Multiplicity in Intergenerational Transmission: Lessons from Surnames

Andrea Del Pizzo Jan Stuhler

June 14, 2024

ABSTRACT

In this paper, we present evidence for multiple channels in intergenerational transmission by analyzing surname-based data. When social status depends on various determinants with differing persistence rates, persistent characteristics become more crucial in explaining correlations with distant relatives. Using US Census data from 1920-1940, we test this hypothesis by examining variations in surname group sizes. Larger surname groups include more distant relatives, while smaller groups reflect closer familial connections. Our findings show varying persistence rates across observable and unobservable characteristics. Persistent traits, such as geographical location and ethnicity, become increasingly important in larger groups, while residual individual traits become less significant. This explains the observed greater surname-level persistence in larger groups. The existence of multiple intergenerational transmission channels has two key implications. First, it explains the high persistence of socioeconomic status across generations due to the increasing importance of persistent factors among distant relatives. Second, it rationalises that the degree of social mobility in different families depends on the persistence of the determinants of their status.

1 INTRODUCTION

The study of the transmission of socioeconomic status (SES) is a key topic across various social sciences. Historically, economists have focused on the impact of family background on the socioeconomic outcomes of offspring (Chetty et al. [2014]). However, recent studies have established intergenerational links across multiple generations, showing a greater persistence in socioeconomic status than previously thought (Lindahl et al. [2015]; Adermon et al. [2021]). Despite evidence contradicting the notion of a geometric decrease in intergenerational correlation over generations (Collado et al. [2023]), the underlying statistical process remains unclear.

To explain the high persistence across multiple generations, recent literature has focused on the transmission of unobservable or latent characteristics, in addition to observable ones (Collado et al.

[2023]). The Latent Factor Model (LFM) (Braun and Stuhler [2018]) suggests that socioeconomic status is transmitted through a single latent construct across generations. While this model accounts for measurement errors and helps quantify the persistence degree over multiple generations, it obscures various potential explanations for this persistence. This paper empirically supports the idea that one source of measurement error in intergenerational transmission is the presence of non-linearities, resulting from variability in persistence rates across different socioeconomic factors. Under such a setting, persistent characteristics become more crucial in explaining correlations with distant relatives. This hypothesis can be tested using surnames since larger surname groups include more distant relatives, while smaller groups reflect closer familial connections. By leveraging variation in surname group size, we demonstrate that persistent characteristics are more important in large groups, consistent with the theory.

First, we argue that existing surname-based evidence (Santavirta and Stuhler [2020]; Güell et al. [2015]; Chetty et al. [2014]) provides a valuable benchmark for testing theories of intergenerational transmission. Using US Census data from 1920-1940, we first document the decreasing pattern of the Informational Content of Surnames (ICS) — the share of the outcome variable explained by surnames for the male working-age population in 1940 — on Occupational Score over surname group size. Second, we observe a slower regression to the mean in larger surname groups by estimating the surname-level Intergenerational Elasticity (IGE).

To explain existing surname-based evidence, we adopt a simultaneous equations model of intergenerational transmission (Conlisk [1974]; Goldberger [1989]). This model assumes that various characteristics, although correlated and with distinct transmission rates, contribute to a single socioeconomic outcome. This approach is particularly useful in the context of surnames, enabling a comprehensive analysis of both individual and aggregate-level mobility (Torche and Corvalan [2018]). Additionally, our model supports the observed heterogeneity in socioeconomic persistence among different families, suggesting that the degree of mobility depends on the family's reliance on persistent factors. While prior literature has explored the theoretical implications of this model, it has received less empirical attention than other alternatives. Our aim is to address this gap.

First, we derive the theoretical implications of variations of surname group size according to our model. It yields two main empirically untestable propositions. First, correlations within surname groups rely on the presence of a common ancestor. Second, the average distance to the common ancestor increases as surname group size grows. In essence, larger surname groups tend to encompass more distant relatives, while smaller groups capture closer familial connections. This proposition hinges on the assumption of positive population growth, consistent with historical trends in the US during the studied period. From these two untestable propositions, we derive two corollaries with testable implications. First, larger surname groups, characterized by weaker familial connections, should exhibit lower within-group correlation of socioeconomic characteristics, assuming that the characteristics of the common ancestor remain independent of surname group size. We provide suggestive evidence for this assumption by demonstrating that parent-child correlations do not consistently vary across surname groups of different sizes. Second, as surname group size increases, within-group correlation of different traits is expected to decrease at a rate inversely proportional to their persistence rate. As this distance grows, highly persistent factors maintain a larger degree of within-group correlation, while less enduring traits tend to swiftly average out. Consequently, while small groups may still encompass characteristics that fade quickly across generations, larger surname groups almost exclusively retain persistent factors.

Then, we study how these two corollaries impact the evolution of the ICS over surname group size and test whether the implications hold in the data. According to the first Corollary the ICS should decrease in the data, as it is an increasing function of within-surname group correlation. This rationalizes the first statistical fact that moved our analysis. According to the second Corollary, the composition of the ICS should vary as surname group size expands. We decompose the ICS into portions attributable to distinct observable characteristics, such as race, state of residence, urban/rural status, etc. We then examine how these components change across different intervals of surname group sizes. We find that the share of the ICS explained by all group-level characteristics increase over surname group size. This indicates that group-level characteristics are more persistent than residual uncorrelated traits.

Then, we examine how these two corollaries affect the Grouping estimator, i.e. the surname-level IGE. First, we show that, under our statistical model, such estimator is a weighted average of the persistence of the relevant characteristics for the outcome variable. In turn, the weights of each factor are a function of its relevance and of its degree of within-group correlation. Two key channels drive changes in these weights. First, averaging over larger surnames reduces the impact of market luck, which previously generated attenuation bias. However, we show that this channel alone cannot fully account for surname-based evidence. Second, even abstracting from finite-sample concerns, within-group correlation depends on the distance to the common ancestor. By Corollary 2, highly persistent factors maintain a higher degree of within-group correlation than less enduring traits with growing surname group size. Consequently, the estimator increasingly assigns more weight to persistent factors as group size expands. This explanation rationalizes the second statistical fact and allows for the empirical estimation of the weights for various observable characteristics. First, we demonstrate that increasing weights with surname group size indicate that the characteristic is more persistent than residual traits. We empirically confirm that all employed group-level characteristics exhibit increasing weighting patterns. Second, by normalizing the weights in each interval and comparing growth curves over surname group size, we directly compare the rates of persistence across observable characteristics. Our findings reveal that race and geographical characteristics, such as state of residence and birthplace, exhibit higher persistence compared to other characteristics such as urban/rural status or education.

Second, we analyse how controlling for different covariates affects the surname-level IGE. We demonstrate that the conditional surname-level IGE identifies the persistence of uncorrelated traits with respect to the controlling variable Therefore, taking the ratio between the conditional and unconditional estimates provides a measure of the persistence of the characteristic itself. On the one hand, increasing ratios over surname group size imply that the covariate is more persistent than residual traits. Once again, we find that all observable characteristics demonstrate higher persistence compared to unobservable ones, validating the previous analysis.

This work contributes to different strands in the intergenerational mobility literature. First, we contribute to a better understanding of the pattern of intergenerational transmission. While the Latent Factor Model is a useful approximation (Braun and Stuhler [2018]), we show that empirical patterns in the data point toward the existence of multiple channels in the intergenerational transmission. This suggests non-linearities in the transmission process across generations and families. On the one hand, we claim that this non-linearity represents a crucial source of excess persistence over multiple generations. Correlations between close generations are mostly accounted for by relevant but not necessarily persistent characteristics, whereas this pattern reverses in correlations between more distant generations. This implies that the decrease in intergenerational correlations shrinks as the distance between generations grows. On the other, we argue that family-level heterogeneity stems from the uneven distribution of persistent factors across the population.

Second, and related, this sheds new light on the interpretation of surname-based estimates of intergenerational mobility, whose interpretation has been controversial (Santavirta and Stuhler [2020]; Güell et al. [2015]; Clark [2014]). We demonstrate that the estimand of surname-based estimators is not constant, but varies across the distribution of group size. On the one hand, relying solely on the Grouping estimator based on large surname groups may provide limited insights into the intergenerational transmission between parents and children. While some factors might be very persistent, they might also be unimportant in the intergenerational process in the short run. On the other hand, Grouping estimator provide valuable insights into transmission processes that direct parent-child estimates cannot fully capture. For instance, we document the significant role played by environmental characteristics, such as geography, in the transmission process in the long run. Despite their modest contribution to parent-child correlations, their significance amplifies over multiple generations. Hence, neglecting these factors could lead to underestimations of socio-economic status persistence, particularly in the long run. In conclusion, surname-based estimates and parent-child correlations complement each other rather than serving as substitute measures of intergenerational transmission.

The paper proceeds as follows. In the next section we present the data. We document surname-based empirical evidences in Section 3. Section 4 introduces the model and discusses the theoretical implications of surname group size. We examine the implications of our model on the Informational Content of Surnames (ICS) in Section 5 and on the Grouping estimator in Section 6. Section 7 validates previous results. Section 8 presents some implications of previous findings on multi-generational inequality and heterogeneity across families. Section 10 concludes.

2 Data

For the empirical analysis, we use data from the US Census of 1920 and 1940, obtained from the IPUMS database. We focus exclusively on the male population for two reasons: first, female employment rates were minimal during this period, and second, surname transmission occurs through the male lineage.

We include the entire male working-age population (18-64 years old) to compute surname averages in both Censuses. In the 1920 US Census, surname averages are used as a regressor for surname-based estimators. In the 1940 US Census, surname averages are necessary to calculate the Informational Content of Surnames at the population level. Our sample reveals a significant majority of white individuals, resulting in a predominantly homogeneous sample in terms of race, as tracking African-Americans during this period is challenging.

The sample for surname-based estimation consists of male children aged between 10 and 20 years in 1920 who were successfully linked to the same individuals in the 1940 Census, when they were between 30 and 40 years old. Within this final dataset, we calculate the rate of overlap, representing the percentage of children for whom we have observed the father's outcome, contributing to the computation of the surname-level average. In our analysis, this rate exceeds 95%, indicating an almost complete overlap in the sample. As noted by Santavirta and Stuhler [2020], the absence of overlap can introduce attenuation bias, potentially impacting the results.

The linking procedure relies on exact matching based on sex, race, and birthplace. Additionally, IPUMS uses an algorithm to evaluate name and age similarity to account for potential errors in the data preparation stage. Similar to the parental sample, we obtain a highly homogeneous sample, overwhelmingly comprised of white individuals.

Due to the absence of income information in the 1920 Census, we use the Occupational Score, which is the median income for a given occupation in that period, as the outcome variable. We provide robustness checks with alternative outcome variables in the Appendix. Additionally, we use education levels from the 1940 Census, geographical location (the state in which the child is living in 1940), birthplace, and urban/rural status as covariates.

Our analysis leverages differences in surname group size to gain insights into intergenerational transmission. In Tables **??**-**??**, we show the characteristics of the sample across surname group sizes. We observe that the sample becomes more rural as surname group size increases. However, we perform a simple reweighting exercise to balance the various groups as a robustness check, and we find that the main results are not affected (Figure 27).

