On the Interpretation of the Intergenerational Elasticity and the Rank-Rank Coefficients for Cross Country Comparison

Nadav Kunievsky

January 27, 2024

Abstract

This paper investigates Intergenerational Elasticity (IGE) and Rank-Rank coefficients, employing Yitzhaki's theorem to express them as weighted averages of underlying causal mechanisms driving mobility. We highlight the challenges of interpreting cross-country comparisons using IGE or Rank-Rank coefficients due to the regression weighting scheme. We also show that, while the Rank-Rank coefficient is more interpretable for positional mobility, it lacks insights into the underlying mechanisms driving mobility across countries. The analysis demonstrates potential drawbacks of using linear regression coefficients as summary statistics in the context of intergenerational mobility comparisons.

Introduction

Numerous studies have examined the relationship between parental income and child income. Two prominent methods for summarizing the joint distribution of these incomes are the Intergenerational Elasticity (IGE) coefficient and the Rank-Rank coefficient (Mogstad and Torsvik (2023)). This paper explores how these measures summarize the joint income distribution and their subsequent connections to the underlying mechanisms that link parental and child income.

Let I_c and I_p denote child and parent income, respectively. The IGE coefficient is the slope coefficient obtained by regressing the logarithm of child income on the logarithm of parent income as follows:

$$\log I_c = \alpha_{IGE} + \beta_{IGE} \log I_p + \epsilon. \tag{1}$$

This regression coefficient captures the persistence between child log income and the parent log income, with higher values indicating stronger persistence.[1] A popular alternative to this method is the Rank-Rank regression, which assesses the correlation between parent and child ranks within their respective income distributions. Assuming a continuous income distribution for both parents and children, let $R_c = F_c(I_c)$ and $R_p = F_p(I_p)$ represent the parent and child ranks in their respective income distributions, where $F_c(x)$ and $F_p(x)$ are the cumulative distribution functions of child and parental income, respectively. Researchers then measure the Rank-Rank relationship by estimating the following regression:

$$R_c = \alpha_r + \beta_r R_p + \varepsilon. \tag{2}$$

The regression slope coefficient quantifies how the child position in the income distribution relates to their parent position in the corresponding income distribution.

The IGE coefficient has been extensively employed in empirical studies to describe intergenerational persistence, dating back to the 1980s (Becker and Tomes (1986), Atkinson (1980)). However, the Rank-Rank coefficient has gained popularity more recently, after Chetty et al. (2014b) applied it to measure social mobility over time in the United States. While both coefficients are used to describe intergenerational mobility, each conveys distinct information about the joint distribution of parental and child income. As demonstrated below, the IGE provides a weighted average of the expected change in child logarithmic income in relation to a change in parent logarithmic income.[2] Consequently, the IGE coefficient is influenced by both the marginal distributions and the dependency structure between parental and child income. In contrast, the Rank-Rank coefficient measures positional mobility across generations, only summarizing the copula while isolating the dependency structure between the incomes and disregarding changes in marginal distributions (Deutscher and Mazumder (2023), Mogstad and Torsvik (2023), Aloni and Krill (2017)). From a practical perspective, the Rank-Rank coefficient has shown to more robust to sample restrictions (Chetty et al. (2014a), Chetty et al. (2014b),

[1] In many cases, the level of intergenerational mobility is reported using (1-β_{IGE})

[2] Mitnik and Grusky (2020) illustrates that the IGE can be considered as the elasticity of the conditional geometric mean, i.e., the expected percentage change in the geometric mean of the child's income with respect to the percentage change in the parental income.

Dahl and DeLeire (2008)). In some countries (although not all; Bratberg et al. (2017), Acciari et al. (2022)), the Rank-Rank relation between parental and child income is almost perfectly linear. On the other hand, the conditional expectation function, $E[\log I_c | \log I_p]$, demonstrates significant nonlinearity (Chetty et al. (2014a), Deutscher and Mazumder (2023)). Moreover, the Rank-Rank coefficient allows researchers to include individuals with no income. This could be important since, as observed by Chetty et al. (2014a), the IGE demonstrates significant sensitivity to the substitution of zeros with ones or $1,000$s.

