Multigenerational mobility refers to the association of socioeconomic status across three or more generations.

Gary Becker & Nigel Tomes, *Journal of Labor Economics* (1986):

"… practically all the advantages or disadvantages of ancestors tend to disappear in only three generations: 'from shirtsleeves to shirtsleeves in three generations.' Parents in such 'open' societies have little effect on the earnings of grandchildren and later descendants."

Gregory Clark, *The Son Also Rises* (2014):

"Surname status shows regression to the mean in all cases, but the process is slow. Elite surnames can take ten or fifteen generations (300-450 years) to become average in status."

What Do We Know So Far about Multigenerational Mobility?

I. Review of the Empirical Literature

II. Some Theoretical Interpretations

III. Clark's Interpretation

I. The Empirical Literature

Many of us have measured intergenerational mobility by estimating first-order autoregressions, e.g., regressions of offspring's log income on parental log income. Is the mobility process really AR (1)?

Note on "misspecification":
For the purpose for which we usually have run these regressions (to characterize the association in socioeconomic status between adjacent generations), these regressions are not misspecified. They measure what we mean them to.

BUT what if instead we want to extrapolate from the two-generation results how the associations die out over additional generations?

E.g., occasionally writers have assumed both stationarity and an AR(1) data-generating process by assuming that intergenerational autocorrelations die out at a geometric rate – say, extrapolating a first-order autocorrelation of 0.4 to imply a second-order autocorrelation of 0.16, a third-order of 0.064, etc.

As we will see later, there is no theoretical presumption in favor of the AR(1) specification. But first, what does the empirical evidence show?

Donald Hodge (University of Chicago!), “Occupational Mobility as a Probability Process,” *Demography* (1966), studied transitions among occupation categories across three generations in the United States.

“Although the discrepancies between the actual and expected values shown in Table 1 clearly indicate that grandfather’s occupation bears some relation to grandson’s occupation which is not fully explained by father’s occupation, we must emphasize that the discrepancies are not large…. Grandfather’s occupation does not have any appreciable direct effect upon a person’s occupation beyond the indirect effect induced by its influence upon father’s occupation.”

Many subsequent studies have estimated multigenerational regressions for years of education or income variables.

For example, Behrman and Taubman, *REStat* (1985), used their NAS-NRC Twins data from the United States to estimate regressions of offspring's years of education on parental and grandparental years of education. The estimated coefficients of grandparental schooling were close to zero and statistically insignificant.

Similar results have been reported in many other studies, e.g., Peters (1992), Warren and Hauser (1997), and Lucas and Kerr (2013).

On the other hand, some other studies estimating multigenerational regressions have gotten non-trivially positive coefficient estimates for grandparental status.

A prominent recent example is the forthcoming *JHR* paper by Mikael Lindahl et al. They use three-generation data from Malmo, Sweden, to estimate regressions of offspring's year of education on parental and grandparental education. The parental coefficient estimate is 0.26 (SE=0.02), and the grandparental one is 0.06 (0.02). Unfortunately, they don't report the analogous regression for log earnings, but they do report that the regression of offspring log earnings on parental log earnings yields a coefficient estimate of 0.30 (0.04), while the regression of offspring log earnings on grandparental log earnings yields an estimate of 0.18 (0.04). That the latter is more than the square of the former suggests that the estimated grandparental coefficient in a three-generation regression would be positive.

Kelly Vosters, a fourth-year Ph.D. student at Michigan State, is presently estimating similar regressions with U.S. PSID data on earnings and income, and her preliminary results are fairly similar to those of Lindahl et al.

So some studies have not found much evidence of a grandparental "effect," and some others have.

An interesting study with a combination of the two is Zeng and Xie, "The Effects of Grandparents on Children's Schooling: Evidence from Rural China," *Demography* (2014). In their regressions of offspring education on parental and grandparental education, they interact grandparental education with whether the grandparents were coresident.

"Although the education of noncoresident and deceased grandparents has little or no effect on grandchildren's dropout rate, the effect of coresident grandparents' education is quite large…. These results suggest that grandparents can play an important role in their grandchildren's schooling if they all live under the same roof."

Two broader lessons:

- We should not be surprised if grandparental coefficients differ across circumstances.
- Such variation might prove to be useful for pointing to likely causal pathways.

II. Theory

My point of departure is the initial model in my 2014 paper in *Research in Social Stratification and Mobility*, which adapted the classic model of Becker and Tomes (1979) to rationalize the double-log functional form of the regression equations typically estimated in empirical studies of intergenerational income mobility.