	size_group							
	25	50	100	150	300	1000	1500	Total
N	227,402 (23.1%)	85,140 (8.6%)	105,590 (10.7%)	70,234 (7.1%)	130,865 (13.3%)	263,225 (26.7%)	101,987 (10.4%)	984,443 (100.0%)
Race								
White	225,080 (99.0%)	84,278 (99.0%)	104,378 (98.9%)	69,319 (98.7%)	128,837 (98.5%)	258,254 (98.1%)	99,695 (97.8%)	969,841 (98.5%)
Black/African American/Negro	1,934 (0.9%)	766 (0.9%)	1,118 (1.1%)	849 (1.2%)	1,909 (1.5%)	4,747 (1.8%)	2,183 (2.1%)	13,506 (1.4%)
American Indian/Alaska Native (AIAN)	323 (0.1%)	73 (0.1%)	74 (0.1%)	42 (0.1%)	91 (0.1%)	196 (0.1%)	96 (0.1%)	895 (0.1%)
Chinese	1 (0.0%)	2 (0.0%)	3 (0.0%)	5 (0.0%)	5 (0.0%)	17 (0.0%)	10 (0.0%)	43 (0.0%)
Japanese	62 (0.0%)	19 (0.0%)	16 (0.0%)	17 (0.0%)	22 (0.0%)	8 (0.0%)	2 (0.0%)	146 (0.0%)
Filipino	2 (0.0%)	2 (0.0%)	1 (0.0%)	1 (0.0%)	1 (0.0%)	2 (0.0%)	1 (0.0%)	10 (0.0%)
Korean	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.0%)	0 (0.0%)	1 (0.0%)
Native Hawaiian	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.0%)
Educational attainment								
No schooling completed	1,308 (0.6%)	443 (0.5%)	569 (0.5%)	401 (0.6%)	883 (0.7%)	1,859 (0.7%)	662 (0.7%)	6,125 (0.6%)
Kindergarten	0 (0.0%)	1 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.0%)
Grade 1	347 (0.2%)	115 (0.1%)	154 (0.1%)	121 (0.2%)	232 (0.2%)	544 (0.2%)	239 (0.2%)	1,752 (0.2%)
Grade 2	754 (0.3%)	289 (0.3%)	345 (0.3%)	272 (0.4%)	549 (0.4%)	1,274 (0.5%)	566 (0.6%)	4,049 (0.4%)
Grade 3	1,630 (0.7%)	601 (0.7%)	794 (0.8%)	542 (0.8%)	1,176 (0.9%)	2,592 (1.0%)	1,131 (1.1%)	8,466 (0.9%)
Grade 4	3,586 (1.6%)	1,359 (1.6%)	1,699 (1.6%)	1,195 (1.7%)	2,331 (1.8%)	5,423 (2.1%)	2,138 (2.1%)	17,731 (1.8%)
Grade 5	4,970 (2.2%)	1,865 (2.2%)	2,312 (2.2%)	1,622 (2.3%)	2,980 (2.3%)	6,612 (2.6%)	2,660 (2.7%)	23,021 (2.4%)
Grade 6	11,342 (5.1%)	3,935 (4.7%)	4,759 (4.6%)	3,122 (4.5%)	5,897 (4.6%)	11,923 (4.6%)	4,651 (4.6%)	45,629 (4.7%)
Grade 7	18,729 (8.4%)	6,661 (8.0%)	8,238 (7.9%)	5,425 (7.9%)	10,090 (7.9%)	19,650 (7.6%)	7,529 (7.5%)	76,322 (7.9%)
Grade 8	77,012 (34.4%)	28,508 (34.1%)	34,243 (33.0%)	22,461 (32.5%)	40,054 (31.2%)	76,453 (29.6%)	28,547 (28.5%)	307,278 (31.8%)
Grade 9	15,269 (6.8%)	5,705 (6.8%)	7,111 (6.9%)	4,709 (6.8%)	8,941 (7.0%)	18,164 (7.0%)	7,038 (7.0%)	66,937 (6.9%)
Grade 10	19,683 (8.8%)	7,464 (8.9%)	9,293 (9.0%)	5,998 (8.7%)	11,487 (8.9%)	22,955 (8.9%)	8,736 (8.7%)	85,616 (8.9%)
Grade 11	9,796 (4.4%)	3,673 (4.4%)	4,879 (4.7%)	3,215 (4.7%)	6,151 (4.8%)	12,883 (5.0%)	5,166 (5.2%)	45,763 (4.7%)
Grade 12	33,932 (15.2%)	13,054 (15.6%)	16,647 (16.1%)	11,448 (16.6%)	21,245 (16.5%)	43,694 (16.9%)	17,204 (17.2%)	157,224 (16.3%)
1 year of college	3,990 (1.8%)	1,641 (2.0%)	2,178 (2.1%)	1,414 (2.0%)	2,742 (2.1%)	5,943 (2.3%)	2,347 (2.3%)	20,255 (2.1%)
2 years of college	5,048 (2.3%)	1,963 (2.3%)	2,578 (2.5%)	1,710 (2.5%)	3,478 (2.7%)	7,379 (2.9%)	2,909 (2.9%)	25,065 (2.6%)
3 years of college	2,337 (1.0%)	942 (1.1%)	1,153 (1.1%)	822 (1.2%)	1,587 (1.2%)	3,386 (1.3%)	1,324 (1.3%)	11,551 (1.2%)
4 years of college	9,214 (4.1%)	3,667 (4.4%)	4,549 (4.4%)	2,986 (4.3%)	5,751 (4.5%)	12,060 (4.7%)	4,890 (4.9%)	43,117 (4.5%)
5+ years of college	3,737 (1.7%)	1,361 (1.6%)	1,695 (1.6%)	1,223 (1.8%)	2,252 (1.8%)	4,536 (1.8%)	1,839 (1.8%)	16,643 (1.7%)
6 years of college (6+ in 1960-1970)	367 (0.2%)	163 (0.2%)	189 (0.2%)	116 (0.2%)	230 (0.2%)	474 (0.2%)	214 (0.2%)	1,753 (0.2%)
7 years of college	273 (0.1%)	103 (0.1%)	135 (0.1%)	98 (0.1%)	169 (0.1%)	340 (0.1%)	137 (0.1%)	1,255 (0.1%)
8+ years of college	366 (0.2%)	151 (0.2%)	178 (0.2%)	122 (0.2%)	240 (0.2%)	452 (0.2%)	143 (0.1%)	1,652 (0.2%)
Urban/rural status								
Rural	89,084 (39.2%)	37,242 (43.7%)	47,724 (45.2%)	32,425 (46.2%)	61,688 (47.1%)	130,914 (49.7%)	51,812 (50.8%)	450,889 (45.8%)
Urban	138,318 (60.8%)	47,898 (56.3%)	57,866 (54.8%)	37,809 (53.8%)	69,177 (52.9%)	132,311 (50.3%)	50,175 (49.2%)	533,554 (54.2%)
State (ICPSR code)								
Connecticut	4,150 (1.8%)	1,185 (1.4%)	1,311 (1.2%)	796 (1.1%)	1,580 (1.2%)	2,844 (1.1%)	1,104 (1.1%)	12,970 (1.3%)
Maine	854 (0.4%)	322 (0.4%)	490 (0.5%)	395 (0.6%)	813 (0.6%)	2,042 (0.8%)	950 (0.9%)	5,866 (0.6%)
Massachusetts	8,579 (3.8%)	2,562 (3.0%)	2,963 (2.8%)	1,931 (2.7%)	3,891 (3.0%)	7,994 (3.0%)	3,264 (3.2%)	31,184 (3.2%)
New Hampshire	697 (0.3%)	249 (0.3%)	292 (0.3%)	216 (0.3%)	474 (0.4%)	1,137 (0.4%)	474 (0.5%)	3,539 (0.4%)
Rhode Island	1,643 (0.7%)	488 (0.6%)	614 (0.6%)	441 (0.6%)	824 (0.6%)	1,549 (0.6%)	681 (0.7%)	6,240 (0.6%)
Vermont	377 (0.2%)	159 (0.2%)	206 (0.2%)	159 (0.2%)	268 (0.2%)	682 (0.3%)	365 (0.4%)	2,216 (0.2%)
Delaware	242 (0.1%)	85 (0.1%)	131 (0.1%)	82 (0.1%)	164 (0.1%)	429 (0.2%)	164 (0.2%)	1,297 (0.1%)
New Jersey	9,517 (4.2%)	2,862 (3.4%)	3,357 (3.2%)	2,058 (2.9%)	3,678 (2.8%)	6,821 (2.6%)	2,434 (2.4%)	30,727 (3.1%)
New York	26,427 (11.6%)	8,136 (9.6%)	9,824 (9.3%)	6,078 (8.7%)	10,513 (8.0%)	19,297 (7.3%)	7,020 (6.9%)	87,295 (8.9%)
Pennsylvania	19,672 (8.7%)	6,662 (7.8%)	8,570 (8.1%)	6,010 (8.6%)	11,352 (8.7%)	21,824 (8.3%)	7,468 (7.3%)	81,558 (8.3%)
Illinois	24,292 (10.7%)	8,170 (9.6%)	9,419 (8.9%)	5,899 (8.4%)	10,667 (8.2%)	19,437 (7.4%)	7,224 (7.1%)	85,108 (8.6%)
Indiana	6,035 (2.7%)	2,702 (3.2%)	3,749 (3.6%)	2,541 (3.6%)	4,958 (3.8%)	10,429 (4.0%)	3,986 (3.9%)	34,400 (3.5%)
Michigan	13,498 (5.9%)	4,935 (5.8%)	5,561 (5.3%)	3,575 (5.1%)	6,322 (4.8%)	11,721 (4.5%)	4,506 (4.4%)	50,118 (5.1%)
Ohio	15,497 (6.8%)	6,225 (7.3%)	8,058 (7.6%)	5,408 (7.7%)	9,894 (7.6%)	19,358 (7.4%)	7,179 (7.0%)	71,619 (7.3%)
Wisconsin	12,531 (5.5%)	5,482 (6.4%)	6,267 (5.9%)	3,826 (5.4%)	6,286 (4.8%)	10,183 (3.9%)	3,554 (3.5%)	48,129 (4.9%)
Iowa	6,739 (3.0%)	3,277 (3.8%)	3,875 (3.7%)	2,498 (3.6%)	4,583 (3.5%)	8,622 (3.3%)	3,228 (3.2%)	32,822 (3.3%)
Kansas	3,559 (1.6%)	1,592 (1.9%)	2,174 (2.1%)	1,515 (2.2%)	2,780 (2.1%)	5,682 (2.2%)	2,328 (2.3%)	19,630 (2.0%)
Minnesota	12,498 (5.5%)	4,797 (5.6%)	5,608 (5.3%)	3,450 (4.9%)	5,548 (4.2%)	9,851 (3.7%)	3,213 (3.2%)	44,965 (4.6%)
Missouri	8,006 (3.5%)	3,297 (3.9%)	3,916 (3.7%)	2,689 (3.8%)	4,882 (3.7%)	10,359 (3.9%)	4,172 (4.1%)	37,321 (3.8%)
Nebraska	4,124 (1.8%)	1,874 (2.2%)	2,243 (2.1%)	1,340 (1.9%)	2,479 (1.9%)	4,519 (1.7%)	1,583 (1.6%)	18,162 (1.8%)
North Dakota	2,707 (1.2%)	1,180 (1.4%)	1,173 (1.1%)	785 (1.1%)	1,239 (0.9%)	2,115 (0.8%)	818 (0.8%)	10,017 (1.0%)
South Dakota	2,361 (1.0%)	1,000 (1.2%)	1,168 (1.1%)	821 (1.2%)	1,250 (1.0%)	2,172 (0.8%)	781 (0.8%)	9,553 (1.0%)
Virginia	1,261 (0.6%)	633 (0.7%)	946 (0.9%)	707 (1.0%)	1,662 (1.3%)	4,191 (1.6%)	1,835 (1.8%)	11,235 (1.1%)
Alabama	1,359 (0.6%)	489 (0.6%)	760 (0.7%)	535 (0.8%)	1,302 (1.0%)	3,417 (1.3%)	1,669 (1.6%)	9,531 (1.0%)
Arkansas	949 (0.4%)	390 (0.5%)	523 (0.5%)	360 (0.5%)	990 (0.8%)	2,468 (0.9%)	1,175 (1.2%)	6,855 (0.7%)
Florida	740 (0.3%)	315 (0.4%)	461 (0.4%)	300 (0.4%)	648 (0.5%)	1,475 (0.6%)	601 (0.6%)	4,540 (0.5%)
Georgia	888 (0.4%)	380 (0.4%)	630 (0.6%)	483 (0.7%)	1,123 (0.9%)	3,207 (1.2%)	1,496 (1.5%)	8,207 (0.8%)
Louisiana	3,068 (1.3%)	1,422 (1.7%)	1,879 (1.8%)	1,255 (1.8%)	2,445 (1.9%)	3,948 (1.5%)	1,373 (1.3%)	15,390 (1.6%)
Mississippi	902 (0.4%)	311 (0.4%)	488 (0.5%)	445 (0.6%)	888 (0.7%)	2,447 (0.9%)	1,046 (1.0%)	6,527 (0.7%)
North Carolina	1,161 (0.5%)	456 (0.5%)	801 (0.8%)	743 (1.1%)	1,783 (1.4%)	5,051 (1.9%)	2,270 (2.2%)	12,265 (1.2%)
South Carolina	586 (0.3%)	202 (0.2%)	391 (0.4%)	322 (0.5%)	746 (0.6%)	1,745 (0.7%)	862 (0.8%)	4,854 (0.5%)
Texas	5,255 (2.3%)	2,397 (2.8%)	3,092 (2.9%)	2,046 (2.9%)	4,120 (3.1%)	9,548 (3.6%)	3,895 (3.8%)	30,353 (3.1%)
Kentucky	2,298 (1.0%)	964 (1.1%)	1,436 (1.4%)	1,120 (1.6%)	2,429 (1.9%)	6,026 (2.3%)	2,636 (2.6%)	16,909 (1.7%)
Maryland	2,728 (1.2%)	1,031 (1.2%)	1,416 (1.3%)	988 (1.4%)	1,947 (1.5%)	3,882 (1.5%)	1,493 (1.5%)	13,485 (1.4%)
Oklahoma	1,575 (0.7%)	624 (0.7%)	1,018 (1.0%)	774 (1.1%)	1,549 (1.2%)	3,741 (1.4%)	1,572 (1.5%)	10,853 (1.1%)
Tennessee	1,357 (0.6%)	588 (0.7%)	992 (0.9%)	802 (1.1%)	1,681 (1.3%)	4,696 (1.8%)	2,213 (2.2%)	12,329 (1.3%)
West Virginia	1,283 (0.6%)	510 (0.6%)	825 (0.8%)	657 (0.9%)	1,404 (1.1%)	3,666 (1.4%)	1,547 (1.5%)	9,892 (1.0%)
Arizona	289 (0.1%)	83 (0.1%)	131 (0.1%)	100 (0.1%)	216 (0.2%)	473 (0.2%)	191 (0.2%)	1,483 (0.2%)
Colorado	1,665 (0.7%)	725 (0.9%)	955 (0.9%)	639 (0.9%)	1,144 (0.9%)	2,433 (0.9%)	985 (1.0%)	8,546 (0.9%)
Idaho	598 (0.3%)	320 (0.4%)	357 (0.3%)	295 (0.4%)	579 (0.4%)	1,166 (0.4%)	437 (0.4%)	3,752 (0.4%)
Montana	1,111 (0.5%)	410 (0.5%)	522 (0.5%)	302 (0.4%)	592 (0.5%)	1,178 (0.4%)	475 (0.5%)	4,590 (0.5%)
Nevada	121 (0.1%)	51 (0.1%)	51 (0.0%)	33 (0.0%)	75 (0.1%)	158 (0.1%)	61 (0.1%)	550 (0.1%)
New Mexico	271 (0.1%)	135 (0.2%)	176 (0.2%)	141 (0.2%)	298 (0.2%)	832 (0.3%)	327 (0.3%)	2,180 (0.2%)
Utah	483 (0.2%)	302 (0.4%)	461 (0.4%)	291 (0.4%)	584 (0.4%)	1,140 (0.4%)	402 (0.4%)	3,663 (0.4%)
Wyoming	317 (0.1%)	150 (0.2%)	177 (0.2%)	130 (0.2%)	226 (0.2%)	521 (0.2%)	192 (0.2%)	1,713 (0.2%)
California	8,462 (3.7%)	3,049 (3.6%)	3,708 (3.5%)	2,612 (3.7%)	4,873 (3.7%)	10,195 (3.9%)	4,074 (4.0%)	36,973 (3.8%)
Oregon	1,517 (0.7%)	634 (0.7%)	843 (0.8%)	585 (0.8%)	1,145 (0.9%)	2,327 (0.9%)	986 (1.0%)	8,037 (0.8%)
Washington	2,730 (1.2%)	1,132 (1.3%)	1,356 (1.3%)	904 (1.3%)	1,647 (1.3%)	3,537 (1.3%)	1,353 (1.3%)	12,659 (1.3%)
District of Columbia	423 (0.2%)	196 (0.2%)	226 (0.2%)	152 (0.2%)	324 (0.2%)	690 (0.3%)	325 (0.3%)	2,336 (0.2%)

	size_group							
	25	50	100	150	300	1000	1500	Total
N	227,402 (23.1%)	85,140 (8.6%)	105,590 (10.7%)	70,234 (7.1%)	130,865 (13.3%)	263,225 (26.7%)	101,987 (10.4%)	984,443 (100.0%)
Occupational income score	24.467 (11.421)	24.477 (11.545)	24.405 (11.452)	24.304 (11.391)	24.302 (11.440)	24.191 (11.429)	24.180 (11.519)	24.324 (11.448)
Duncan Socioeconomic Index	29.072 (22.761)	29.491 (22.953)	29.440 (22.939)	29.207 (22.892)	29.277 (22.997)	29.225 (23.069)	29.227 (23.236)	29.242 (22.970)
Occupational prestige score, Siegel	34.513 (13.886)	35.199 (13.744)	35.195 (13.724)	35.067 (13.692)	35.086 (13.781)	35.112 (13.824)	35.155 (13.932)	34.988 (13.820)
Occupational education score, 1950 basis	48.307 (183.504)	43.892 (171.502)	44.281 (172.488)	43.903 (171.834)	44.794 (173.815)	44.767 (173.540)	46.117 (176.706)	45.539 (175.858)
Occupational earnings score, 1950 basis	49.074 (29.339)	48.647 (29.956)	48.511 (29.968)	48.290 (30.012)	48.251 (30.031)	47.884 (30.123)	47.784 (30.276)	48.360 (29.912)
Nam-Powers-Boyd occupational status score, 1950 basis	83.448 (177.689)	78.622 (166.385)	78.855 (167.365)	78.321 (166.790)	79.051 (168.690)	78.620 (168.512)	79.712 (171.574)	79.910 (170.604)

Moreover, we complement the empirical analysis with a simulated population. For each statistical process of intergenerational transmission, we generate a fictitious population spanning ten generations. The initial generation's surname distribution comprises a fifth with unique surnames, followed by distributions of two, three, five, and ten for the subsequent fifths.

The population's evolution is modeled through a fertility process and a surname mutation process. Regarding fertility, we generate a growing population where 30% have no male child, 20% have one male child, 30% have two, 20% have three, and 10% have four. Concerning the surname mutation process, we assume that in each new generation, surnames randomly mutate with a 2% probability. This simulation feature is crucial for reproducing the empirical distribution of surnames in western countries, preventing convergence to a few large and uninformative surnames.

3 Statistical Facts - Surname Based

In this section, we present evidence on surname-based estimation in our data, linking it to the broader discussion in the literature. Specifically, we document two key stylized facts within our context. First, the proportion of the outcome variable explained by surname decreases with surname size. Second, the persistence of surname-level averages increases with surname size. While these facts have been noted in previous studies, we provide further insights.

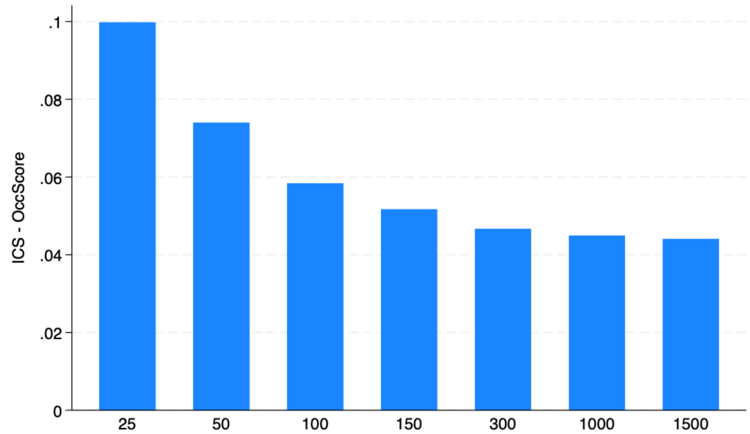
3.1 Statistical Fact 1 - Decreasing ICS

The Informational Content of Surnames (ICS) represents the share of the outcome variable explained by surname groups Güell et al. [2015]. The analytical procedure involves an initial regression of the outcome variable on surname indicators, producing the associated R_S^2. A subsequent regression of the outcome variable against a randomly generated indicator variable, which has the same distribution as the surname indicator, yields R_R^2. This method accounts for potential random explanations, particularly relevant in smaller groups. The ICS is then expressed as follows:

$$ICS = R_S^2 - R_R^2$$

We document the empirical relationship between ICS and surname group size in our data, focusing on the male working-age population in 1940 across various surname group size intervals (e.g., smaller than 25, between 25 and 50). The results are presented in the following figure:

FIGURE 1: ICS over Surname Size



Note: The figure depicts the ICS of the Occupational Score for the male working-age population in the 1940 US Census. Each data point represents the ICS, specifically focusing on surnames whose size falls within the range defined by the two preceding points.

This pattern aligns with findings from previous studies. For instance, Güell et al. [2015] used the ICS to deduce parent-child correlation through model calibration. Their approach used surnames as a proxy to measure standard Intergenerational Mobility, allowing them to extract the parameter of interest from a single cross-sectional wave of census data. The premise is that a surname's predictiveness of the outcome reflects the extent of intergenerational transmission.