This paper examines the challenges that are inherent to using IGE and Rank-Rank coefficients for cross-country mobility comparisons. We express these coefficients as weighted averages of causal factors affecting intergenerational mobility using Yitzhaki's theorem (Yitzhaki (1996)), demonstrating that these coefficients assign varying weights across the parental income distribution. This helps to explain certain properties that were shown in the existing literature. We further explore how the parental income distribution influences the IGE and Rank-Rank coefficients, complicating cross-country comparisons, particularly when mobility occurs in different segments of the parental income distribution in each country. A related study (Maasoumi et al. (2022)) also employs Yitzhaki's theorem, framing the IGE coefficient weighting scheme as a special case within a broader class of intergenerational mobility measures that captures different preference relations over income distributions. The authors show that the IGE coefficient corresponds to a specific case of a preference relations that places higher weight on the mobility of wealthier households. In contrast, our study focuses on interpreting the coefficients as a weighted average of the underlying causal mechanisms and examines how these interpretations are important for cross-country comparison.

Decomposing the IGE coefficient

We begin by examining the β_{IGE} coefficient. Let us assume that (I_c, I_p) are i.i.d, $E[|\log I_c|], E[|\log I_p|] < \infty$, and $E[\log I_c | \log I_p = t]$ exists and is differentiable for all t. According to Yitzhaki's theorem, we can express β_{IGE} as a weighted average of the derivative of the conditional expectations:

$$\beta_{IGE} = \frac{\mathrm{Cov}(\log I_c, \log I_p)}{\mathrm{Var}(\log I_p)} = \int_{-\infty}^{\infty} \frac{\partial E[\log I_c | \log I_p = t]}{\partial t} w(t) dt,$$

where

$$w(t) = \frac{E\left[\log I_p - \mu_{I_p} | \log I_p > t\right] \mathrm{P}(\log I_p > t)}{\mathrm{Var}(\log I_p)}, \quad \int_{-\infty}^{\infty} w(t) dt = 1, \quad \mu_{I_p} = E[\log I_p].$$

We can then interpret the IGE coefficient as a summary statistic of the underlying function $E[\log I_c | \log I_p]$, where the weights depend on the distribution of parental income. Specifically, these weights are maximized at $E[\log I_p]$ and approach zero at the boundary of the support (Yitzhaki (1996), Heckman et al. (2006)). Thus, β_{IGE} assigns higher weight to households with the mean parental log income[3] and lower weights to households at the top and bottom of the parental log income distribution.

The fact that the IGE coefficient assigns lower weights to households at the extremes may be concerning in cases in which a significant portion of mobility occurs for children from very poor or very rich families. This can potentially occur as a result of policies aimed at reducing poverty or simply through regression to the mean. The fact that the weights depend on the underlying parental log income distribution implies that comparisons of the IGE coefficients that are cross-country or over time can be difficult to interpret. For instance, without knowing the exact weights, differences between two countries may simply arise from differences in the weighting schemes used by the IGE coefficient, even if the conditional expectation function $E[\log I_c | \log I_p]$ is the same across both countries.

Notably, the fact that the IGE coefficient assigns higher weights to mobility around the mean may explain why the IGE is considered sensitive to sample definitions and restrictions. Some sample restrictions, such as excluding households with zero income or those with very high income, can significantly impact the mean of the distribution. As a result, households that receive higher weights change and the IGE coefficient also changes.

To better understand how the IGE coefficient relates to the Rank-Rank coefficient and underlying income elasticity,[4] we aim to decompose the integrand into the expected parent-child income elasticity and additional correlative mechanisms. Let the causal model governing child income be given as follows:

$$I_c = h(I_p, u), \tag{3}$$

where u represents other unobserved factors that affect child income. Let $\epsilon_{I_c,I_p}(u) = \frac{\partial \log I_c}{\partial \log I_p}$ be the elasticity of child income with respect to parent income for given unobserved factors, u, evaluated at I_p. Let $I_p(t) = exp(t)$ denote the

[3]Note that this is generally not the same as families with mean income.

