

A Group-Specific Measure of Intergenerational Persistence

Tom Hertz
Department of Economics, American University
4400 Massachusetts Avenue NW, Washington DC 20016

hertz@american.edu
Tel: 202-885-2756
Fax:202-885-3790

July 2, 2007

Abstract

This paper outlines a decomposition technique that allows for inter-group comparisons of intergenerational persistence in economic status, for groups whose means are different. It takes account of both within- and between-group sources of the covariance of parents' and children's outcomes, and yields results that are quite different, and arguably more meaningful, than the simple comparison of within-group regression estimates.

Many thanks to Bhashkar Mazumder and Gary Solon for helpful comments on an earlier draft.

The usual disclaimers apply.

JEL: J62

Keywords: Intergenerational mobility; between-group decomposition.

Length, excluding title page and abstract, but including all else: 1957 words.

1. Introduction

Recent interest in the question of intergenerational economic mobility has generated a body of research that seeks to quantify the economic resemblance between parents and children (see, for example, Corak 2004; Bowles, Gintis, and Osborne 2005). These analyses often report the elasticity between the long-run incomes of parents and children, or, alternatively, the correlation in their log incomes, both of which measure what Fields (2000) terms “time dependence,” or persistence, of income across generations. These are interpreted as inverse measures of intergenerational economic mobility, where “mobility” refers to movements in either direction, relative to the generational means (rendering these statistics uninformative about changes in those means over time). High mobility (low persistence), in turn, is considered an indicator of greater equality of economic opportunity, although the one need not always imply the other (Jencks and Tach 2005).

These elasticities, however, often conceal considerable heterogeneity between groups, raising the question: can we determine which are the most and least mobile groups? Hertz (2005) discusses this problem and notes that stratifying and calculating within-group elasticities does not produce a meaningful answer. The reason is that any long-run difference in group means (the between-group component) is ignored under stratification, and this can lead to the false impression that a persistently disadvantaged group is actually highly mobile. Children of the disadvantaged group may be quite mobile with respect to their parents when measured by their within-group intergenerational elasticity. But if both parents and children in that group are generally confined to the lower end of the income distribution, then the children will typically remain relatively disadvantaged, which is to say, they will display little mobility.

The challenge is not only to acknowledge the between-group effect, but also to allocate it in a coherent way across groups, in order to develop group-specific statistics that measure both the degree to which parents and children have similar incomes within a group, as well as the degree to which people in a given group tend to fall above or below the sample mean. In this paper I propose a simple decomposition that does just that, and is applicable to either the intergenerational elasticity or the correlation coefficient. I first express the pooled elasticity (or correlation) as a weighted average of the within-group elasticities (or correlations), plus a term that is proportional to the weighted between-group estimator, i.e. the coefficient from a weighted regression of group means of the second generation outcomes against those of the first generation, with the group shares being the weights. My innovation, however, is to demonstrate that the between-group component can itself be allocated across groups, yielding separate estimates of persistence for each group that capture *both* the within-group component *and* that group's contribution to the between-group estimator. These group-specific statistics are a metric by which income persistence may meaningfully be compared across groups, even when those groups have very different average outcomes.

In the remainder of the paper I outline the technique, and provide examples of its application using the dataset found in Hertz (2005). Although I focus on the issue of the intergenerational resemblance, the technique is clearly applicable to other analyses that estimate pairwise similarities in an outcome, and compare these for members of different groups.

2. The Decomposition

Let groups be indexed by $g = 1 \dots G$, and let π_g represent group g 's share of the total number of parent-child pairs. Let x and y be the outcomes in the first and second generations, with means \bar{x} and \bar{y} and variances¹ $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$; \bar{x}_g represents group g 's mean and $\hat{\sigma}_{x(g)}^2$ its variance, for parents, with analogous definitions in y for the children. Let $\hat{\beta}_g$ represent the within-group regression coefficient (an elasticity if x and y are in logarithms); and $\hat{\delta}_G$ be the weighted between-group regression coefficient: $\hat{\delta}_G = \sum_g \pi_g (\bar{y}_g - \bar{y})(\bar{x}_g - \bar{x}) / \sum_g \pi_g (\bar{x}_g - \bar{x})^2$.

The pooled regression coefficient may then be written:

$$[1] \quad \hat{\beta} = \sum_g \pi_g \left(\hat{\beta}_g \frac{\hat{\sigma}_{x(g)}^2}{\hat{\sigma}_x^2} \right) + \hat{\delta}_G \frac{\sum_g \pi_g (\bar{x}_g - \bar{x})^2}{\hat{\sigma}_x^2}.$$

The final term can then be decomposed by group, and brought within the summation sign:

$$[2] \quad \hat{\beta} = \sum_g \pi_g \left(\hat{\beta}_g \frac{\hat{\sigma}_{x(g)}^2}{\hat{\sigma}_x^2} + \frac{(\bar{x}_g - \bar{x})(\bar{y}_g - \bar{y})}{\hat{\sigma}_x^2} \right).$$