The assumptions, spelled out fully in the paper, include these:

- A single parent divides her income between her own consumption and investment in a single child's human capital so as to maximize a Cobb-Douglas utility function in which the two goods are the parent's consumption and the child's adult income.
- The specifications of the human capital production function and the earnings function are such that the elasticity of the child's adult income with respect to parental investment in the child's human capital is a constant γ.
- The human capital production function includes an *e* term that denotes the human capital endowment the child receives regardless of the family's conscious investment choices. This endowment is intergenerationally correlated because of both inheritance of genetic traits and cultural inheritance, such as the effects of parental role-modeling. My initial model follows Becker and Tomes in assuming that inheritance of endowment follows the AR(1) process

  $$e_{it} = \delta + \lambda e_{i,t-1} + v_{it}$$

  where the heritability coefficient λ lies between 0 and 1.

A bunch of math shows that these assumptions lead to this steady-state intergenerational income elasticity:

$$\beta = (\gamma + \lambda)/(1 + \gamma\lambda)$$

This equation shows that the intergenerational income elasticity is positive for both of two reasons – because γ is positive (i.e., richer parents' greater investment in their children's human capital makes their children richer) and because λ is positive (i.e., richer parents tend to have more favorable endowments, which tend to be passed on to their children through genetic and cultural inheritance).

Suppose, for example, that $\gamma=0.3$ and $\lambda=0.2$ (or vice versa). Then the intergenerational income elasticity is $\beta=(0.3+0.2)/[1+(0.3)(0.2)]$, which is about 0.47.

What does this model imply about the role of grandparents? Still more math yields the second-order autoregression

$$\log y_{it} = \text{intercept} + (\gamma + \lambda)\log y_{i,t-1} - \gamma\lambda \log y_{i,t-2}$$
$$+ \text{ white-noise error term}$$

In this regression of the child's log income on both parental and grandparental log income, the coefficient of parental log income is positive, but the coefficient of grandparental log income is a small *negative* quantity! For example, with γ=0.3 and λ=0.2, the coefficient of parental log income is 0.50, and the coefficient of grandparental log income is -0.06.

This implication of a negative coefficient for grandparental income, first noted by Becker and Tomes (1979), is initially surprising, but it does not really mean that an exogenous increase in grandparental income harms the child's income. Rather, it reflects a subtle implication of higher grandparental income *conditional on the amount of parental income*. If the parent did not earn more despite the advantages of higher grandparental income, this signals that the parent got a poor draw on her or his genetic/cultural endowment, and that poor draw tends to be passed on to some extent to the child.

Note that, if the multigenerational mobility process is really AR(2) with a negative coefficient for grandparental status, then intergenerational autocorrelations decline *more* rapidly than geometrically. For example, with $\gamma=0.3$, $\lambda=0.2$, and hence about a 0.47 correlation between parent and child log incomes, the correlation between the grandparent's and child's log incomes is about 0.18, somewhat less than the square of 0.47. And the correlation between the great-grandparent's and child's log incomes is only about 0.06.

This explains why Becker and Tomes made their "shirtsleeves to shirtsleeves in three generations" remark despite referring only to two-generation evidence. They thought (a) that the intergenerational correlation for adjacent generations is no more than 0.2 AND (b) that intergenerational correlations die out *more* rapidly than geometrically.

We now see the flaws in both premises. With regard to (b), there is hardly any evidence that the grandparental coefficient really is negative. Rather, there is growing evidence to suggest that, in some times and places, the grandparental coefficient is positive, and hence that multigenerational correlations die out *less* rapidly than geometrically.

This doesn’t mean that the basic Becker-Tomes analysis is wrong as far as it goes. Rather, it doesn’t go far enough. It is incomplete, leaving out additional ways in which grandparents’ status may foretell children’s outcomes.

In other words, the theory needs to be extended. And some of the most plausible extensions actually are foreshadowed in Becker and Tomes’s writings.

Extending the Theory

Example #1: As Zeng and Xie suggested, grandparents sometimes contribute to cultural inheritance.

Accordingly, in chapter 6 of *A Treatise on the Family* (1981), Becker explicitly entertained extending the endowment transmission model from the AR(1) specification

$$e_{it} = \delta + \lambda e_{i,t-1} + v_{it}$$

to more complex specifications such as this AR(2):

$$e_{it} = \delta + \lambda_1 e_{i,t-1} + \lambda_2 e_{i,t-2} + v_{it}$$

where $0 \leq \lambda_2 < \lambda_1 < 1$.