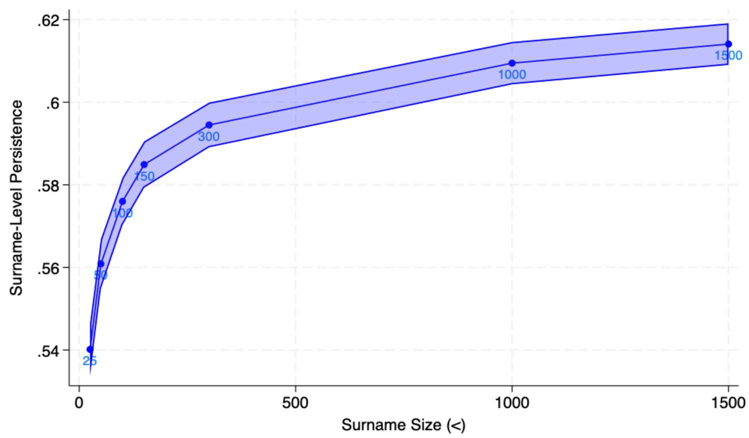
Previous research highlighted that the explanatory power of surnames decreases with increasing surname size. Larger surname groups exhibit greater within-group heterogeneity, reducing the correlation between surname averages and individual outcomes within the group. Consequently, even within a simplistic AR(1) model of transmission, averaging over larger groups mechanically diminishes the numerator. We extend this literature by illustrating that a model tailored for multi-generational evidence can explain this observed pattern.

3.2 STATISTICAL FACT 2 - INCREASING PERSISTENCE

The persistence of surname-level socioeconomic outcomes is estimated through the Intergenerational Elasticity (IGE) at the surname level, which amounts to a Two-Sample Two-Stage Least Squares (TSLS) method. This involves regressing the child outcomes onto the surname-level average in the parental generation, referred to as *Grouping* Santavirta and Stuhler [2020], following a three-stage procedure Hull [2017]. In the first stage, we regress parental outcomes on surname indicators. In the second stage, we use the resulting fitted values as an instrument for the IGE. However, a direct relationship exists between parental outcome (y_{it-1}) and parental surname average ($\overline{y}_{it-1}$), especially in small groups.

We document the empirical relationship between these estimators of persistence at the surname level and surname group size. Our analysis is based on a nearly fully overlapping sample, where extrapolation from the first to the second stage is limited, making our empirical patterns potentially contingent on a high overlapping rate. We estimate persistence by initially restricting our sample to small surname groups and then gradually enlarging the sample. For instance, we run the surname IGE using only surname groups smaller than 25, then gradually include surname groups with fewer than 50 members, and so forth. The results are presented in the figure below.

FIGURE 2: Persistence over Surname Size



*Note:*The figure illustrates the Persistence of Occupational Score at the surname level. We calculate the Grouping estimator on surname groups smaller than any specified size. Our analysis focuses exclusively on male workers aged between 30 and 40 in the 1940 US Census, linked to their fathers. The regressor employs the surname average computed from the working-age population in the 1920 US Census.

Previous studies used surname group averages to estimate Intergenerational Mobility for two reasons. First, in the absence of family links, surname instruments were the only viable option. Second, their goal was to explain long-term intergenerational transmission dynamics. For example, Clark [2014] employs a Latent Factor Model, where intergenerational transmission occurs at a latent level, using surname averages under the assumption that this latent component is shared within the surname group. His results indicate significant persistence in socioeconomic status transmission, sparking debate on the empirical design's validity. Replicating these findings, Chetty et al. [2014] shows that the persistence of surname averages tends to increase with surname size, though not to the extent observed in Clark's results.

To our knowledge, a formal explanation for this phenomenon has not been proposed. However, Chetty et al. [2014], Güell et al. [2018], and Torche and Corvalan [2018] argue that this trend can be explained by the discrepancy between aggregate mobility and individual-level mobility. They claim that surname-level persistence is driven by regional or ethnic characteristics, depending on the context.

We integrate Clark's original intuition with recent criticisms within a formal framework. We argue that surname-level analysis is still relevant for understanding long-term intergenerational transmission mechanisms. While surnames increasingly capture environmental characteristics as their size expands, we assert that this phenomenon arises from their higher persistence over multiple generations. Neglecting this aspect leads to a significant understatement of intergenerational persistence, especially in the long run. To substantiate this argument, we construct a flexible model encompassing multiple channels of socioeconomic persistence, formalizing the rationale behind this observed empirical trend

4 MODEL - MULTIPLE LATENT FACTOR MODEL

We present a model encompassing multiple factors which are passed through generations and affect the outcome variable at different rates. Such multiplicity can reconcile both surname-based and multigenerational evidence. Here, we illustrate a simplified representation with only two factors, though the model can accommodate more:

$$\begin{aligned} y_{ist} &= \rho_1 f_{1ist} + \rho_2 f_{2ist} + u_{it} \\ f_{1ist} &= \lambda_1 f_{1ist-1} + \epsilon_{1ist} \\ f_{2ist} &= \lambda_2 f_{2ist-1} + \epsilon_{2ist} \end{aligned}$$

The outcome variable y_{ist} for each individual i and surname s is influenced by latent factors f_{jit} through the relevance parameter ρ_j and a random shock u_{it}, assumed to be independent and identically distributed across individuals and generations. For instance, an individual's education (f_{1it}) impacts their income (y_{it}) based on the returns of education ρ_1. However, income is also shaped by other factors, including a stochastic component.

Intergenerational transmission occurs at the level of these factors. Each individual's factor (f_{jist}) is connected to their father's factor f_{jist-1} through the parameter λ_j and a random component ϵ_{ist} again assumed to be independent and identically distributed across individuals and generations. For example, an individual's education (f_{1ist}) depends on their father's education (f_{1ist-1}) with a persistence rate λ_1. Another factor, such as geographical location (f_{2ist}), may be transmitted at a different rate (λ_2) from the father's (f_{2ist-1}).

The introduction of multiple factors in intergenerational transmission offers a novel explanation for multigenerational evidence. Recent studies have noted excess persistence in the long run, where the correlation between the grandfather and the child exceeds the square of the parent-child correlation, as an AR(1) model would predict. This model provides a straightforward interpretation: the intergenerational transmission process is nonlinear. While some relevant characteristics fade rapidly over generations, others persist longer. This implies that the correlation between parent and child is driven by different characteristics than the correlation between grandparent and child. In the former case, highly relevant (high ρ) factors are more important, while in the latter, highly persistent factors (high λ) may prevail.

This model generalizes the Latent Factor Model (LFM) employed by Clark [2014], where intergenerational transmission occurs at a latent level. This latent factor is passed down through generations at a given rate (λ) and affects the outcome variable, such as income, through the parameter (ρ). The fundamental premise of this model is that using parental income as a metric for socioeconomic status is imprecise. The latent factor being transmitted actually encompasses a broader set of characteristics, which, if neglected, leads to attenuation bias in the estimation. The understatement of persistence in the long run results from incorrectly iterating the impact of measurement error over multiple generations.

While the LFM offers a useful approximation for quantifying multigenerational correlations, it simplifies many potential mechanisms by assuming a single construct of socioeconomic status, making its economic interpretation complex. In contrast, the Multiple Latent Factor Model (MLFM) is highly flexible, positing that any single socioeconomic outcome results from multiple characteristics, both observable and unobservable. The model only assumes linear transmission for each characteristic. However, the heterogeneity in parameters across different factors is sufficient to generate nonlinear transmission in the outcome variable. Thus, the observed excess persistence across multiple generations can be attributed to not accounting for these nonlinearities. Some characteristics persist more strongly over time than others, and failing to consider this leads to an underestimation of long-term persistence. Furthermore, this approach allows for the investigation of heterogeneity in social mobility across families.

In the following sections, we demonstrate how this model can rationalise existing surname-based evidence and unveil important aspects of the intergenerational transmission process

4.1 Surname Group Size

In this paper, we exploit variation in surname group size to investigate the existence of multiple intergenerational transmission channels. The underlying idea is that smaller surname groups denote closer familial ties, while larger groups encompass more distant relationships, leading to variations in the types of traits shared within these groups. Small groups tend to maintain correlations even for traits that fade over time, whereas large groups predominantly share more persistent traits. First, we formalize this intuition within the framework of the MLFM. Then, we examine its implications on empirical data to test its reliability.

Within the MLFM framework, we introduce the notion of a *common ancestor*. Under the premise of tracing solely male lineage, we assume that each surname group originates from an initial common ancestor whose traits and characteristics are inherited by current members of the surname group. Thus, correlations within the surname group are determined by the presence of this common ancestor. Although we assume a single common ancestor in the following discussion, we allow for the existence of a limited number of distinct common ancestors to address concerns regarding the origins of surnames linked to professions or geographical proximity. This leads to the first proposition.

Proposition .1 *Common traits within a surname group arise from the existence of a common an-*

cestor, whose traits are inherited by the present generation:

$$E\left[f_{kist}|sur = s\right] = \lambda_k^{t-\tau_s} f_{kis\tau}$$

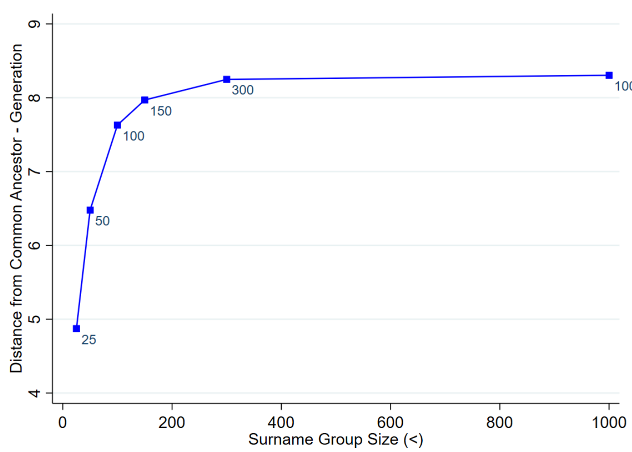
The traits shared within a surname group depend on those of the common ancestor, denoted as $f_{kis\tau_s}$ and taken as given, with τ_s representing the generation of the common ancestor. These traits are then transmitted to future generations in the current generation t according to the transmission parameter λ_k. Hence, the distance to the common ancestor $(t - \tau_s)$ becomes a critical component of our theoretical framework.

However, the relationship between the distance to the common ancestor and surname group size hinges on the nature of the fertility process and a positive probability of surname mutation. Regarding the former, positive population growth, usual in the time-span we consider, implies a positive relationship between group size and distance to the common ancestor. Formally, we assume that the probability of an increase in the distance to the common ancestor monotonically grows with surname group size.

$$P(t - \tau_s = k + \epsilon|n_s) - P(t - \tau_s = k|n_s) < P(t - \tau_s = k + \epsilon|n'_s) - P(t - \tau_s = k|n'_s)$$

Regarding the latter, surname mutations are essential for understanding why surnames still convey socio-economic information and, although rare today, were relatively common in the past. Güell et al. [2015] indicate that in the absence of these mutations, surname distributions would converge to a few very common surnames with low predictive power for individual outcomes. Thus, incorporating this feature in simulations is crucial for replicating the observed surname distribution in Western countries.

FIGURE 3: Average Distance to Common Ancestor as function of Surname Size



Note: The figure depicts the average distance to the common ancestor in a simulated population of 10 generations with growing population and 2% probability of surname mutation. Each data point represents the average distance to the common ancestor, specifically focusing on surnames whose size falls within the range defined by the two preceding points.

This leads us to a second non-testable proposition. We show analytically that relaxing the constraint to small surname groups ($n_s < N$) increases the average distance to the common ancestor.

Proposition .2 *The larger the size of the surname group, the greater the average distance to the common ancestor*

$$\frac{\partial E\left[t - \tau_s | n_s < N\right]}{\partial N} \geq 0$$

This proposition formalizes the intuitive argument presented earlier. Smaller surname groups tend to reflect closer familial ties, thus resulting in a smaller distance to the common ancestor. Conversely, larger surname groups indicate weaker familial connections, as the common ancestor is more distant.

Consequently, surname group size is associated with a variation in the distance to the common ancestor that we can exploit to explore fundamental mechanisms of intergenerational transmission. In particular, this enables testing for the presence of heterogeneity in the persistence rates of distinct characteristics. This concept, **multiplicity**, carries significant implications for understanding the persistence of socio-economic status across multiple generations and diverse families.

From these two propositions, we can deduce two primary implications that can be tested using empirical data. First, we demonstrate that within-group correlation decreases with surname group size.

Corollary .1 *The overall within-group correlation decreases with larger surname group sizes for any given characteristic:*

$$V(E[f_{ist}|sur]|n_s < N_1) > V(E[f_{ist}|sur]|n_s < N_2) \Longleftrightarrow N_1 < N_2$$

The model's explanation of this phenomenon is similar to Güell et al. [2015]. As the size of surnames expands, shared components among group members decrease, mainly due to a larger average distance to the common ancestor. However, Corollary 1 hinges on an additional assumption: the value of the factor for the common ancestor ($f_{kis\tau_s}$) is independent of surname group size (n_s).

This assumption generates two distinct implications. First, it requires no association between the generation of the common ancestor (τ_s) and the characteristic's value ($f_{kis\tau_s}$). In simpler terms, the era in which the common ancestor lived does not indicate the relative position of that ancestor in the distribution for any characteristic. Second, it dismisses any direct influence of surname group size on the factors of the common ancestor. Consequently, families across different surname group sizes can be considered comparable. While previous literature highlighted variations in socioeconomic status across surname frequencies (cite), we provide two reassuring pieces of evidence. First, we concentrate on overall rare surnames, ensuring comparability. For example, we do not compare a rare surname with "Smith" but rather with one that is only slightly less rare. However, we still observe differences in the urban/rural status of the sample across surname group size. In particular, the sample becomes more rural as surname group size increases. To address this issue, we recompute Figure 2 balancing the sample over urban-rural status and obtain a qualitatively similar pattern (Figure 27).

Second, we demonstrate that the standard Intergenerational Elasticity (IGE) does not significantly increase across surname group sizes (Figure 12), whereas the surname-level IGE does. Significant differences across surname group sizes would affect even the parent-child correlation. Therefore, their absence assures us that changes in within-group correlation across surname group sizes stem from variations in the distance to the common ancestor rather than differences in the common ancestors across surname group sizes.

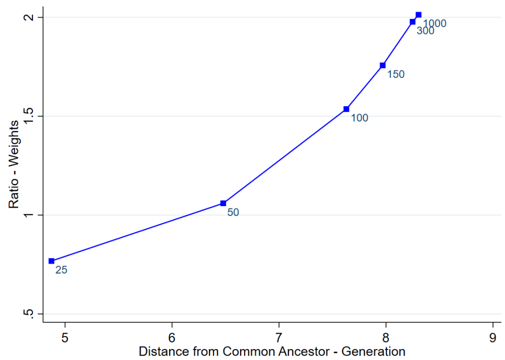
Along with a decrease in the overall size of within-group correlation, we also expect a shift in its composition. More persistent characteristics tend to remain more correlated across generations relative to less persistent ones.

Corollary .2 *As surname group size increases, more persistent characteristics become relatively more important:* $\forall\ N_1 < N_2\ \wedge\ \lambda_k > \lambda_l$

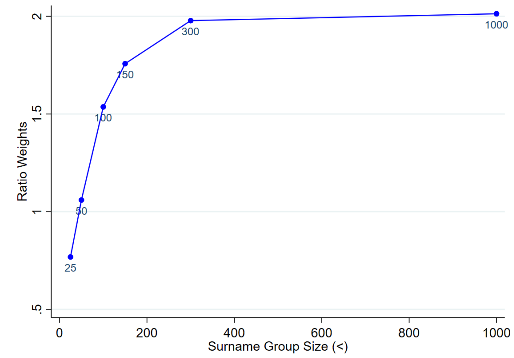
$$V(E[f^k_{ist}|sur]|n_s < N_1) - V(E[f^k_{ist}|sur]|n_s < N_2) < V(E[f^l_{ist}|sur]|n_s < N_1) - V(E[f^l_{ist}|sur]|n_s < N_2)$$

The absolute decrease in within-group correlation for the persistent characteristic k is smaller than the absolute decrease for the less persistent characteristic l as we relax the restriction on surname group size. This suggests a change in the composition of within-group correlation outcomes, with persistent characteristics comprising a larger share of the surname-level average outcome. Finally, these two corollaries can be validated against empirical data by examining the behavior of the ICS and the surname-level IGE. This will help determine whether the model's implications align with observed data. However, this analysis considers population quantities, while the sample variability presents finite-sample noise, which we will address further in the section on the Grouping estimator.

FIGURE 4: Simulation - Weight Ratio



(A) Weight Ratio as function of the Distance to the Common Ancestor



(B) Weight Ratio as function of Surname Group Size

Note: In the left Panel: The figure depicts the ratio of the weight of the persistent factor to the weight of the less persistent one over the average distance to the common ancestor. In the right Panel: The figure depicts the ratio of the weight of the persistent factor to the weight of the less persistent one over surname group size. Each data point represents the weight ratio, specifically focusing on surnames whose size falls within the range defined by the two preceding points. We use a simulated population of 10 generations with growing population and 2% probability of surname mutation. For the MLFM, we assume $\rho_1 = 0.8$ and $\lambda_1 = 0.4$ for the first factor and $\rho_2 = 0.4$ and $\lambda_2 = 0.8$ for the second.

