is not straightforward to assess a priori what values the Gini correlations will take. The length of the various periods taken into account may affect the value of the Gini correlations in several ways. First, we would assume that with observations corresponding to longer periods of time, there will be less noise in the data, so that the longer the period the larger will be the denominator in the expression of the correlation (as would the variance under less noisy data), and therefore the absolute value of the Gini correlations between longer periods will be higher. On the other hand, the same reduction in noise may also increase the absolute value of the numerator, since less noise will tend to increase the covariance between income in one period and the rank of the individual in the distribution of income in the other period. The net effect of these two factors related to the impact of noise in the data is difficult to determine a priori. Independently of the issue of noise in the data, when working with longer periods of time, employment mobility over time (with some workers gaining and some loosing) could also lead to lower correlations between two income observations. This is because our guess is that the impact of employment mobility would tend to be larger when the time periods under review are themselves longer. In general, experiments with different data sets are needed to enable us to get a better understanding of the issues at hand, and a better feeling for the values that Gini correlations may take in various settings.

3. Illustration

We use data from Mexico's 1996 National Urban Employment Survey, a panel of individuals living in 16 different metropolitan areas. The individuals are interviewed on a quarterly basis for five quarters. Table I gives the various statistical results. The first part of the table gives the Gini indices of wage inequality for each quarter among males aged 15 to 65 (we have 6262 observations in the sample). Wages for the five quarters are converted into real terms using the Consumer Price Index. The quarterly Gini indices vary from 0.5672 to 0.6153. Giving equal weight to the five periods, the income shares per quarter vary from 19.30 percent to 20.99 percent. The second part of the table gives the matrix of pair-wise correlations Γ_{ij} between the various periods. As expected, the farther the periods, the lower the correlations, with the exception that for most quarters, the correlation with the quarter four periods later (or earlier) increases a bit, which suggests mild seasonality in earnings. Another interesting property is that for most periods, there are differences between Γ_{ij} and Γ_{ji}, which means that the marginal distributions are not exchangeable. This implies that the shape of the marginal distribution change in a non-linear way.[1]

The third and fourth parts of the table provides the pair-wise correlations Γ_{i0} and Γ_{0i}. The correlations with aggregate income tends to be higher than the correlations with another period, which is reasonable since the aggregate income includes as a component the periodical income. The larger the number of quarters included in average income, the lower the correlations are, since each period of time represents a smaller share of overall income.

The last part of the table provides the decomposition from equation (4) applied to the various accounting periods, from two to five quarters. Figure 1 shows how the Gini index of inequality is reduced by taking more periods in consideration. Specifically, the Gini index for average income between quarters one and two is 0.5855. For the first three quarters, it is 0.5711. For the first four periods, it is 0.5551. For all five quarters, it is 0.5475.

It is interesting to note, that all the terms reflecting the deviation of the Gini correlations of each period with the overall distribution are positive, which according to equation (4) imply that they increase the Gini of the overall period. It is not clear whether this property is typical to other data sets, or whether one should expect all the terms to have the same sign.

With five quarters, a total of 48 correlations must be computed in order to implement the decomposition in equation (4). In the variance-like decomposition of equation (5), the number of correlations required is reduced to 20 for five quarters. How good an approximation is equation

[1] If the only change that occurs between the periodical distributions is a linear shift, then the Gini correlations will be equal. Further research is needed to develop a statistical test to show whether the difference between Gini correlation coefficients is significant.

(5)? Table 1 can be used to show that using equation (5) would have yielded a Gini for the five periods of 0.5388, which accounts for 98.4 percent of the actual Gini obtained with equation (4). Equation (5) accounts for an even larger share of the actual Gini for fewer periods. Hence, it is a useful first order approximation for evaluating the impact of the accounting period on the Gini.

4. Conclusion

The Gini index of inequality for a sum of random variables can be decomposed in a way that resembles the decomposition of the variance, plus an additional term, which reflects the deviation of the underlying distributions from exchangeability up to a linear transformation. To be able to make quantitative inferences on the effect of the accounting period on the Gini coefficient, we should evaluate the Gini correlation for different types of variables (households or individual, monthly or quarterly etc.). If we find the magnitude of those correlations, and if they are relatively stable over time (and possibly over countries), we may be able to predict the impact of the accounting period on inequality in quite general settings.

Appendix: Proof of equations (2) and (3).