[4]As noted by Mitnik and Grusky (2020), the IGE coefficient does not actually provide information about the parent-child income elasticity, which is evident in our setup, as $\frac{\partial E[\log I_c | \log I_p]}{\partial \log I_p} \neq E\left[\frac{\partial \log I_c}{\partial \log I_p} | \log I_p\right]$

inverse of $\log I_p$, where $\log I_p = t$. We can then rewrite the integrand as follows:

$$\begin{aligned}\frac{\partial E[\log I_c | \log I_p = t]}{\partial t} &= \int_{-\infty}^{\infty} \frac{\partial \log h\left(I_p\left(t\right), u\right) P(u | \log I_p = t)}{\partial t} du \\ &= \underbrace{E\left[\epsilon_{I_c, I_p(t)}(u) \middle| \log I_p = t\right]}_{\text{Causal IGE}} + \underbrace{\int_{-\infty}^{\infty} \log h\left(I_p\left(t\right), u\right) \frac{\partial P(u | \log I_p = t)}{\partial t} du}_{\text{Other Factors}},\end{aligned} \tag{4}$$

where the second equality follows from the product rule. The first component captures the conditional expected causal IGE, while the second component captures how changes in income are associated with changes in other factors that affect income.[5] Therefore, β_{IGE} can be expressed as a summation of the weighted causal intergenerational elasticities (causal IGE) and an additional term that captures how parental income is correlated with other factors that affect child income. In most studies of intergenerational mobility, both terms are crucial as researchers are interested in measuring how parent income is associated with child income, through either the causal effect of parental income or the association between parental income and other factors such as neighborhood quality, quality of schools, inherited human capital, and peer effects.

Decomposing the Rank-Rank coefficient

We now turn our attention to the Rank-Rank coefficient. Using Yitzhaki's theorem once more, we have the following:

$$\beta_r = \frac{\text{Cov}(R_c, R_p)}{\text{Var}(R_p)} = \int_{t=0}^{1} \frac{\partial E[R_c | R_p = t]}{\partial t} w(t) dt,$$

where, using the fact that the rank distribution is uniform, the exact weighting scheme is as follows:

$$w(t) = \frac{12(1-t)t}{2}, \quad \int_0^1 w(t) dt = 1.$$

Comparing the weights of the IGE coefficient to the Rank–Rank coefficient, the Rank–Rank weights place most of the weight on households at the median of the parental income distribution. In contrast, the IGE assigns most of the weight to households closer to the mean of the distribution. In addition, weights decline symmetrically as we

[5] This decomposition of the β_{IGE} can be thought of as an omitted variable bias. In this case, bias is taken with respect to the Ordinary Least Squares weighted causal effects of log parental income, as implied in Yitzhaki's theorem.

move further away from the median and toward the extremes. Thus, similar to the β_{IGE} coefficient, the Rank–Rank coefficient assigns lower weights to households closer to the top and bottom of the parental income distribution. Notably, since the median is usually less sensitive to changes in sample restrictions at the top and bottom of the distributions, this weighting scheme might explain why the Rank–Rank coefficient has been documented to be more robust for different sample restrictions (Dahl and DeLeire (2008), Chetty et al. (2014b)). Finally, compared with the IGE coefficient, the weights for cross-country comparisons are more consistent, assigning similar weights to households at the same rank of the income distribution. Note that, if the marginal distributions differ across countries, this implies that the Rank–Rank weighting scheme assigns different weights to households with the same income levels. Whether this is desirable depends on the researcher's questions and objectives.