5 ICS - INFORMATIONAL CONTENT OF SURNAMES

5.1 ICS - THEORETICAL IMPLICATION

In this section, we analyze the response of the ICS (Intergenerational Correlation of Status) to variations in surname group size. First, we provide an explanation for the decreasing trend of the ICS in accordance with Corollary 1. Second, we examine the composition of the ICS. According to Corollary 2, persistent factors should increase their relative share of the ICS as surname group size grows.

For this analysis, we focus on a specific interval in surname group size, restricting our sample to surname groups whose sizes fall between N_1 and N_2.

$$Sz(N1, N2) = 1(N_1 < n_s < N_2)$$

We then illustrate the analytical result for the ICS under this model and its assumptions. The ICS can be expressed as follows, with a more detailed proof available in the Appendix:

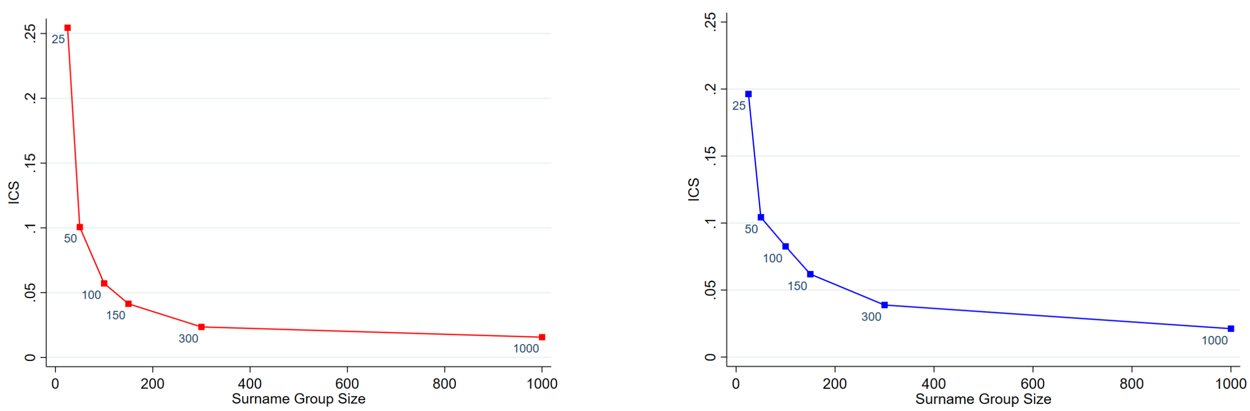
$$\begin{aligned} ICS_{N_1,N_2} &= \frac{V(E_n\left[y_{it}|sur\right]|Sz(N_1,N_2)=1) - V(E_n[y_{it}|fakesur]\left|Sz(N_1,N_2)=1\right)}{V(y_{it}|Sz(N_1,N_2)=1)} \\ &= \frac{\rho_1^2 E\left[\lambda_1^{2(t-\tau_s)}\left(\frac{n_s-1}{n_s}\right)|Sz(N_1,N_2)=1\right] + \rho_2^2 E\left[\lambda_2^{2(t-\tau_s)}\left(\frac{n_s-1}{n_s}\right)|Sz(N_1,N_2)=1\right]}{\rho_1^2+\rho_2^2+V(u_{it})} \end{aligned}$$

The ICS_{N_1,N_2} hinges on several crucial components, particularly the exponent to the transmission parameter λ_k, representing the distance between the current generation t and the generation of the common ancestor τ_s. Nonetheless, to address potential random group effects, we adjust each term by $\left(\frac{n_s-1}{n_s}\right)$, where n_s is the surname group size within the interval $[N_1, N_2]$. This adjustment is larger when the random grouping is finer, as smaller groups tend to carry the information of their members. As surname size increases, this correction term approaches unity, making random grouping almost entirely uninformative.

Consistent with Corollary 1, we expect the ICS_{N_1,N_2} to decrease over surname group size. Larger group sizes are associated with a greater average distance to the common ancestor $t - \tau_s$ (Proposition 2), which results in a weakening of characteristics inherited from the common ancestor. Equation 1 represents the first testable evidence that rationalizes the observed trend depicted in Figure 1:

$$ICS_{N_1,N_2} > ICS_{N_3,N_4} \iff N_1 < N_3 \wedge N_2 < N_4 \tag{1}$$

FIGURE 5: Simulation - Comparison ICS LFM vs HLFM



(A) LFM (B) HLFM

Note: The figure depicts the ICS of the outcome variable in a simulated population of 10 generations with growing population and 2% probability of surname mutation. Each data point represents the ICS, specifically focusing on surnames whose size falls within the range defined by the two preceding points. For the LFM, we assume $\rho = 0.8$ and $\lambda = 0.8$. For the HLFM, we assume $\rho_1 = 0.8$ and $\lambda_1 = 0.4$ for the first factor and $\rho_2 = 0.4$ and $\lambda_2 = 0.95$ for the second.

However, simpler statistical processes of intergenerational transmission generate the same pattern of the overall ICS across surname size. In other words, the introduction of non-linearity in the intergenerational transmission does not yield qualitative differences compared to previous models. The mechanism for the reduction of ICS over surname size remains primarily driven by the distance to the common ancestor. Nonetheless, the presence of multiplicity may be necessary to explain the difference in slope between small and large surname groups. In small groups, the decrease in ICS tends to be relatively steep, while in large groups, the decrease is more gradual.

Corollary 2 produces a testable implication that distinguishes between different mechanisms of intergenerational transmission. Unlike other models, Corollary 2 predicts a change in the composition of the ICS over surname group size. We analyze the ICS by breaking it down based on the trait that generates it:

$$ICS_{N_1,N_2} = \underbrace{\frac{\rho_1^2 E\left[\lambda_1^{2(t-\tau_s)}\left(\frac{n_s-1}{n_s}\right)|Sz(N_1,N_2)=1\right]}{V(y_{it})}}_{=ICS^1_{N_1,N_2}} + \underbrace{\frac{\rho_2^2 E\left[\lambda_2^{2(t-\tau_s)}\left(\frac{n_s-1}{n_s}\right)|Sz(N_1,N_2)=1\right]}{V(y_{it})}}_{=ICS^2_{N_1,N_2}}$$

In small groups, characterized by closer familial connections, we observe a greater contribution of characteristics transmitted from father to child. We argue that within-group correlation comprises a relatively larger proportion of characteristics that are highly relevant but less persistent. Conversely, in large surname groups, the composition of the ICS is primarily influenced by a few persistent factors that maintain a sufficient degree of within-group correlation. As the distance to the common ancestor $(t-\tau_s)$ increases, the numerator shrinks relatively less for more persistent characteristics (larger λ) compared to factors associated with lower λ. For instance, a small surname group may accurately predict both an unobserved personality trait and the state of residence of its members, whereas a larger group may only predict the state of residence. This implies that the share of the ICS generated by persistent characteristics grows with surname group size. Equation 2 provides a testable implication, which can be examined in the data:

$$\frac{ICS^1_{N_1,N_2}}{ICS^2_{N_1,N_2}} < \frac{ICS^1_{N_3,N_4}}{ICS^2_{N_3,N_4}} \iff n_1 < n_3 \wedge n_2 < n_4 \quad \wedge \quad \lambda_1 > \lambda_2 \tag{2}$$

The presence of this pattern represents the first indication of multiplicity in intergenerational transmission. According to the model's prediction, we could compare the persistence of distinct characteristics depending on the steepness of the ICS related to them with surname group size. A steeper decrease would imply a lower λ, while a flatter one would imply a higher one.

5.2 ICS - Testable Implication

In this section, we test the implications on the ICS described in Section 5.1 regarding variations in surname group size as discussed in Section 4.1. We examine how the share of the ICS generated by observable characteristics, such as geography or ethnicity, varies across different surname group sizes. Our findings confirm that environmental characteristics exhibit greater persistence than individual-level traits, providing evidence for the existence of multiplicity in intergenerational transmission.

To begin, we decompose the ICS into two parts: one generated by observable characteristics and the other by individual-specific unobservable traits. This decomposition is performed directly on the observable variables and confirmed using a factor model to capture the primary variation due to group-level characteristics.

Our approach involves several steps. First, we regress the outcome variable on a chosen characteristic, such as geography. Assuming linearity, the fitted values represent the conditional expectation of the outcome variable with respect to the chosen regressor:

$$y_{it}^g = E[y_{ist}|f_{ist}^g] = \rho_g f_{ist}^g$$
$$y_{it}^r = y_{ist} - E[y_{it}|f_{ist}^g] = \rho_r f_{ist}^r$$

The residual f_{ist}^r is a comprehensive factor encompassing any characteristic uncorrelated with the observable characteristic of interest, affecting the outcome variable. Second, we average over surnames each part of the outcome variable:

$$E_n[y_{ist}^g|sur] = \overline{y}_{ist}^g = \rho_g \overline{f}_{ist}^g \qquad\qquad E_n[y_{ist}^r|sur] = \overline{y}_{ist}^r = \rho_r \overline{f}_{ist}^r$$

Finally, we compute the variance of these objects in each interval of surname group size.

$$V(\overline{y}_{ist}^g) = \rho_g^2 V(\overline{f}_{ist}^g) \qquad\qquad V(\overline{y}_{ist}^r) = \rho_r^2 V(\overline{f}_{ist}^r)$$

To compute the ICS, we adjust for potential random grouping effects by repeating these steps on randomly shuffled surnames and subtracting the results from each part of the ICS.

$$E_n[y_{ist}^g|fakesur] = \tilde{y}_{ist}^g = \rho_g \tilde{f}_{ist}^g$$
$$V(\tilde{y}_{ist}^g) = \rho_g^2 V(\tilde{f}_{ist}^g)$$

This allows us to express each component of the ICS as a function of observable characteristics that can be estimated.

$$ICS_{N_1,N_2}^g = \frac{V(\overline{y}_{ist}^g|Sz(N1,N2)=1) - V(\tilde{y}_{ist}^g|Sz(N1,N2)=1)}{V(y_{ist}|Sz(N1,N2)=1)}$$
$$ICS_{N_1,N_2}^r = \frac{V(\overline{y}_{ist}^r|Sz(N1,N2)=1) - V(\tilde{y}_{it}^r|Sz(N1,N2)=1)}{V(y_{ist}|Sz(N1,N2)=1)}$$

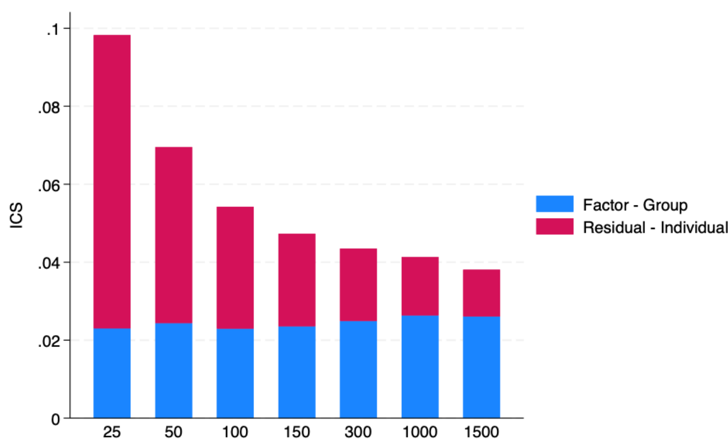
We conduct this analysis using various group-level characteristics such as state of residence, birthplace, ethnicity, and urban/rural status. Additionally, we use factor analysis to capture the common variation among these characteristics, primarily reflecting geographical variation adjusted for other variables (Figure 13). We repeat this exercise with the estimated common factor and with all the characteristics combined.

We test the prediction of changing composition embedded in Equation 2. Beyond confirming that the overall ICS decreases as per Equation 1, we find that, while group-level traits contribute relatively modestly to the total ICS in small groups, they become a predominant component in larger surname groups. Our findings confirm the predictions from the model, as shown in Figures 14-19.

First, we confirm that the overall ICS decreases, consistent with Equation 1. Second, we observe that the composition of the ICS varies with surname group size, in line with Equation 2. Group-level traits contribute more significantly to the ICS in larger surname groups. Additionally, we can compare the persistence rates of these observables. Steeper declines indicate lower persistence rates, while flatter declines suggest higher persistence rates, though the characteristics' relevance might confound the results.

Overall, environmental characteristics tend to show more persistence compared to residual factors, although there are considerable variations among them. For example, the influence of urban/rural status on the ICS declines less compared to residual traits but more than other characteristics like birthplace or state of residence. This suggests that transitions from rural to urban areas might have been more frequent than movements across states. Similarly, while the impact of counties diminishes with surname group size, the effect of states remains consistent, likely because moving across counties is more frequent than moving across states.

FIGURE 6: Decomposition ICS - Factor



Note: The figure depicts the decomposition of the ICS of the Occupational Score in the male working age population in 1940 US Census. Each data point represents the ICS, specifically focusing on surnames whose size falls within the range defined by the two preceding points. We decompose the ICS in a part correlated to a factor capturing most of the variation in group-level characteristics and a residual one.

6 Surname-level IGE

6.1 Surname-level IGE - Theoretical Implication

In this section, we investigate how the Surname-level Intergenerational Elasticity (IGE) responds to changes in surname group size. First, we demonstrate that the second Corollary provides a coherent explanation for the statistical trend discussed in Section 3.2. As surname group size increases, the surname-level IGE increasingly reflects persistent traits, explaining the overall increasing trend. Existing studies suggest that this phenomenon stems from surnames capturing aggregate-level mobility rather than individual-level mobility (Santavirta and Stuhler [2020]; Chetty et al. [2014]; Güell et al. [2015]). We formalize these insights by illustrating that this is due to the greater persistence of environmental characteristics.

Second, we propose a novel testable implication of Corollary 2 related to surname-based estimation. The model predicts that the Grouping estimator's weights assigned to persistent characteristics should rise alongside surname group size. Our findings demonstrate that as surname group size expands, surname-based estimators increasingly incorporate persistent factors. Furthermore, we provide evidence suggesting that environmental characteristics likely contribute to the factors increasingly captured by the grouping estimator

Initially, we rationalize the increasing pattern of the surname-based estimator across surname group size. Intuitively, surname-based estimates capture the persistence of factors, weighting them by their degree of within-group correlation. As discussed in Section 4.1, this correlation arises from shared ancestry, i.e., having a common ancestor (Proposition 1). However, larger surname groups entail weaker familial connections compared to smaller ones (Proposition 2). Consequently, only highly persistent factors maintain correlation within the surname group, while less persistent ones average out (Corollary 2). Therefore, estimates increasingly reflect the persistence of enduring factors

Formally, we document the value of persistence implied by the model for each surname size. We begin with a sample limited to very small surname groups and gradually relax this constraint to ensure smoothness in the estimation pattern. The model-implied formula for the Grouping estimator is expressed as:

$$\begin{aligned}
\beta_N^G &= \frac{Cov(E_n\left[y_{ist}|sur\right], E_n\left[y_{ist-1}|sur\right]|n_s < N)}{V(E_n\left[y_{ist-1}|sur\right]|n_s < N)} \\
&= \frac{\lambda_1\rho_1^2 V(\overline{f}_{1ist-1}|n_s < N) + \lambda_2\rho_2^2 V(\overline{f}_{2ist-1}|n_s < N))}{\rho_1^2 V(\overline{f}_{1ist-1}|n_s < N) + \rho_2^2 V(\overline{f}_{2ist-1}|n_s < N) + V(\overline{u}_{it-1}|n_s < N)}
\end{aligned}$$

For readability, we denote the surname sample average with an overline. The variation of $\beta^G N$ over surname size can occur through two distinct channels. First, the variability of the stochastic component decreases with increasing surname size. Assuming that the errors are iid, we know that:

$$V(\overline{u}_{it-1}|n_s < N)) = E\left[\frac{V(u_{it-1})}{n_s}|n_s < N\right]$$

where n_s is the surname group size, given that N is the maximum surname group size. A larger N decreases the average inverse of surname size $E\left[\frac{1}{n_s}|n_s < N\right]$, which overall decreases this term. This means that market luck is diluted when averaging over larger surname groups. This mechanism is also at play in the simple Latent Factor Model and follows the original intuition from Clark [2014].