For the correlation coefficient we have:

$$[3] \quad \hat{r} = \sum_g \pi_g \left(\hat{r}_g \frac{\hat{\sigma}_{x(g)} \hat{\sigma}_{y(g)}}{\hat{\sigma}_x \hat{\sigma}_y} \right) + \hat{\tau}_G \frac{\sqrt{\sum_g \pi_g (\bar{x}_g - \bar{x})^2 \sum_g \pi_g (\bar{y}_g - \bar{y})^2}}{\hat{\sigma}_x \hat{\sigma}_y}, \quad \text{and:}$$

$$[4] \quad \hat{r} = \sum_g \pi_g \left(\hat{r}_g \frac{\hat{\sigma}_{x(g)} \hat{\sigma}_{y(g)}}{\hat{\sigma}_x \hat{\sigma}_y} + \frac{(\bar{x}_g - \bar{x})(\bar{y}_g - \bar{y})}{\hat{\sigma}_x \hat{\sigma}_y} \right)$$

¹ The decomposition is greatly simplified by the use of the ‘‘population’’ variance estimator, which divides the sum of squared deviations from the mean by N , rather than $N-1$ (the usual degrees-of-freedom correction).

where \hat{r}_g is the within-group correlation coefficient, and $\hat{\tau}_G$ is the correlation-equivalent of $\hat{\delta}_G$,

$$\text{given by: } \hat{\tau}_G = \frac{\sum_g \pi_g (\bar{y}_g - \bar{y})(\bar{x}_g - \bar{x})}{\sqrt{\sum_g \pi_g (\bar{x}_g - \bar{x})^2 \sum_g \pi_g (\bar{y}_g - \bar{y})^2}} .$$

Equations [1] and [3] break the pooled regression (or correlation) coefficient into a weighted sum of the within-group coefficients (with weights reflecting group shares and variances) plus a term that includes the between-group estimator. In [2] and [4] each group is allocated its share of the between-group term, as desired. This piece is the product of differences between the group mean and the sample mean in each generation, which is just that group's (unweighted) contribution to the between-group covariance, $\sum_g \pi_g (\bar{y}_g - \bar{y})(\bar{x}_g - \bar{x})$.

This product will be near zero if the group mean approximates the sample mean in either generation. On the other hand, if a group's mean lies above (or below) the sample mean in both generations, this contribution will be positive; and if there are just two groups this term would be positive both for the lower-income *and* the higher-income group.² Last, for a group whose mean differs from that of the other group(s) in both generations, the smaller the group, the larger will be its distance from the pooled sample mean, and the larger will be the contribution of this term of the equation.

I will refer to the contents of the parentheses in [2] or [4], prior to weighting by π_g , as the group-specific measures of persistence; their weighted sum across groups then yields the standard pooled estimator.

² If the group mean switches from lying below the sample mean in one generation to above it in the next, or vice versa, then this term would be negative. The within-group term can also, in principle, be negative: this need not worry us in the intergenerational context, but might matter in other applications.

3. Examples

Two applications of this decomposition are illustrated, both relating to parents and children found in the dataset of Hertz (2005), which contains estimates of long-run age-adjusted family income in two generations, for a sample of 6,273 families from the Panel Study of Income Dynamics. The first two columns divide the sample according to race (only Black and White families appear in the sample), while the remaining columns divide the sample into the four major census regions. Sample weights are applied for all calculations, which causes no problems so long as all terms in equations [1] through [4] are also calculated using those weights. Table 1 presents the results for equations [1] and [2]; results for the correlation coefficients were qualitatively similar and are omitted.

The first panel shows that Blacks made up 18 per cent of the sample's weighted number of parent-child pairs (first row) and that their mean family income was seventy to eighty log points below that of Whites in both generations (second and third rows). The second panel reports the intergenerational regression coefficients for the full sample (0.534), and by group (0.317 and 0.393). In the row labeled "A", these figures are scaled by the variance terms from equation [1].

The large gap between the pooled and within-group estimates implies that between-group effects are important. In the second panel, we see that the value of $\hat{\delta}_G$ is greater than one, confirming that the racial (log) income gap grew over time.³ In the next row this figure is scaled by the appropriate variances, revealing that the between-group effect accounts for two-fifths of the overall intergenerational coefficient (0.231 / 0.534).

³ In the two-group (b, w) case, the between-group estimator simplifies to $(\bar{y}_w - \bar{y}_b) / (\bar{x}_w - \bar{x}_b)$.

In the row labeled “B”, the between-group effect is split into its Black and White components, with the effect for Blacks being much the larger (1.046 versus 0.051). This occurs because their (weighted) share of the sample is small, so their group mean lies farther from the sample mean than does the mean for Whites. Next, these between-group components are weighted by their group shares, and their sum is indeed 0.231. This confirms that equations [1] and [2] are indeed equivalent.