As we did for the IGE coefficient, we can express the Rank-Rank coefficients in terms of the underlying parent-child income elasticities. Let ϵ_c and ϵ_p be the elasticities of rank with respect to income for the child and parents, respectively. Let $R_c\epsilon_c = R_c\frac{\partial R_c}{\partial I_c}\frac{I_c}{R_c}$ and $R_p\epsilon_p = R_p\frac{\partial R_p}{\partial I_p}\frac{I_p}{R_p}$ represent the semi-elasticities of rank with respect to income. These quantities measure how the rankings of parents and child change in response to a percentage variation in their respective incomes. We can then rewrite, with a slight abuse of notation, the integrand as follows:[6]

$$\begin{aligned}
\frac{\partial \mathrm{E}[R_c|R_p=t]}{\partial t} &= \frac{\partial \mathrm{E}[F_c\left(h\left(F_p^{-1}(t),u\right)\right)|R_p=t]}{\partial t} \\
&= \int_{-\infty}^{\infty}\frac{\partial F_c\left(h\left(F_p^{-1}(t),u\right)\right)P(u|R_p=t)}{\partial t}du \\
&= \mathrm{E}\left[\frac{\partial R_c}{\partial h}\frac{\partial h}{\partial I_p}\frac{1}{\frac{\partial R_p}{\partial I_p}}\bigg|R_p=t\right] + \int_{\infty}^{\infty}F_c(h(F_p^{-1}(t),u)\frac{\partial P(u|R_p=t)}{\partial t}du \\
&= \mathrm{E}\left[\frac{\partial I_c}{\partial I_p}\frac{\frac{\partial R_c}{\partial I_c}}{\frac{\partial R_p}{\partial I_p}}\frac{I_c}{I_c}\frac{I_p}{I_p}\frac{R_c}{R_c}\frac{R_p}{R_p}\bigg|R_p=t\right] + \int_{\infty}^{\infty}F_c(h(F_p^{-1}(t),u)\frac{\partial P(u|R_p=t)}{\partial t}du \\
&= \underbrace{E\left[\frac{R_c}{R_p}\frac{\epsilon_c}{\epsilon_p}\epsilon_{I_c,I_p}(u)\bigg|R_p=t\right]}_{\text{Re-Scaled Causal IGE}} + \underbrace{\int_{\infty}^{\infty}F_c(h\left(F_p^{-1}(t),u\right))\frac{\partial P(u|R_p=t)}{\partial t}du}_{\text{Other factors}},
\end{aligned} \tag{5}$$

where the third equality is due to the product rule and the chain rule. The fourth equality results from dividing and multiplying by parents and child income and ranks and the definition of the parents and child ranks.[7] The final

[6] Maasoumi et al. (2022) expresses the Rank-Rank coefficient as a weighted average of $\frac{\partial \mathrm{E}[\log I_c|\log I_p=t]}{\partial t}$, with weights that are generally positive but do not necessarily sum to 1. In contrast, we express the Rank-Rank coefficient as a weighted average of $\frac{\partial \mathrm{E}[R_c|R_p=t]}{\partial t}$ with weights that sum to 1.

[7] For the sake of clarity, we slightly abuse notation and denote $I_c = h\left(F_p^{-1}(t),u\right)$, $R_c = F_c\left(h\left(F_p^{-1}(t),u\right)\right)$, $I_p = F_p^{-1}(t)$, and $R_p = t$.

equality follows from the definition of semi-elasticities. Expressing the integrand in this way reveals the similarities and differences between the IGE coefficient and the Rank-Rank coefficient. First, as child income cumulative distribution function is monotonic, similar to the log function, the effects of other factors on income have remained the same, except that we use child marginal income distribution to transform the income instead of log. Likewise, the Rank-Rank coefficient is also affected by the causal effects of the IGE, but now the IGE is multiplied by a "translation" term that converts the income elasticities to rank elasticities.

If we are using the Rank–Rank coefficient for cross-country comparisons, the decomposition we derived above explicitly demonstrates that the Rank–Rank coefficient is only useful for comparisons of positional mobility. However, It cannot speak to how similar or different the mechanisms driving this mobility are across countries.[8] For example, consider two countries with the same underlying causal mechanisms $h(I_p, u)$ and assume that $I_p \perp\!\!\!\perp u$, which implies that the second term is zero. If the parental income distributions differ across the two countries, the Rank-Rank coefficient would still be different for two reasons. The first reason is that, although the weighting scheme is the same for households with the same income rank, the regression weighting scheme weights households with the same income level differently. The second and more substantial reason is that the way that the causal mechanisms affect rank would differ between the two countries as the semi-elasticities are different in the causal IGE term in equation 5. Therefore, although we might motivate the use of the Rank-Rank coefficient as a means to abstract away from the marginals, we cannot avoid considering the marginals if we want to use the Rank-Rank coefficient to think about differences in the driving mechanisms of mobility between two countries.