Second, even abstracting from finite sample issues, multiplicity still generates an increasing pattern of the Grouping estimator over surname group size. Assuming knowledge of the conditional expectation of any characteristic with respect to the surname, multiplicity in intergenerational transmission may underlie this pattern. First, we demonstrate that any intergenerational estimator in this model represents a differently weighted average of the persistence of various factors. In particular, the weighting structure of the Grouping estimator hinges on the relevance of the factor (ρ_k) and its degree of between variability ($V(E[f_{kist-1}|sur])$).

$$\beta_N^G = \lambda_1\omega_{1N} + \lambda_2\omega_{2N}$$

where:

$$\omega_{1N} = \frac{\rho_1^2 V(E[f_{1ist-1}|sur]|n_s < N)}{\rho_1^2 V(E[f_{1ist-1}|sur]|n_s < N) + \rho_2^2 V(E[f_{2ist-1}|sur]|n_s < N)}$$

From the analytical expression of the weights, we observe that, even in the absence of finite sample concerns, the between variability of the factor depends on the surname group size. To investigate this mechanism further, we represent the estimator as a function of the weight of the more persistent factor. Without loss of generality, let's assume that $\lambda_2 > \lambda_1$, allowing us to rewrite the estimator as:

$$\beta_N^G = \lambda_1 + (\lambda_2 - \lambda_1)\omega_{2N}$$

As a result, we can establish a direct relationship between the magnitude of the IG elasticity at the surname level and the degree to which it weights persistent characteristics:

$$\frac{\partial \beta_N^G}{\partial \omega_{2N}} \geq 0$$

Therefore, to comprehend how multiplicity influences the pattern of the grouping estimator, we need to examine how the weighting structure responds to variations in surname group size. By Proposition 2, a larger surname group size implies an increase in the distance to the common ancestor, leading to an uneven impact on the between variability of factors with different transmission rates. By Corollary 2, a characteristic with a higher persistence rate will exhibit a smaller fall in the between variability compared to a low persistence rate characteristic. It follows that, as surname group size increases, the weighting ratio shifts in favor of the more persistent characteristic:

$$\omega_{2N_2} - \omega_{2N_1} \geq 0 \quad \Leftrightarrow \quad \lambda_2 \geq \lambda_1 \ \wedge \ N_2 \geq N_1 \tag{3}$$

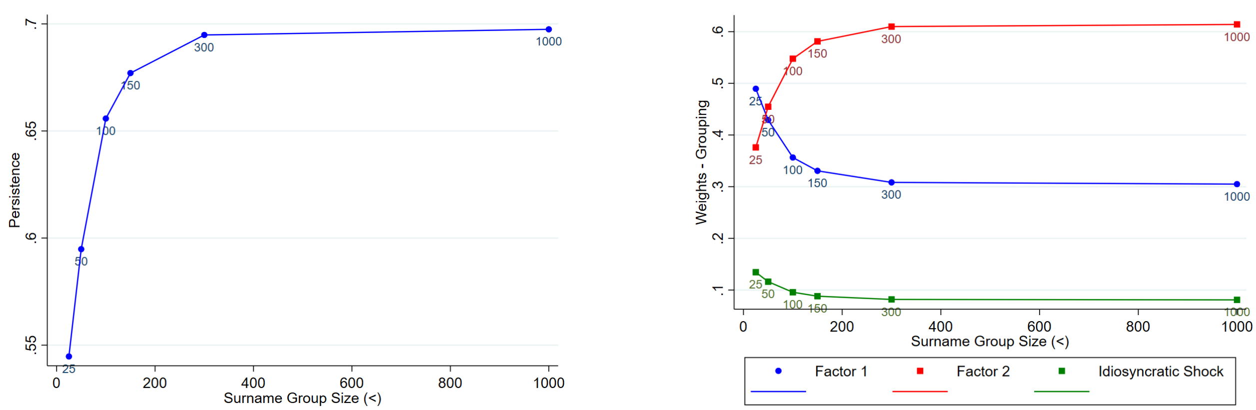
While this explanation holds true in the population, simulations demonstrate that finite sample disturbances do not qualitatively alter the result.

This result yields two testable implications. First, it represents the mechanism driving the growth of the estimator with increasing surname size. Summarizing, larger surname size implies a larger average distance to the common ancestor, i.e., weaker familial connections. Consequently, the weakening of these connections results in a shift in the weighting of the estimator. Specifically, the weight of the persistent factor increases relative to the less persistent factor, generating the empirical pattern in Figure 2. Equation 4 thus represents a testable and validated implication of the model.

$$\frac{\partial \beta_N^G}{\partial N} \geq 0 \tag{4}$$

Second, it serves as additional suggestive evidence for the existence of multiple channels in intergenerational transmission. We can directly test Equation 3 by studying whether the weight associated with distinct characteristics increases over surname group size.

FIGURE 7: Simulation - Persistence over Surname Size



(A) Persistence

(B) Decomposition of the Estimator

Note: In the left Panel: The figure illustrates the Persistence of Occupational Score at the surname level. We calculate both the Grouping estimator and the JIVE estimator on surname groups smaller than any specified size. In the right Panel: The figure depicts the decomposition of the grouping estimator on surname groups smaller than any specified size. We use a simulated population of 10 generations with growing population and 2% probability of surname mutation. For the HLFM, we assume $\rho_1 = 0.8$ and $\lambda_1 = 0.4$ for the first factor and $\rho_2 = 0.4$ and $\lambda_2 = 0.8$ for the second.

However, weights are a function of two distinct parameters: relevance ρ and persistence λ. To isolate the latter, we normalize the weight for characteristic g based on the weight on very small surnames $N_1 < N$. By doing so, we eliminate the relevance parameter, making it a function solely of the transmission parameter:

$$\frac{\omega_{g,N}}{\omega_{g,N_1}} = \underbrace{\frac{V(E[f_{gist}|sur]|n_s < N)}{V(E[f_{gist}|sur]|n_s < N_1)}}_{=h(\lambda_g)} \frac{V(E[y_{gist}|sur]|n_s < N_1)}{V(E[y_{gist}|sur]|n_s < N)}$$

By comparing these ratios for different characteristics, we can directly compare the rate of persistence of distinct observables while abstracting from considerations of relevance.

$$\frac{\omega_{g,N}}{\omega_{g,N_1}} \geq \frac{\omega_{r,N}}{\omega_{r,N_1}} \quad \Leftrightarrow \quad \lambda_g \geq \lambda_r \tag{5}$$

The empirical analogue of Equation 5 allows us to gain insights into the persistence of various observable characteristics, corroborating the evidence from previous sections. We expect the percentage change in the ratio of the weights for all observables to grow over surname group size. Furthermore, such growth should be particularly pronounced for persistent characteristics such as state of residence or race.

6.2 Surname-level IGE - Testable Implication

In this section, we test the model predictions elaborated earlier. First, we estimate the weights for various environmental characteristics. We find that these weights increase with surname group size, indicating greater persistence than average, consistent with previous empirical findings. We also compare the persistence rates of these characteristics by plotting the normalized growth of weights over surname size. Consistent with evidence from the ICS, race and geographical location demonstrate the highest levels of persistence, while urban/rural status, for instance, shows lower persistence. Residual characteristics are even less persistent.

To explore the weighting structure in the data and assess Equation 3, we follow the methodology outlined in Section 5.2. We decompose the surname-level average of the outcome variable into two components: $\overline{y}^g_{ist}$, representing the part correlated to the observable characteristic, and $\overline{y}^2_{ist}$, representing the uncorrelated component. These components are defined in terms of the model as follows:

$$\overline{y}^g_{ist} = \rho_g \overline{f}_{gist} \qquad\qquad \overline{y}^r_{ist} = \rho_r \overline{f}_{rist}$$

To replicate the weights used by the Grouping estimator, we compute the variance of each component and normalize it by the variance of the overall surname-level average of the outcome variable:

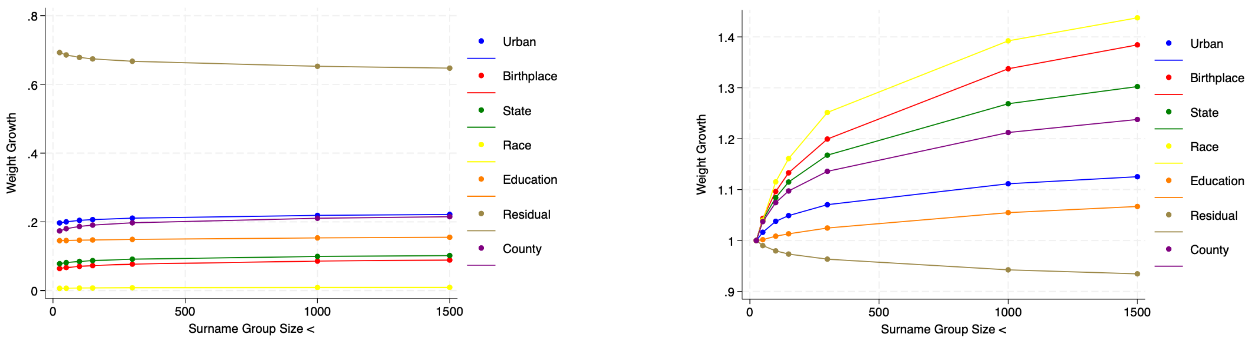
$$\hat{\omega}_{g,N} = \frac{\rho_g^2 V(\overline{y}^g_{ist}|n_s < N)}{V(\overline{y}_{ist}|n_s < N)} \qquad\qquad \hat{\omega}_{r,N} = \frac{\rho_r^2 V(\overline{y}^r_{ist}|n_s < N)}{V(\overline{y}_{ist}|n_s < N)}$$

We evaluate Equation 3 by plotting the evolution of these weights over surname size for the previous set of characteristics. However, as discussed in Section 6.1, interpreting weight patterns solely as reflections of persistence can be misleading due to the relevance parameter. For instance, urban/rural status shows a significant increase over surname group size, contrary to ICS decomposition evidence, but its high relevance may confound our interpretation.

To isolate persistence from relevance, we normalize these weights relative to the first surname group size interval, as in Equation 5:

$$\frac{\hat{\omega}_{g,N}}{\hat{\omega}_{g,25}} = \frac{V(\overline{f}_{gist}|n_s < N)}{V(\overline{f}_{gist}|n_s < 25)} \frac{V(\overline{y}_{ist}|n_s < 25)}{V(\overline{y}_{ist}|n_s < N)}$$

FIGURE 8: Weight's Growth over surname size



(A) Non-Normalized

(B) Normalized

Note: In the left Panel: The figure illustrates the evolution of the weight of every single characteristic as surname size grows. We calculate the weight for any characteristic on surname groups smaller than any specified size at the **child** generation. In the right Panel: We divide the weights for each characteristic with respect to the weight in the first surname group size interval. Our analysis focuses exclusively on male workers aged between 30 and 40 in the 1940 US Census, linked to their fathers.

We plot the weights' growth for all characteristics in Figure 8, resolving the apparent contradiction between ICS decomposition results and non-normalized weights. From the left panel, we gather two crucial pieces of information. First, we assess the magnitude, noting that non-observed characteristics explain most of the variability in the occupational score. Among observables, urban/rural status and county are most relevant, as urban or rural residence significantly shapes occupation. In rural areas, many people are in farming, and we cannot differentiate among them, so we impute the median to all. This likely overstates the importance of urban status for socioeconomic outcomes. We test this prediction by replicating the analysis using log-wages rather than occupational scores, as this information is available in the cross-section. In Figure 26, we confirm that urban status is less relevant for wages than for occupational variables. Furthermore, we find similar persistence patterns across surname group size, validating the overall methodology. Conversely, race is least relevant due to the sample's ethnic homogeneity, and geographical characteristics like state of residence or birthplace have moderate importance.

Second, we observe trends in weights over surname group size. All observed variables, from urban status to race, show a growing trend, suggesting they are more persistent than non-observed characteristics, which show a decreasing trend. This result is intuitive, as offspring are more likely to live in the same area as their parents than to inherit traits like personality. However, comparing the absolute change in weights across surname group size solely as a function of persistence is misleading, given that these changes are scaled by relevance.

The right panel's normalization reveals a clear order of persistence among observable characteristics. Race and geographical characteristics show the highest persistence, while urban/rural status, despite being more relevant, exhibits less. For example, a father in rural California is more likely to have offspring remain in California than remain in a rural area. Moreover, it shows that environmental factors are generally more persistent than individual characteristics. For instance, education, though correlated with environmental factors, demonstrates lower persistence. Orthogonal characteristics to these environmental factors are even less persistent and thus decrease in importance across surname group size.

7 Conditional Surname-level IGE

In previous sections, we identified two channels through which Grouping estimates increase with surname group size: finite-sample noise and multiplicity in intergenerational transmission. Our analysis of weights provides evidence of the latter. This analysis also offers insights into the parameters associated with each observable characteristic, specifically relevance (ρ) and transmission (λ). For instance, the weights' analysis indicates that ethnicity and geographical characteristics are highly persistent (high λ) but have relatively low relevance (low ρ). In this section, we examine how surname-level IGE responds when controlling for covariates, serving a dual purpose. First, we demonstrate that finite sample noise alone cannot explain the observed pattern, indicating the necessity of multiplicity. Differentiating between these channels is crucial as it allows us to distinguish between a simple Latent Factor Model (LFM) (Clark [2014]) and a process involving multiplicity. Second, we validate the results from the weights' analysis, showing that the estimator behaves consistently with the parameters associated with each characteristic. Thus, we simultaneously validate the previous analysis and differentiate between two key statistical processes in intergenerational transmission. We conclude that the LFM alone does not adequately explain surname-based evidence or the broader intergenerational process.

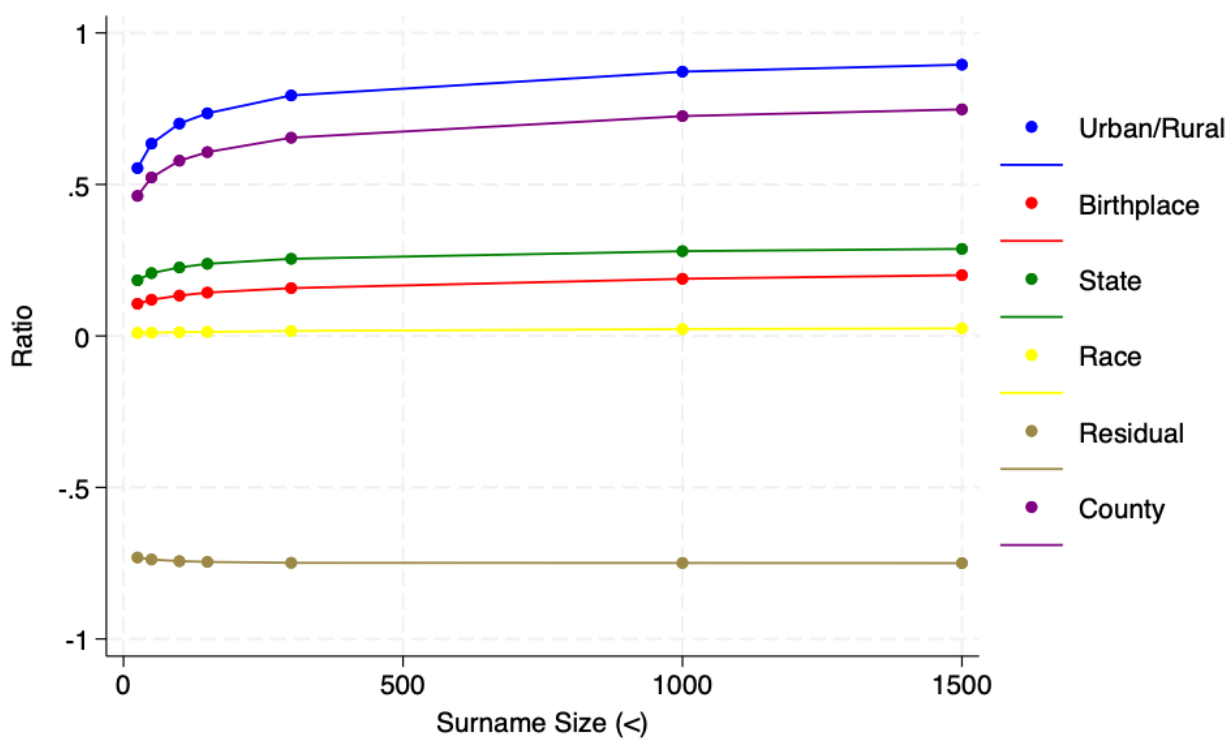
To this end, we use the ratio between the controlled surname-level IGE and the standard surname-level IGE in the sample. For the analytical derivation, we include three factors for greater flexibility. However, we can treat the third factor as an error term by setting the persistence parameter to zero and the relevance parameter to one.

$$R^*_{n,g} = \left(\frac{1}{V(\overline{y}_{it-1}|n_s < N)}\right)\left(\frac{(\lambda_1 - \lambda_2)V_{1n}V_{2n} + (\lambda_1 - \lambda_3)V_{1n}V_{3n}}{\lambda_2 V_{2n} + \lambda_3 V_{3n}}\right)$$

where $V_j = \rho_j^2 V(\overline{f}_{jit-1}|n_s < N)$ is a measure of the relevance of characteristics j. The sign and magnitude of this ratio are both of interest. On the one hand, the sign depends only on the persistence of the characteristic being controlled for, as the denominator is always positive. We expect the ratio to be positive when controlling for persistent characteristics and negative otherwise. The intuition underlying this result is that the conditional estimate captures the persistence rate of characteristics orthogonal to the controlled one. If we control for something persistent, the remaining factors will be less persistent, thus decreasing the conditional estimmate and viceversa. From our analysis of the weights, we anticipate positive ratios for all observable characteristics except for the residual, which we expect to be negative.