Then redoing the math generalizes the multigenerational mobility equation from our previous AR(2) model

$$\log y_{it} = \text{intercept} + (\gamma + \lambda) \log y_{i,t-1} - \gamma\lambda \log y_{i,t-2} + \text{white-noise error term}$$

to this AR(3) model:

$$\log y_{it} = \text{intercept} + (\gamma + \lambda_1) \log y_{i,t-1} + (\lambda_2 - \gamma\lambda_1) \log y_{i,t-2} - \gamma\lambda_2 \log y_{i,t-3} + \text{white-noise error term}$$

Example #2: Group effects [also anticipated in section VI of Becker and Tomes (1979)]

Various types of "ethnic capital" might cause the intercept in a multigenerational mobility equation such as

$$\begin{aligned}\log y_{it} = \text{intercept} &+ (\gamma + \lambda_1)\log y_{i,t-1} \\ &+ (\lambda_2 - \gamma\lambda_1)\log y_{i,t-2} - \gamma\lambda_2 \log y_{i,t-3} \\ &+ \text{white-noise error term}\end{aligned}$$

to differ across subpopulations. For example, if racial discrimination in the United States causes African-American families to have a lower earnings function intercept, they would also have a lower intercept in the multigenerational mobility equation [Duncan (1968), Hertz (2005)].

A failure to model such group-specific intercepts in the multigenerational mobility equation amounts to omission of group fixed effects. Applying the usual omitted-variables-bias analysis shows that, because parental log income, grandparental log income, and great-grandparental log income all have positive partial correlations with the omitted group effect, all the ancestral coefficients are pushed in a positive direction. And this would be a force towards slower-than-geometric decay in intergenerational autocorrelations.

Example #3: Measurement error

Even if the grandparental coefficient is not really positive, it might appear to be because of measurement error.

E.g., suppose the true process is AR(1) with a 0.5 parental coefficient. Then the true intergenerational autocorrelations would decline geometrically – a 0.5 correlation between child and parents, 0.25 between child and grandparents, 0.125 between child and great-grandparents, etc.

But now suppose each generation's status is measured with purely classical measurement error, and suppose the measured variation in each generation consists 80% of true variation and 20% of measurement noise. Then the attenuation factor for each measured autocorrelation would be 0.8. Consequently, the *measured* autocorrelations would tend towards 0.4 between child and parents, 0.2 between child and grandparents, 0.1 between child and great-grandparents.

Because these *measured* autocorrelations decline more slowly than geometrically, fitting an autoregression of child's status on both parental and grandparental status would result in a spuriously positive coefficient estimate for grandparental status.

III. Clark's Interpretation

Greg Clark and his collaborators have gathered data, from many countries over many centuries, on socioeconomic outcomes for individuals with rare surnames. Clark et al. find that, when they aggregate individuals by surname, the group-level correlation between adjacent generations in various SES measures tends to be about 0.75. The autocorrelations decline approximately geometrically as they are measured across 2, 3, 4, etc. generations.

Clark and his co-authors interpret these high and persistent intergenerational autocorrelations as reflecting the "true rate of social mobility." They claim that conventional family-level measures give a misleading impression of mobility due to a sort of errors-in-variables bias: An individual-level measure of income, education, occupational prestige, or whatever is merely a noisy measure of the individual's "fundamental social competence or status." Aggregating individuals by surname treats this errors-in-variables bias by averaging out the noise.

Clark concludes that an AR(1) intergenerational mobility regression with a coefficient around 0.75 is a "law of social mobility" applicable to all societies in all eras:

"Surname evidence shows that all social mobility can essentially be reduced to one simple law,

$$x_{t+1} = bx_t + e_t \,,$$

where *x* is the underlying social competence of families. The persistence rate, *b*, is always high relative to conventional estimates, generally 0.7-0.8. It seems to be little affected by social institutions."

This is the basis for the Clark quotation at the beginning of this talk. If the autoregressive coefficient is always as high as 0.7-0.8, it takes hundreds of years for regression to the mean to play out.

This is a fascinating and provocative reinterpretation of the large body of empirical evidence on intergenerational mobility. But is this interpretation empirically supported?

To his credit, Clark notes some testable implications of his hypothesis.