For simplicity, we define $a_i = b_i (\mu_i/\mu_0)$, and provide the decomposition for $G_0 = 2COV[a_1Y_1+ a_2Y_2, F(Y_0)]$. This normalization enables us to work with variables with unit means, but it does not affect the generality of the proof. Using the properties of the covariance we can write:

$$G_0 = 2COV[a_1Y_1+ a_2Y_2, F(Y_0)] = a_1\, 2COV[Y_1, F(Y_0)] + a_2\, 2COV[Y_2, F(Y_0))] \qquad (A.1)$$
$$= a_1\, \Gamma_{10}G_1 + a_2\, \Gamma_{20}\, G_2\ .$$

Define the identity:

$$\Gamma_{i0} = \Gamma_{0i} + D_{i0} \text{ for } i=1,2, \qquad (A.2)$$

where D_{i0} is the difference between the two Gini correlations defined between Y_0 and Y_i . Using (A.1) and (A.2), we get:

$$G_0 = a_1 (\Gamma_{01}+ D_{10})\, G_1 + a_2 (\Gamma_{02} + D_{20})\, G_2.$$

Rearranging terms:

$$G_0 - a_1\, D_{10}G_1 - a_2D_{20}G_2 = a_1\, \Gamma_{01}\, G_1 + a_2\, \Gamma_{02}\, G_2.$$

Using the properties of the covariance:

$$\Gamma_{01} = \frac{\text{cov}(Y_0, F(Y_1))}{\text{cov}(Y_0, F(Y_0))} = \frac{1}{\text{cov}(Y_0, F(Y_0))}\{a_1\text{cov}(Y_1, F(Y_1)) + a_2\text{cov}(Y_2, F(Y_1))\} =$$
$$= \frac{a_1G_1 + a_2G_2\Gamma_{21}}{G_0}\ .$$

Writing Γ_{02} in a similar manner, we get equation (2):

$$G_0^2 - [a_1D_{10}G_1 + a_2D_{20}G_2]G_0 = a_1G_1(a_1G_1 + a_2G_2\Gamma_{21}) + a_2G_2(a_1\Gamma_{12}G_1 + a_2G_2)$$
$$= a_1^2G_1^2 + a_2^2G_2^2 + a_1a_2G_1G_2(\Gamma_{12} + \Gamma_{21}).$$

Assuming equality of the Gini correlation coefficients between Y_0 and Y_1 sets $D_{10} = 0$. A similar assumption for Y_2 and Y_0 sets $D_{20} = 0$. The assumption $\Gamma = \Gamma_{12} = \Gamma_{21}$ completes the proof of (3).

References

Behrman, J. R. and P. Taubman (1989) "Is Schooling 'Mostly in the Genes'?" *Journal of Political Economy* **97**: 1425-46.

Burkhauser, R. V. and J. G. Poupore (1997) "A Cross-National Comparison of Permanent Inequality in the United States and Germany" *Review of Economics and Statistics* **79**: 10-17

Bowles, S. and H. Gintis (2002) "Intergenerational Inequality" *Journal of Economic Perspectives*, forthcoming.

Lerman, R. and S. Yitzhaki (1984) "A Note on the Calculation and Interpretation of the Gini Index" *Economics Letters* **15**: 363-68.

Creedy, J. (1979) "The Inequality of Earning and the Accounting Period" *Scottish Journal of Political Economy* **26**: 89-96.

Creedy, J. (1991). "Lifetime Earning and Inequality," *Economic Record* 67: 46-58.

Gibson, J., Huang, J. and Rozelle, S. (2001) "Why is Income Inequality so Low in China Compared to Other Countries? The Effect of Household Survey Method" *Economics Letters* **71**: 329-333.

Schechtman, E. and S. Yitzhaki (1987) "A Measure of Association Based on Gini's Mean Difference" *Communications in Statistics*: Theory and Methods **A16**: 207-231.