On the other, the magnitude of the ratio is influenced by both the persistence ($\lambda_1 - \lambda_2$) and the relevance of the controlled characteristic (V_{1n}), with relevance likely playing a larger role. For instance, race is highly persistent but not very relevant in our sample, so we expect little difference between conditional and unconditional estimates. Conversely, urban status, while less persistent, is highly relevant, leading to a larger difference between estimates.

FIGURE 9: Ratio over surname size - Occupational Score



Note: In the left Panel: The figure illustrates the evolution of the ratio between the unconditional and the conditional surname-level IGE for every single characteristic as surname size grows. We restrict the sample on surname groups smaller than any specified size at the **father** generation. Our analysis focuses exclusively on male workers aged between 30 and 40 in the 1940 US Census, linked to their fathers.

We observe that the behavior of each characteristic aligns with our theoretical predictions and the parameters inferred from the weight analysis. Race, persistent but not very relevant in our sample, shows a positive but small ratio. Birthplace and state of residence, both persistent and moderately relevant, yield a positive and significant ratio. County and urban status, though not extremely persistent, are highly relevant, resulting in a large ratio. Conversely, the residual is highly relevant but less persistent, leading to a large negative ratio.

Additionally, this ratio helps differentiate the influence of finite-sample noise and multiplicity. We derive the model-implied formula for the LFM and show that it cannot explain the empirical pattern. Assuming a model with two independent factors with the same transmission parameter (λ) and an error term representing finite-sample attenuation bias:

$$R^*_{n,g} = \frac{\beta_n}{\beta_{n,1}} - 1 = \frac{V_1 V_u}{V_1 V_2 + V_2^2 + V_2 V_u} > 0$$

This ratio is always positive due to stronger attenuation bias in the conditional estimate, which uses smaller variability than the unconditional one, amplifying the bias's relative impact. However, the residual generates a negative ratio, contradicting the model. Moreover, according to Corollary 2, market luck's relevance (V_u) should decrease faster with increasing surname group size than the relevance of transmittable factors. This would imply the ratio converging to zero, which we do not observe. Instead, the persistent positive ratios suggest that while finite-group disturbances exist, multiplicity in intergenerational transmission is necessary to explain the patterns.

In conclusion, analyzing how the estimator responds to controlling for various characteristics achieves two goals. First, it validates our weight results, showing that the grouping estimator weights characteristics as predicted by theory. Second, it provides clear evidence that the LFM alone cannot explain the empirical patterns, confirming the presence of multiplicity in intergenerational transmission. This mechanism's implications extend beyond surname-based estimation, indicating that correlations between distant relatives are driven by different characteristics than those between close relatives. For instance, geography tends to be more persistent and hence more relevant in explaining long-term intergenerational mobility. This highlights the importance of a comprehensive assessment of social mobility that considers both group and individual-level attributes.

8 Implications

These findings have implications that extend beyond the sole rationalization of surname-based evidence. We identify the multiplicity of mechanisms as a crucial feature of the nature of intergenerational transmission. Multiple studies (cite) in the literature have addressed the attenuation bias in standard estimates of intergenerational transmission by postulating the presence of a latent factor with higher persistence than observable measures of socioeconomic outcomes. While the latent factor model remains a useful statistical tool for quantifying intergenerational transmission, its interpretation is complex. It is challenging to define the latent factor and to intuitively understand its economic significance.

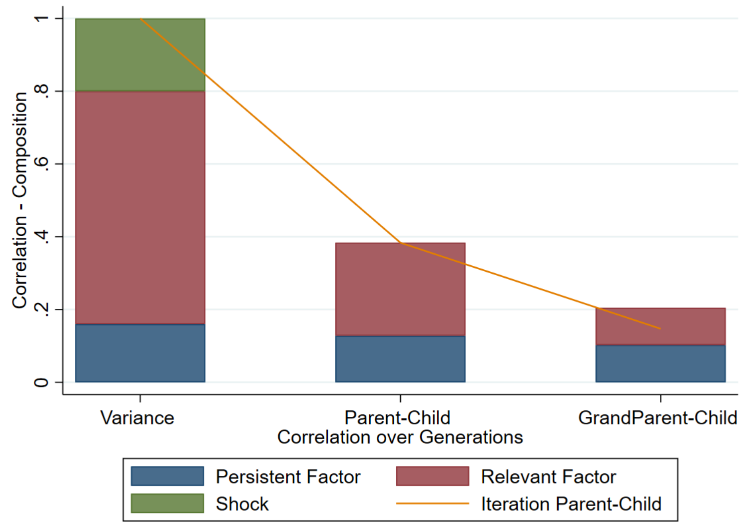
Through the analysis of surname-based evidence, we offer an explanation that makes estimates from other studies more concrete. Instead of relying on a vague concept of a socio-economic factor, we return to a straightforward framework where socio-economic outcomes are the function of multiple different characteristics, both observable and unobservable. The presence of distinct persistence rates across these characteristics generates non-linearities, the neglect of which can lead to underestimation of intergenerational transmission.

We discuss two cases where non-linearities play a clear role. First, in the literature on multigenerational inequality, the decrease in intergenerational correlation slows across multiple generations as more persistent characteristics drive the correlation between distant relatives more than between close relatives. In this scenario, the implications of the single latent factor model are not easily distinguishable from the multiple factor model. However, while the single factor model obscures many potential mechanisms, the multiple factor model provides a much clearer interpretation. Second, different degrees of intergenerational persistence can be observed across families, depending on whether the socio-economic outcome is generated by more or less persistent sources.

8.1 1 - Multigenerational Inequality

As anticipated, multiplicity provides a straightforward explanation for the evidence from multi-generational studies. The observed excess persistence across multiple generations results from the non-linearity introduced by distinct persistence parameters. This causes a shift in the composition of multigenerational correlation: close relatives exhibit correlations primarily driven by relevant but fading factors, while distant relatives are influenced by more persistent characteristics. Consequently, the decline in intergenerational correlations diminishes as the distance between generations increases. For instance, in the context of surname-based evidence, environmental factors play an increasingly important role in shaping within-group correlation as surname group size increases.

FIGURE 10: Simulation - Correlation over Generations - Composition



Note: This figure depicts the decomposition of the variability of the outcome variable. Then, we decompose the parent-child correlation and the grand-parent child correlation. We use a simulated population of 10 generations and 2% probability of surname mutation. For the HLFM, we assume $\rho_1 = 0.8$ and $\lambda_1 = 0.4$ for the first factor and $\rho_2 = 0.4$ and $\lambda_2 = 0.8$ for the second.

The Multiple Latent Factor Model (MLFM) both complements and contrasts with alternative theories of multigenerational transmission. While our findings confirm that measuring intergenerational mobility through parent-child correlation is unsuitable for analyzing long-term dynamics, we deviate from the simple Latent Factor Model (LFM). According to the LFM, transmission occurs over a single construct of socio-economic status, of which any socioeconomic outcome is an imprecise proxy. This framework is useful for quantification and highlights the presence of measurement error in both surname-based estimates and broader intergenerational persistence. However, it is agnostic about features of the transmission process, such as multiplicity. In the context of surname-based estimates, the LFM provides inadequate implications as it cannot explain the variation in the Grouping estimator over surname size, attributing it solely to attenuation due to finite-sample noise, which we prove incorrect.

Moreover, the MLFM challenges the notion that direct effects from distant relatives on child outcomes uniquely cause excess persistence. The theory suggesting causality as the sole basis for correlations between distant relatives and child outcomes (Mare [2011]) is inconsistent with surname-based evidence. While we cannot entirely dismiss second-order effects from relatives other than parents on child outcomes, we can reject a pure AR(2) model where the grandparental outcome is the sole omitted variable underlying the measurement error of parent-child correlation. The increasing pattern of surname-based estimates over surname size cannot be adequately described by the stationary version of this model. Furthermore, this perspective overlooks the distinction in the nature of correlations between close and distant relatives.

In summary, the analysis of surname-based evidence provides two main insights. First, multiplicity implies that the intergenerational transmission process is non-linear. Initially, intergenerational correlations decrease rapidly due to the relevance of fading characteristics, but over time, they stabilize as persistent factors become predominant. Failure to account for this nonlinearity in the transmission process results in the observed excess persistence in distant relatives.

Second, a thorough analysis of intergenerational transmission should consider both aggregate and individual-level mobility. Group membership and environmental factors transmitted from parents to children are likely key contributors to this nonlinearity. Given their high persistence over generations, these factors play an even larger role in explaining long-term social mobility. For example, in the surname setting, geographical location is crucial for understanding the pattern of intergenerational estimates. However, these factors are not the only contributors to the nonlinearities in the transmission process. Wealth, for instance, is a characteristic that can generate this non-linear pattern both over time and across individuals.

8.2 2 - Heterogeneity across families

The relevance of the Multiplicity Latent Factor Model (MLFM) extends beyond its ability to explain multigenerational evidence. The existence of multiple channels in intergenerational transmission implies inherent heterogeneity across different families, a phenomenon not accounted for by the alternative data-generating processes considered.

To illustrate this, we derive a naive measure of intergenerational persistence within a given family s. It's important to note that this is a theoretical construct, and its estimation carries significant uncertainty.

$$\phi_s = E\left[\left(\frac{y_{ist} - y_{ist-1}}{y_{ist-1}}\right)^2 |f_{1ist-1}, f_{2ist-1}\right] = \kappa + (\lambda_1 - 1)^2\omega_1 + (\lambda_2 - 1)^2\omega_2$$

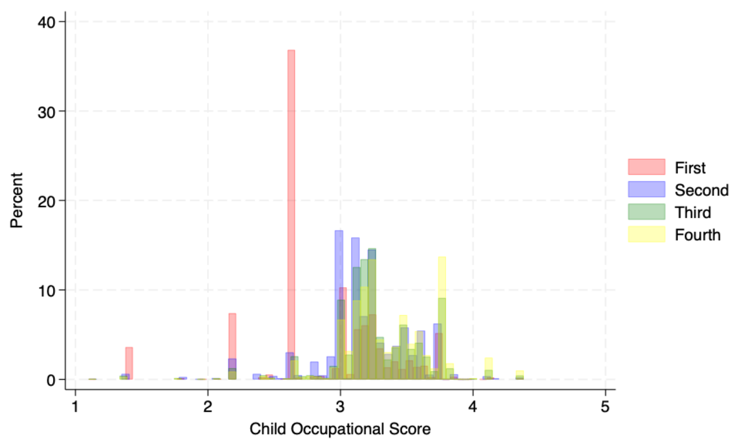
This measure represents the squared expected relative deviation of the child's outcome from the father's. The smaller this deviation, the greater the persistence of the outcome variable. It shows that intergenerational persistence is a function of the relevance of each factor in determining the outcome, even if the outcome level remains the same. If a persistent factor contributes significantly to the outcome, we should expect a lower deviation. Hence, intergenerational transmission will be stronger in families where persistent factors are more influential. Given that the weights sum to one, we can rearrange and take the derivative with respect to the persistent factor:

$$\frac{\partial \phi_s}{\partial \omega_2} = (\lambda_2 - 1)^2 - (\lambda_1 - 1)^2 < 0 \iff \lambda_2 > \lambda_1$$

This reasoning formalizes existing views on social mobility. For instance, consider two families with fathers at the same low income level: one is a white family living in a dynamic state, and the other is an African-American family living in a depressed state. According to the theory, we should expect higher upward mobility in the former, where the low income is less influenced by persistent characteristics like race and geography. This aligns with Chetty et al. [2014]'s conclusions on the importance of place-based policies targeting low mobility areas.

However, assuming independence and normality in the factors limits the ability to explain varying persistence rates across the outcome distribution. Relaxing these assumptions would reveal variability in persistence across the distribution. For example, if a persistent factor like wealth is concentrated at the top of the income distribution, children from these families would deviate little from their father's socioeconomic status. In Figure 11, we document this heterogeneity, noting high persistence at the bottom of the outcome distribution.

FIGURE 11: Histogram according to Father position



Note: This figure depicts the distribution of the Occupational Score in 1940 for the child generation according to the quartile of the father's Occupational Score in 1920.

Although the estimator is sensitive to covariance among factors, examining this heterogeneity remains relevant. First, it provides additional evidence of multiplicity. For example, this model explains Clark's findings, extending insights from Chetty et al. [2014] and Güell et al. [2015]. Clark [2014] identified exceptionally high persistence at the surname level among rare surnames from distinct socioeconomic strata, which Chetty et al. [2014] could not replicate in their data. We suggest that Clark's results may be due to averaging over families whose outcomes were driven by highly persistent characteristics, capturing the persistence of very specific factors.

Second, the model provides an intuitive interpretation for high intergenerational transmission at the extremes of the outcome distribution. The presence of "Poverty Traps" or "Golden Cages" is of significant interest from both normative and policy perspectives.

9 CONCLUDING REMARKS

REFERENCES

Adrian Adermon, Mikael Lindahl, and Mårten Palme. Dynastic human capital, inequality, and intergenerational mobility. *American Economic Review*, 111(5):1523–48, 2021.

Joshua D Angrist. The perils of peer effects. *Labour Economics*, 30:98–108, 2014.

Sebastian Till Braun and Jan Stuhler. The transmission of inequality across multiple generations: testing recent theories with evidence from germany. *The Economic Journal*, 128(609):576–611, 2018.

Raj Chetty, 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.

Gregory Clark. The son also rises. In *The Son Also Rises*. Princeton University Press, 2014.

M Dolores Collado, Ignacio Ortuño-Ortín, and Jan Stuhler. Estimating intergenerational and assortative processes in extended family data. *The Review of Economic Studies*, 90(3):1195–1227, 2023.

John Conlisk. Can equalization of opportunity reduce social mobility? *The American Economic Review*, pages 80–90, 1974.

Arthur S Goldberger. Economic and mechanical models of intergenerational transmission. *The American Economic Review*, 79(3):504–513, 1989.

Maia Güell, José V Rodríguez Mora, and Christopher I Telmer. The informational content of surnames, the evolution of intergenerational mobility, and assortative mating. *The Review of Economic Studies*, 82(2):693–735, 2015.

Maia Güell, Michele Pellizzari, Giovanni Pica, and José V Rodríguez Mora. Correlating social mobility and economic outcomes. *The Economic Journal*, 128(612):F353–F403, 2018.

Peter Hull. Examiner designs and first-stage f statistics: A caution. *Unpublished Working Paper*, page 5, 2017.

Mikael Lindahl, Mårten Palme, Sofia Sandgren Massih, and Anna Sjögren. Long-term intergenerational persistence of human capital an empirical analysis of four generations. *Journal of Human Resources*, 50(1):1–33, 2015.

Robert D Mare. A multigenerational view of inequality. *Demography*, 48(1):1–23, 2011.

Torsten Santavirta and Jan Stuhler. Name-based estimators of intergenerational mobility. *Technical Report*, 2020.

Florencia Torche and Alejandro Corvalan. Estimating intergenerational mobility with grouped data: A critique of clark's the son also rises. *Sociological Methods & Research*, 47(4):787–811, 2018.