First, intergenerational mobility coefficient estimates should come out to about 0.75 when the data are aggregated not only by rare surnames, but by *any* grouping – "race, religion, national origin, or even common surnames." Is this really borne out?

E.g: Card, DiNardo, and Estes (2000) use U.S. decennial censuses to estimate intergenerational regressions of years of education or log weekly earnings for immigrants, grouping by country of origin. Their estimates run at "only" about 0.45. Their estimates are similar to previous results reported by Borjas (1993).

E.g.: Aaronson and Mazumder (2008) also use U.S. decennial censuses to estimate intergenerational regressions of men's log annual earnings on the log of the average income for their parents' generation in the men's state of birth. Their estimates also average at about 0.45.

E.g.: The forthcoming *QJE* article in which Chetty et al. use U.S. tax return data to study intergenerational income mobility includes an appendix with group-average regressions based on surnames. Using all surnames, they estimate an intergenerational income elasticity of 0.42. Their estimates using only rare surnames are about 0.35.

Second, the forthcoming *EJ* article by Clark and Cummins says that, "if we were to measure the social status of families as an aggregate of earnings, wealth, education, occupation, and health, then observed social mobility even in parent child studies would decline. For such an aggregation would reduce the variance of the error component in measured status. Thus the measured rate of persistence, even in one generation, will be much closer to that of the underlying latent variable."

Kelly Vosters, the Michigan State Ph.D. student who is starting to run three-generation income regressions with the PSID, also has written an excellent paper that uses the PSID to test the Clark-Cummins prediction. When she uses the Lubotsky-Wittenberg multiple-proxies method to construct an aggregate parental index supplementing log income with education and occupation measures, the intergenerational elasticity estimate goes up only a little, from 0.44 to 0.47. It is not "much closer" to the allegedly universal value of 0.75.

So when we follow up on testable predictions of the Clark hypothesis, it doesn't perform so well. But if Clark doesn't have the right story for why his group-level autocorrelations for rare surnames come out so high, what is the alternative explanation?

To illustrate that there could be other possibilities, let's return to the group-effects explanation of positive grandparental coefficients.

Suppose that the SES of family *i* in group *g* in generation *t* can be decomposed as

$$y_{igt} = a_{gt} + b_{igt}$$

where the *a* term is a group-level (e.g., surname) average and the *b* term is an orthogonal family-specific deviation from the group average.

Following a suggestion in footnote 13 in Becker and Tomes (1979), suppose that the intergenerational process for the group average *a* is AR(1), with a coefficient of 0.8 (thus according with Clark's group-level evidence). Also suppose that *b* separately follows an AR(1) process with a coefficient of 0.3, and that the cross-sectional variance of *y* is 60% within-group and 40% between-group.

Then it is easy to calculate that, at the family level, the first-order intergenerational correlation is 0.5 (a weighted average of 0.8 and 0.3). The higher-order autocorrelations are 0.31 for two generations apart, 0.22 for three, 0.17 for four, and 0.13 for five.

Note the following points about this illustrative example:

- By construction, it accords with Clark's group-level evidence. It involves a 0.8 first-order autocorrelation at the group level, which declines geometrically at higher orders.
- At the individual level, the first-order autocorrelation is much smaller. Unlike in Clark's interpretation, in this story the smaller individual autocorrelations are not spuriously attenuated by errors-in-variables bias, but reflect the true individual-level social mobility.
- In accordance with some of the multigenerational evidence discussed earlier, the individual-level autocorrelations decline more slowly than at a geometric rate.

Of course, none of this proves that the model I just made up is definitive. But it illustrates that Clark's theory need not be the only possible explanation of his rare surnames evidence and other empirical regularities.

What Do We Know So Far about Multigenerational Mobility?

1. Some fragments of multigenerational evidence are roughly consistent with an AR(1) mobility process. Others suggest that the coefficient of grandparental status is positive, so that intergenerational autocorrelations decay more slowly than geometrically.

2. Where positive grandparental coefficients are estimated, there are many possible sources. These include direct causal effects from grandparents (e.g., cultural inheritance effects when grandparents are present in the children's lives), group effects, and errors-in-variables bias.

3. The version of the errors-in-variables story due to Clark is rejected by empirical tests suggested by Clark himself.

4. That highlights some good news: Furthering our understanding of multigenerational mobility is amenable to empirical research. We know much more about social mobility than we did 30 years ago, and we are continuing to make progress.