[TABLE 1 ABOUT HERE]

In the final panel I report the statistic that I have defined as the group-specific measure of persistence, which is the sum of the entries in rows A and B. This stands at 1.367 for Blacks and 0.349 for Whites, implying that once within- and between-group effects are both accounted for, generational income persistence for Blacks is roughly four times as high as for Whites. Another way to express this fact is to look at the group-share weighted estimates in the final row (which sum to the pooled intergenerational elasticity), and to note that while Blacks make up just 18 per cent of the (weighted) sample, they generate nearly half (0.248 / 0.534) of the parent-child covariance that underlies the intergenerational elasticity. This is not due to a larger within-group effect, but to the significant gap in mean incomes across groups.

The regional decomposition involves four groups of more nearly equal size, and smaller differences in means. Still, incomes were higher in the Northeast, and lower in the South, so the between-group effect is not zero: it contributes 0.064 to the total intergenerational elasticity. As expected, when this between-group component is allocated across regions (row B), it has a noticeable effect on estimated persistence for the Northeast and South (0.073 and 0.140), but little effect for the other regions, which lie closer to the mean. Overall persistence is lowest in the West and highest in the South. Had we made this comparison on the basis of the within-group estimates alone, we would have instead concluded that both the Northeast and the Midwest had higher persistence than the South. Note also that, like Blacks, Westerners are a small minority; yet the West does not account for much of the between-group effect, as its mean is near the sample mean.

This technique highlights the importance of between-group effects when group means differ, and shows how to handle them when making inter-group comparisons of intergenerational income mobility. It could be applied to many such situations: recent analyses whose findings might (or might not) be altered by the adoption of this method include Björklund and Chadwick (2003), Björklund, Lindahl, and Plug (2006), and Sacerdote (2007) all of which compare within-group intergenerational coefficients for children in different family types (e.g. living with biological parents or not).

Table 1

Decompositions of Intergenerational Persistence Measures By Race and Region

	Pooled	Black	White	North-East	Mid-West	South	West
Shares: π_g	1.000	0.181	0.819	0.243	0.306	0.299	0.151
Means: \bar{x}_g	10.66	10.09	10.79	10.82	10.75	10.40	10.75
Means: \bar{y}_g	10.68	10.00	10.83	10.84	10.71	10.47	10.73
Within-Group Regressions: $\hat{\beta}, \hat{\beta}_g$	0.534	0.317	0.393	0.585	0.556	0.483	0.381
A $\hat{\beta}_g \hat{\sigma}_{x(g)}^2 / \hat{\sigma}_x^2$		0.321	0.298	0.459	0.424	0.626	0.270
Between-Group Regression: $\hat{\delta}_G$		1.180			0.794		
$\hat{\delta}_G \text{Var}(\bar{x}_g) / \hat{\sigma}_x^2$		0.231			0.064		
B $(\bar{x}_g - \bar{x})(\bar{y}_g - \bar{y}) / \hat{\sigma}_x^2$		1.046	0.051	0.073	0.008	0.140	0.012
$\pi_g (\bar{x}_g - \bar{x})(\bar{y}_g - \bar{y}) / \hat{\sigma}_x^2$		0.189	0.042	0.018	0.002	0.042	0.002
		$\Sigma=0.231$			$\Sigma=0.064$		
Group Persistence: A+B		1.367	0.349	0.532	0.432	0.766	0.282
$\pi_g (A+B)$	$\Sigma=0.534$	0.248	0.286	0.129	0.133	0.229	0.043

References

- Björklund, Anders and Laura Chadwick. 2003. "Intergenerational Income Mobility in Permanent and Separated Families." *Economics Letters*, 80:239–246.
- Björklund, Anders, Mikael Lindahl, and Erik Plug. 2006. "The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data." *The Quarterly Journal of Economics*, 121(3):999-1028.
- Bowles, Samuel, Herbert Gintis, and Melissa Osborne, eds. 2005. *Unequal Chances: Family Background and Economic Success*. New York: Russell Sage and Princeton University Press.
- Corak, Miles, ed. 2004. *Generational Income Mobility in North America and Europe*. Cambridge: Cambridge University Press.
- Fields, Gary S. 2000. "Income Mobility: Concepts and Measures," in *New Markets, New Opportunities? Economic and Social Mobility in a Changing World*, edited by Nancy Birdsall and Carol Graham, pp. 101-132. Washington: Carnegie Endowment for International Peace and Brookings Institution Press.

Hertz, Tom. 2005. "Rags, Riches and Race: The Intergenerational Economic Mobility of Black and White Families in the United States," in *Unequal Chances: Family Background and Economic Success*, edited by Samuel Bowles, Herbert Gintis and Melissa Osborne, pp. 165-191. New York: Russell Sage and Princeton University Press.

Jencks, Christopher and Laura Tach. 2005. "Would Equal Opportunity Mean More Mobility?" John F. Kennedy School of Government, Faculty Research Working Paper Series RWP05-037, May 2005.

Sacerdote, Bruce. 2007. "How Large Are the Effects from Changes in Family Environment? A Study of Korean American Adoptees." *The Quarterly Journal of Economics*, 122(1):119-157.