10 Appendix

10.1 Images

FIGURE 12: Parent-Child Correlation

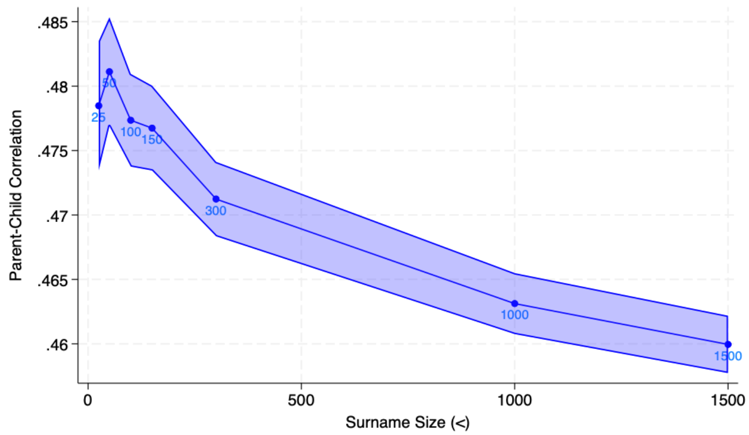


FIGURE 13: Correlation of Factor Scores

FIGURE 14: Decomposition ICS - State of Residence

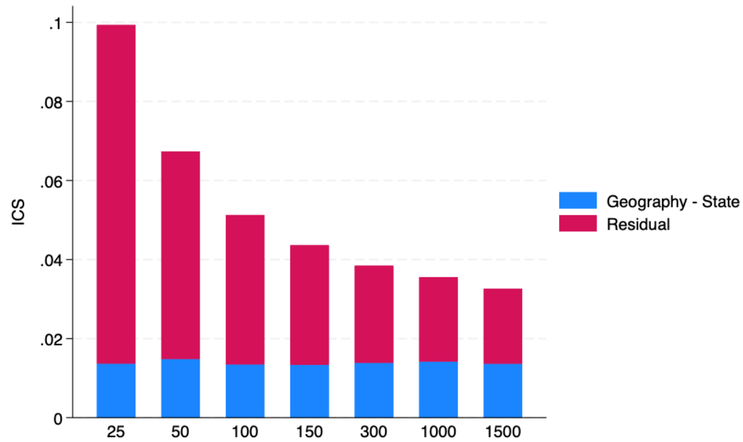


FIGURE 15: Decomposition ICS - Urban/Rural Status

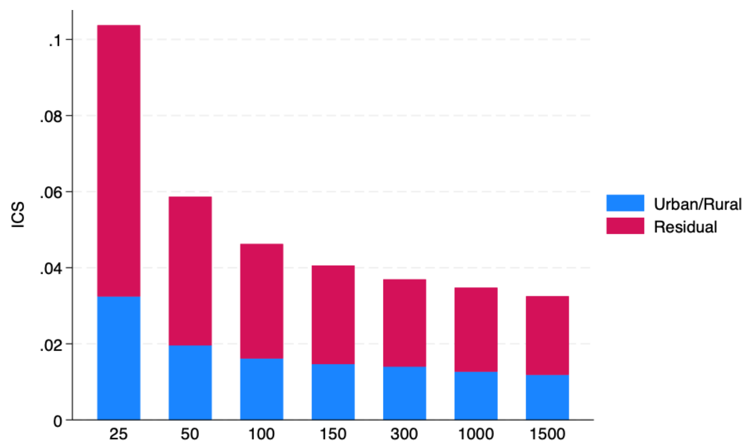


FIGURE 16: Decomposition ICS - Race

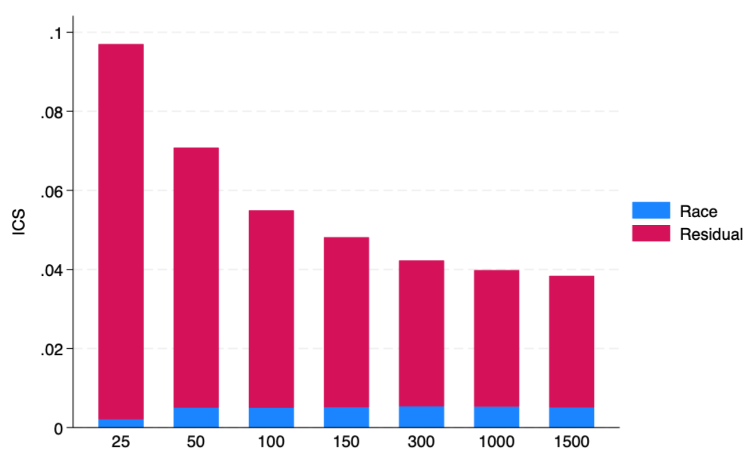


FIGURE 17: Decomposition ICS - Factor

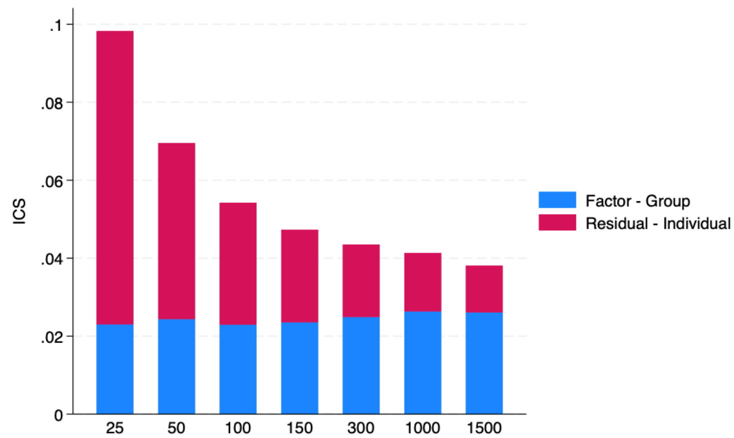


FIGURE 18: Decomposition ICS - All group-level characteristics

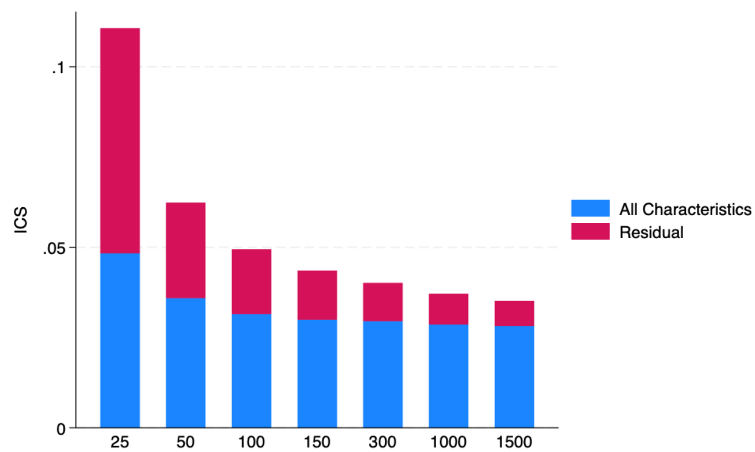


FIGURE 19: Decomposition ICS - Birthplace

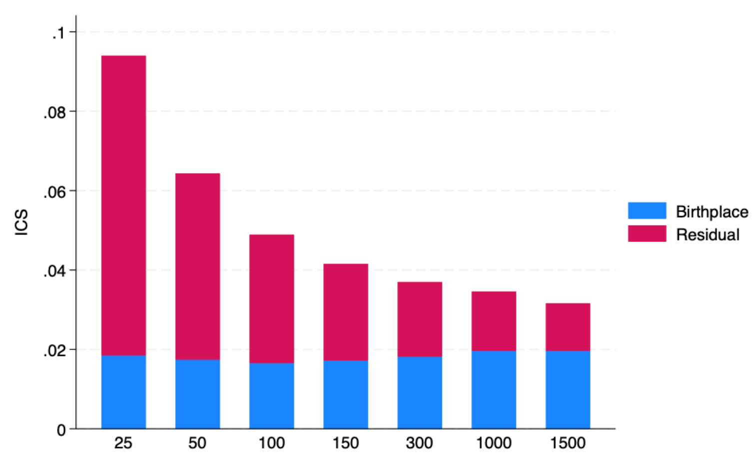


FIGURE 20: Weight - Birthplace

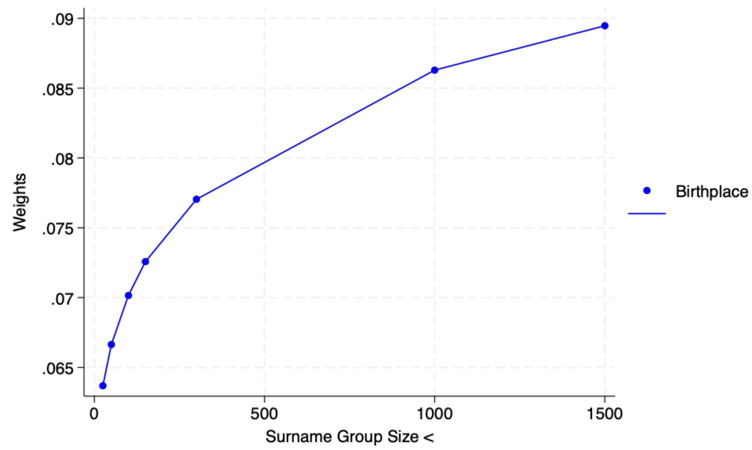


FIGURE 21: Weight - Race

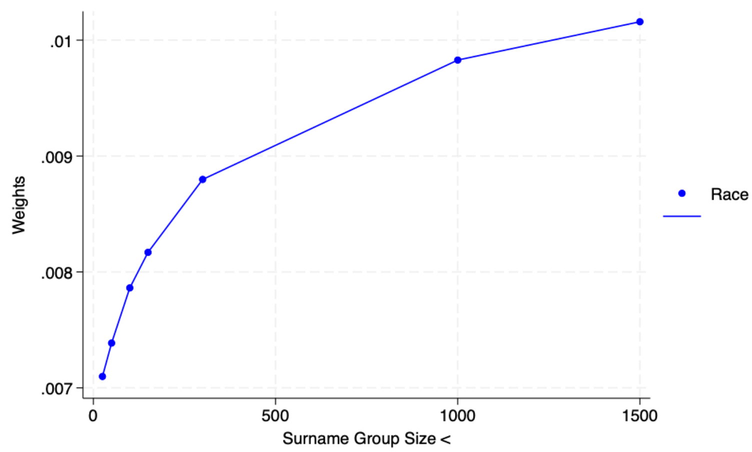


FIGURE 22: Weight - Urban/Rural Status

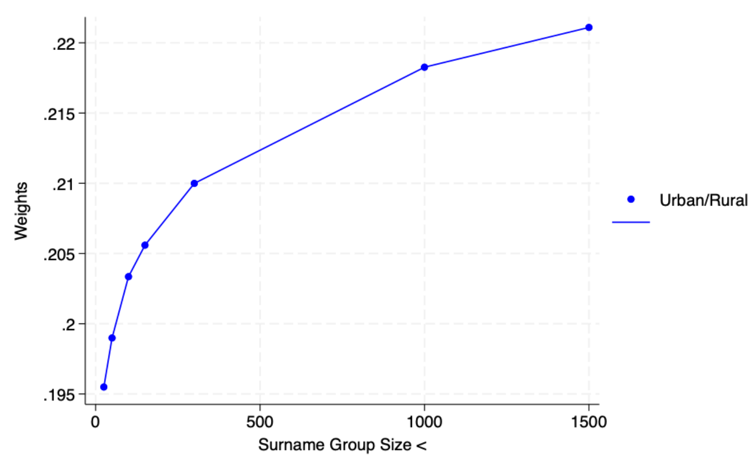


FIGURE 23: Weight - State of Residence

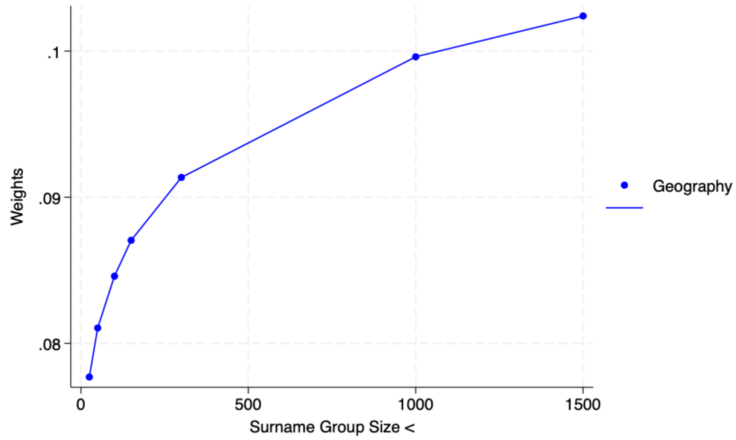


FIGURE 24: Weight - Factor

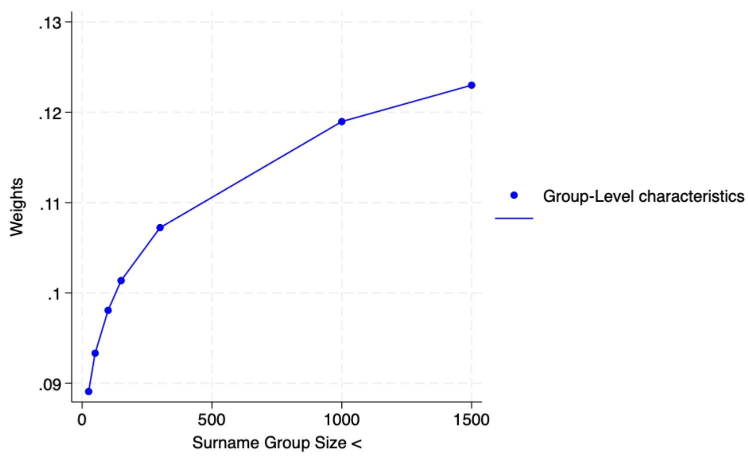
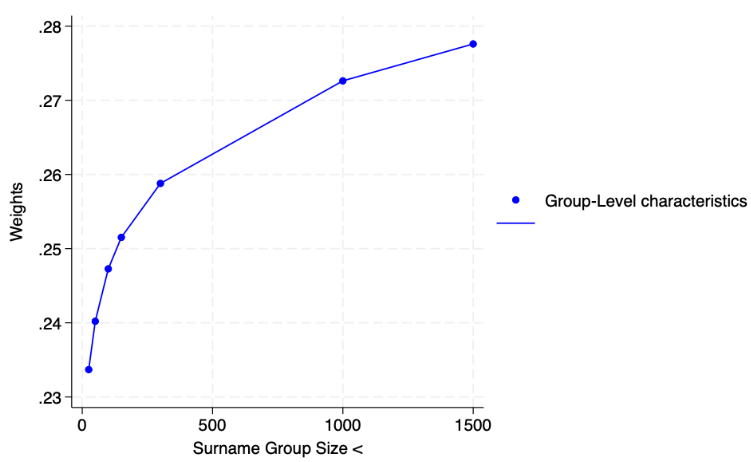
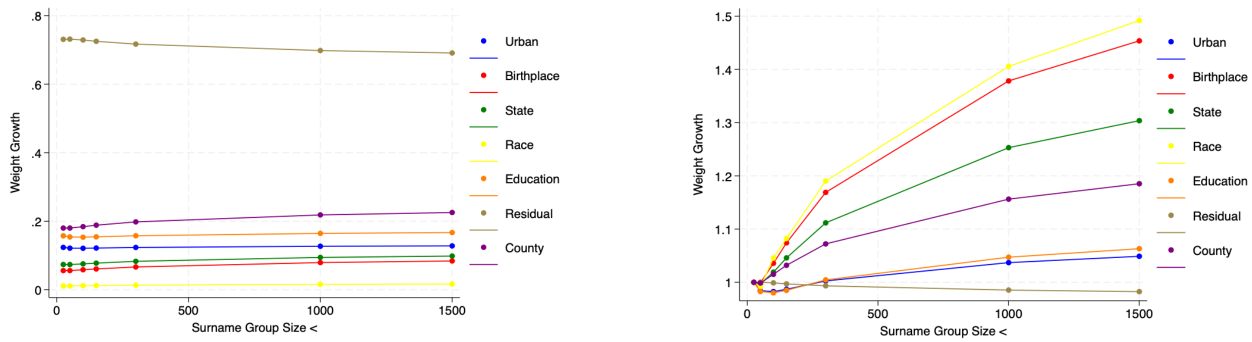


FIGURE 25: Weight - All Characteristics



$$\frac{\hat{\omega}_{g,N}}{\hat{\omega}_{g,25}} = \frac{V(\overline{f}_{gist}|n_s < N)}{V(\overline{f}_{gist}|n_s < 25)} \frac{V(\overline{y}_{ist}|n_s < 25)}{V(\overline{y}_{ist}|n_s < N)}$$

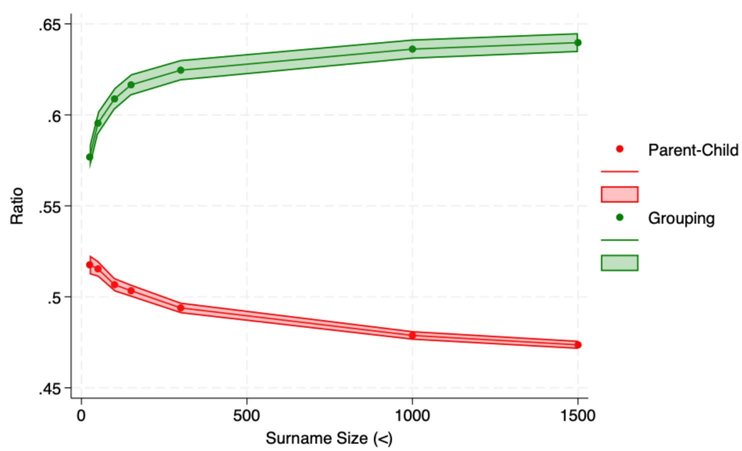
FIGURE 26: Weight's Growth over surname size - Wage Income



(A) Non-Normalized (B) Normalized

Note: In the left Panel: The figure illustrates the evolution of the weight of every single characteristic as surname size grows. We calculate the weight for any characteristic on surname groups smaller than any specified size at the **child** generation. In the right Panel: We divide the weights for each characteristic with respect to the weight in the first surname group size interval. Our analysis focuses exclusively on male workers aged between 30 and 40 in the 1940 US Census, linked to their fathers.