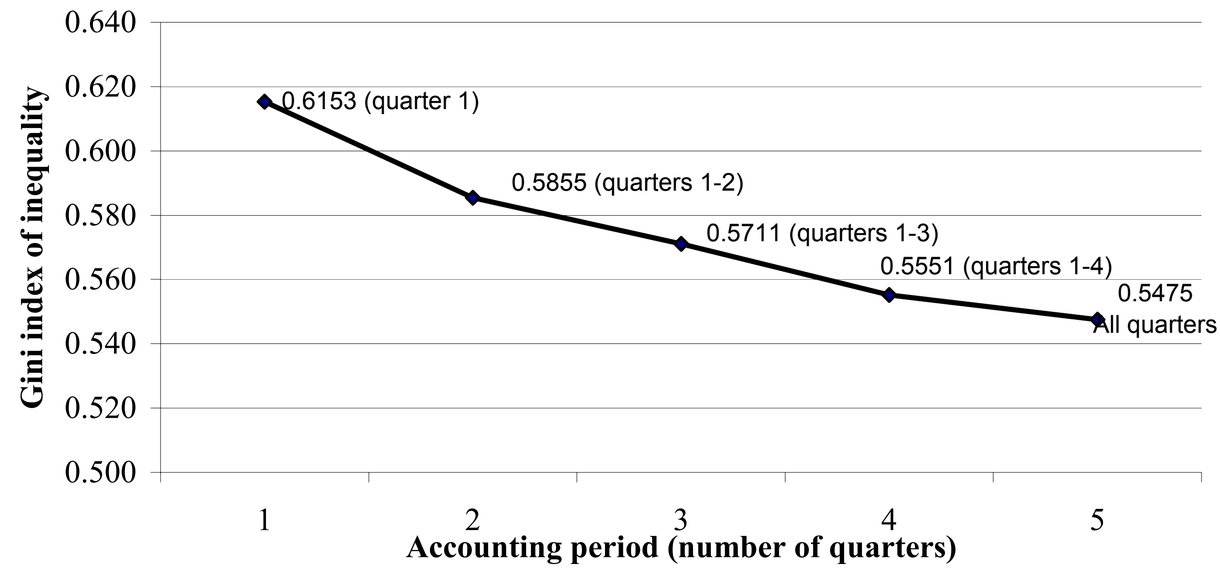
Schechtman, E. and S. Yitzhaki (1999) "On The Proper Bounds of The Gini Correlation" *Economics Letters* **63**: 133-138.

Table I: Gini index of inequality and accounting period, Mexico 1996

	Gini indices of inequality (G_i) and income shares (a_i)				
	1st quarter	2nd quarter	3rd quarter	4th quarter	5th quarter
Gini index	0.6153	0.6027	0.5967	0.5672	0.5789
Income share (five quarters)	0.2037	0.1930	0.1997	0.1937	0.2099
	Gini correlations matrix (Γ_{ij})				
	1st quarter	2nd quarter	3rd quarter	4th quarter	5th quarter
1st quarter	1.0000	0.7908	0.7669	0.7044	0.7200
2nd quarter	0.8267	1.0000	0.8480	0.7462	0.7700
3rd quarter	0.7703	0.8143	1.0000	0.7796	0.8064
4th quarter	0.7961	0.8024	0.8158	1.0000	0.8082
5th quarter	0.7833	0.7938	0.7851	0.7722	1.0000
	Gini correlations with aggregate income (Γ_{i0})				
	One quarter	Two quarters	Three quarters	Four quarters	Five quarters
1st quarter	1.0000	0.9662	0.9416	0.9322	0.9232
2nd quarter	-	0.9556	0.9454	0.9371	0.9309
3rd quarter	-	-	0.9448	0.9392	0.9318
4th quarter	-	-	-	0.9178	0.9136
5th quarter	-	-	-	-	0.9224
	Gini correlations with aggregate income((Γ_{0i})				
	1st quarter	2nd quarter	3rd quarter	4th quarter	5th quarter
One quarter	1.0000	-	-	-	
Two quarters	0.9537	0.9276	-	-	
Three quarters	0.9199	0.9175	0.9216	-	
Four quarters	0.9137	0.9134	0.9200	0.8616	-
Five quarters	0.9058	0.9079	0.9113	0.8615	0.8881
	G_0	G_0^2	$G_0\sum_{i=1}^{k} a_i D_{i0} G_i$	$\sum_{i=1}^{k} a_i^2 G_i^2$	$\sum_{i=1}\sum_{i\neq j} a_i a_j G_i G_j \Gamma_{ij}$
One quarter	0.6153	-	-	-	-
Two quarters	0.5855	0.3428	0.0071	0.1858	0.1499
Three quarters	0.5711	0.3261	0.0084	0.1221	0.1956
For quarters	0.5551	0.3082	0.0095	0.0889	0.2097
All quarters	0.5475	0.2998	0.0094	0.0703	0.2201

Source: Authors' estimation from ENEU data.

Figure 1: Accounting period and wage inequality in Mexico, 1996



Source: Authors' estimation from ENEU data.