Discussion

This paper employs Yitzhaki's theorem to express IGE and Rank-Rank coefficients as weighted averages of the causal mechanisms driving income and positional mobility. We demonstrate that interpreting cross-country comparisons using the IGE coefficient can be challenging due to the regression weighting scheme. Additionally, we establish that the Rank-Rank coefficient is readily interpretable only when researchers focus on positional mobility, without providing insights into the similarities or differences in the underlying mechanisms driving mobility across countries.

We highlight the potential drawbacks of using linear regression coefficients as summary statistics. Linear regression

[8] In theory, the Rank–Rank coefficient can be more informative on causal mechanisms that operate directly from parent income rank to child income rank, bypassing income levels. We leave this observation for future research.

may be preferred in certain cases for its efficiency and stability, even with a small number of observations. However, it seems that in the context of intergenerational mobility comparisons, this is not always warranted. Recent research has shifted to using large administrative datasets that can provide precise estimates of the relation between parent and child income. Consequently, the practice of reporting regression coefficients over estimates from more flexible and transparent methods may not always be well justified.

References

Acciari, Paolo, Alberto Polo, and Giovanni L. Violante. And Yet it Moves: Intergenerational Mobility in Italy. *American Economic Journal: Applied Economics*, 14(3):118–163, 2022.

Aloni, Tslil, and Zeev Krill. Intergenerational Income Mobility in Israel-International and Sectoral Comparisons. *Ministry of Economics, Office of Chief Economist*, 2017.

Atkinson, Anthony B. On Intergenerational Income Mobility in Britain. *Journal of Post Keynesian Economics*, 3(2):194–218, 1980.

Becker, Gary S., and Nigel Tomes. Human Capital and the Rise and Fall of Families. *Journal of Labor Economics*, 4(3, Part 2):S1–S39, 1986.

Bratberg, Espen, Jonathan Davis, Bhashkar Mazumder, Martin Nybom, Daniel D. Schnitzlein, and Kjell Vaage. A Comparison of Intergenerational Mobility Curves in Germany, Norway, Sweden, and the US. *The Scandinavian Journal of Economics*, 119(1):72–101, 2017.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez. Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *The Quarterly Journal of Economics*, 129(4):1553–1623, 2014.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, Emmanuel Saez, and Nicholas Turner. Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility. *American Economic Review*, 104(5):141–147, 2014.

Dahl, Molly W., and Thomas DeLeire. The Association between Children's Earnings and Fathers' Lifetime Earnings: Estimates Using Administrative Data. *University of Wisconsin-Madison, Institute for Research on Poverty*, Madison, WI, USA, 2008.

Deutscher, Nathan, and Bhashkar Mazumder. Measuring intergenerational income mobility: A synthesis of approaches. *Journal of Economic Literature*, 61(3):988–1036, 2023.

Heckman, James J., Sergio Urzua, and Edward Vytlacil. Understanding Instrumental Variables in Models with Essential Heterogeneity. *The Review of Economics and Statistics*, 88(3):389–432, 2006.

Maasoumi, Esfandiar, Le Wang, and Daiqiang Zhang. Generalized Intergenerational Mobility Regressions. *Working paper, Emory University*, 2022.

Mitnik, Pablo A., and David B. Grusky. The Intergenerational Elasticity of What? The Case for Redefining the Workhorse Measure of Economic Mobility. *Sociological Methodology*, 50(1):47–95, 2020.

Mogstad, Magne, and Gaute Torsvik. Family Background, Neighborhoods, and Intergenerational Mobility. In Shelly Lundberg and Alessandra Voena (Eds.), *Handbook of the Economics of the Family*, Volume 1, Issue 1, pp. 327–387. Elsevier Science & Technology, 2023.

Yitzhaki, Shlomo. On Using Linear Regressions in Welfare Economics. *Journal of Business & Economic Statistics*, 14(4):478–486, 1996.