FIGURE 27: Grouping estimator and Parent-Child Correlation - Reweighted Regression

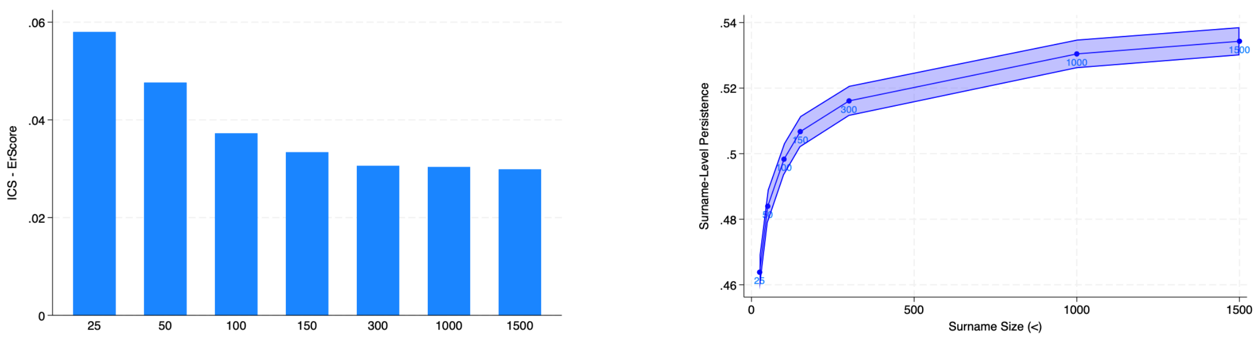


10.2 ALTERNATIVE OUTCOMES

In this section we show that our results do not hinge on the choice of the outcome variable. Recent studies *(cite)* discussed the validity of the variable here chosen as outcome, i.e. the Occupational Score. It represents the median income for a given occupation in 1950, hence it might not represent faithfully the economic structure of past decades. Consequently, we replicate the main empirical evidences (ICS and Surname-level IGE) with distinct outcome variables.

First, we employ the Earning Score which represents the rank of occupations by earnings in 1950. We choose this to show that the validity of our strategy holds even when using rank variables and not only income in logs (or proxies of it).

FIGURE 28: Main Results - Earning Score 1950

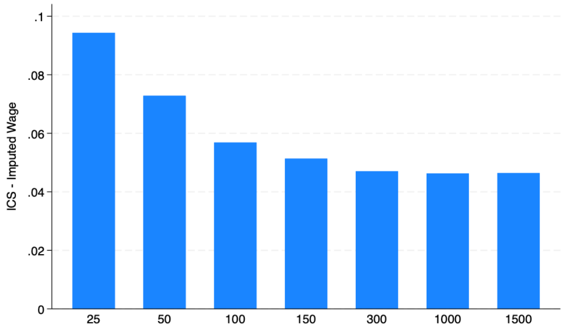


(A) ICS (B) Persistence - Surname-level IGE

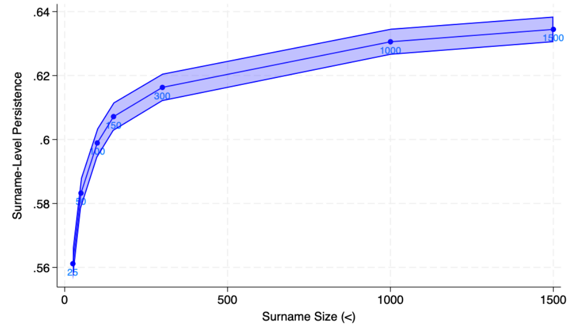
Note: In the left Panel: The figure illustrates the ICS over surname group size for the whole male working-age population in 1940. In the right Panel: The figure depicts the Persistence of Earning Score at the surname level on surname groups smaller than any specified size. Our analysis focuses exclusively on male workers aged between 30 and 40 in the 1940 US Census, linked to their fathers. The regressor employs the surname average computed from the working-age population in the 1920 US Census.

Second, we employ actual wages in 1940 imputed to each occupation. In this way, we control for the issue of the Occupational Score being based in 1950, a decade later with respect to our analysis, and for the reliability of Occupational Score as a proxy for income inside each occupation.

FIGURE 29: Main Results - Imputed Occupational Wages 1940



(A) ICS

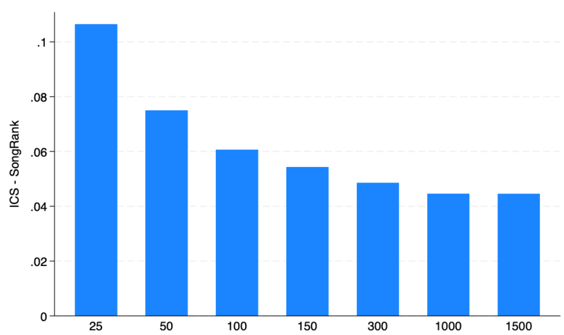


(B) Persistence - Surname-level IGE

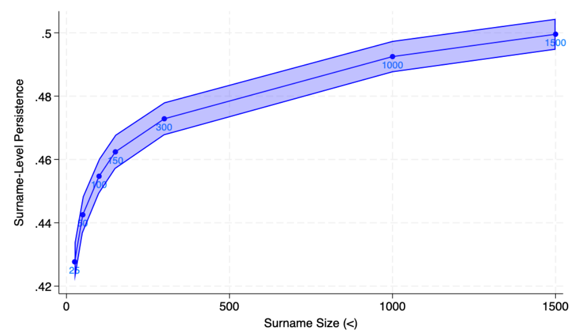
Note: In the left Panel: The figure illustrates the ICS over surname group size for the whole male working-age population in 1940. In the right Panel: The figure depicts the Persistence of Imputed Occupational Wages in 1940 at the surname level on surname groups smaller than any specified size. Our analysis focuses exclusively on male workers aged between 30 and 40 in the 1940 US Census, linked to their fathers. The regressor employs the surname average computed from the working-age population in the 1920 US Census.

Finally, we employ an alternative version of the Occupational Score suggested by (cite), i.e. the Song Rank, which controls for time and spatial variation in the Occupational Earnings.

FIGURE 30: Main Results - Song Rank



(A) ICS

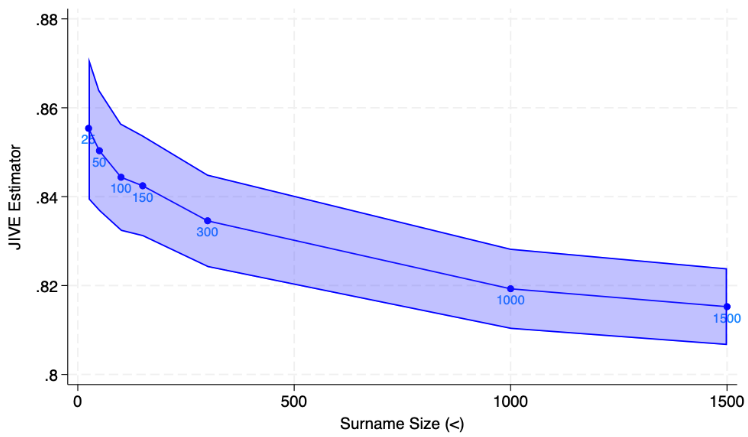


(B) Persistence - Surname-level IGE

Note: In the left Panel: The figure illustrates the ICS over surname group size for the whole male working-age population in 1940. In the right Panel: The figure depicts the Persistence of Song Ranks at the surname level on surname groups smaller than any specified size. Our analysis focuses exclusively on male workers aged between 30 and 40 in the 1940 US Census, linked to their fathers. The regressor employs the surname average computed from the working-age population in the 1920 US Census.

10.3 PATTERN OF THE JIVE

FIGURE 31: JIVE Estimator



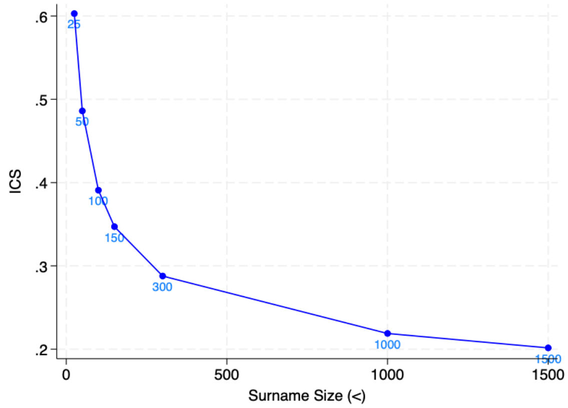
In this section we show the theoretical reason why, even in a MLF, the JIVE can generate a decreasing pattern over surname size. First, from Angrist [2014] we obtain the following formula relating the coefficient of any outcome regressed to its own leave-one-out mean. In our case this appears at the denominator of the JIVE estimator.

$$Cov(y_{it-1}; \overline{y}_{(i)t-1}) = \tau^2 - \frac{1-\tau^2}{N-1}$$

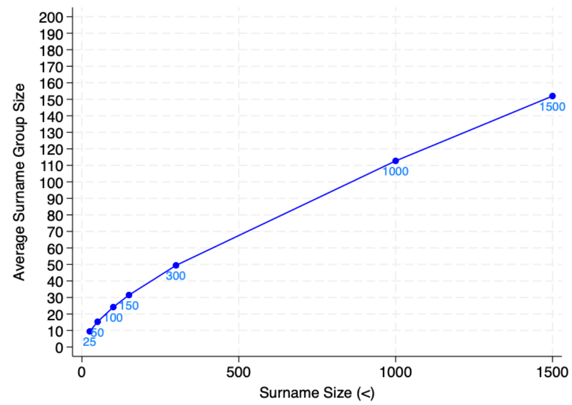
where $\tau^2 = \frac{V(\overline{y}_{it-1})}{V(y_{it-1})}$ and N represents the average group size. The bias is given by the negative term and it grows in τ^2 and it decreases in N. In our setting, τ^2 represents the ICS without the refinement for potential random grouping and N represents the average surname size. The part explained by the random grouping is likely to be the cause of the problem in JIVE estimates. As follows, first, we show that the bias is decreasing in surname size in our setting. The impact of surname size more than offsets the impact of decreasing τ^2. Second, we specify why this term can be considered a bias and what are its consequences in terms of the MLF model.

We compute the τ^2 in the estimation sample and N and we compute the overall value of the "bias". Recall that τ^2 is much larger than the ICS for two reasons. First, we are not accounting for potential random grouping. Second, in the estimation sample the groups are smaller than in the census. Hence, they pick up a larger share of the variability of the outcome, although in a mechanic way. I compute this τ^2 only beacuse is the variability used by the estimator, not for any other substantive reasons. Below, we show both the evolution of τ^2 and N over surname group size.

FIGURE 32: τ^2 and N in the estimation sample



(A) "ICS" - τ^2



(B) Average Surname Size (N)

Note: In the left Panel: The figure illustrates the τ^2 for surname groups smaller than any specified size. In the right Panel: The figure depicts the average surname size N for surname groups smaller than any specified size. Our analysis focuses exclusively on male workers aged between 30 and 40 in the 1940 US Census, linked to their fathers.

So we compute the bias and we notice that indeed decreases in magnitude over surname size.

TABLE 1: Bias over Surname Size

Size Restriction	τ^2	N	Bias	Bias in percentage of τ^2
25	.6	10	0.044	7.4
50	.49	15	0.036	7.4
100	.39	25	0.0254	6.5
150	.35	30	0.0224	6.4
300	.29	50	0.0144	4.9
1000	.21	110	0.007	3.3
1500	.2	150	0.005	2.5

Hence, we see that this negative deviation tends to fade over larger surname sizes. The impact of the increase in surname size more than offsets the decrease in τ^2. This explains why JIVE estimates decrease over surname size. However it is still not clear how to link this negative term to the theoretical model.

According to the MLF, the simple grouping estimator captures the persistence of factors common inside the surname group. This is due to the fact that each characteristic is weighted by $\rho^2 V(\overline{f}_{it-1})$, i.e. its relevance and its degree of within correlation. If the between variance is very small, which means that the surname group is not an indicator for that factor, the weight will converge to zero.

$$\omega = \frac{\rho^2 V(\overline{f}_{it-1})}{V(\overline{y}_{it-1})} \to 0$$

Consequently, the grouping estimator remains a weighted average of the persistence of the factors with all the weights being strictly non-negative.

On the contrary in the JIVE the weight is:

$$\omega = \frac{\rho^2 Cov(f_{it-1}; \overline{f}_{(i)t-1})}{V(\overline{y}_{it-1})}$$

and it might not be strictly non-negative. For instance, in a fully random grouping estimate, the leave-one-out average and the outcome are mechanically negatively correlated. This example would amount to have $\tau \approx 0$.

$$Cov(y_{it-1}; \overline{y}_{(i)t-1}) = -\frac{1}{N-1}$$

If the surname group is not indicative of the factor itself and its distribution is almost independent of the surname indicator then the weight can even become negative.

$$\omega = \frac{\rho^2 Cov(f_{it-1}; \overline{f}_{(i)t-1})}{V(\overline{y}_{it-1})} \to -\frac{\rho^2}{N-1}\frac{1}{V(\overline{y}_{it-1})}$$

At that point, the JIVE might not be anymore a weighted average of the factors' persistence, with all the weights strictly non-negative. Moreover, we know that factors not shared inside the surname group are usually less persistent than the captured ones. Consequently the estimate gets biased upward due to this mechanism.

As N grows, we can expect both ω_1 and ω_2 to grow. However their impact on the estimate is going to be in opposite directions. While the increase in ω_1 pushes the estimate upward, the increase in ω_2 (reduction in absolute value) shrinks it toward zero. Whenever the second channel is larger in magnitude than the first the evolution of the JIVE, even under MLF, becomes decreasing in average surname group size.

10.4 PROOFS

10.4.1 PROPOSITION 1

In this section, we describe the steps that lead to Proposition1 in section 4.1. From the model equation on transmission for characteristic k:

$$E[f_{kist}|sur = s] = \lambda_k E[f_{kist-1}|sur = s] + \underbrace{E[\epsilon_{kist}|sur = s]}_{=0}$$

By independence of the error in the transmission $E[\epsilon_{kist}|sur = s] = E[\epsilon_{kist}] = 0$. We iterate this procedure until we encounter the common ancestor for surname s in generation τ_s, whose value for characteristic k is taken as given. Hence,

$$E[f_{kis\tau_s}|sur = s] = f_{kis\tau_s}$$

The resulting equation becomes then:

$$E[f_{kist}|sur = s] = \lambda_k^{t-\tau_s} f_{kis\tau_s}$$

10.4.2 PROPOSITION 2

In this section we describe the steps that lead to Proposition 2:

$$E[t - \tau_s|n_s < N_2] - E[t - \tau_s|n_s < N_1] > 0 \quad \forall N_1 < N_2$$

First we can rewrite the first component as:

$$\begin{aligned} E[t - \tau_s|n_s < N_2] = E[t - \tau_s|n_s < N_1]P(n_s < N_1|n_s < N_2) \\ +E[t - \tau_s|N_1 < n_s < N_2]P(N_1 < n_s < N_2|n_s < N_2) \end{aligned}$$

We define $P(N_1 < n_s < N_2|n_s < N_2) = P(\delta)$ Plugging in and rearranging we get:

$$\underbrace{P(\delta)}_{>0}(E[t - \tau_s|N_1 < n_s < N_2] - E[t - \tau_s|n_s < N_1]) > 0$$

Finally we can rewrite this difference in expectations as follows, with K being the maximum distance to common ancestor:

$$\sum_{k=1}^{K} k \left(P(t - \tau_s = k|N_1 < n_s < N_2) - P(t - \tau_s = k|N_1 < n_s)\right) > 0$$

Now we take an assumption which is untestable, although credible inside a fertility process with increasing population. We define this property as monotonicity. For any possible level k of $t - \tau_s$ such that $P(t - \tau_s = k|n_s) \neq 0$, for any $\epsilon > 0$ and for any $n'_s > n_s$, the following holds.

$$P(t - \tau_s = k + \epsilon|n_s) - P(t - \tau_s = k|n_s) < P(t - \tau_s = k + \epsilon|n'_s) - P(t - \tau_s = k|n'_s)$$

This means that the difference between the probabilites of observing a given distance to common ancestor with a smaller one is decreasing in surname group size. Now we want to bring this assumption from each single surname group size to sets of surname group sizes. To this end, we sum in the left hand side the difference for all $n_s < N_1$ and we multiply by $P(n_s = n|n < N_1)$ both sides. Since each term of the summation is smaller than the right hand side assuming $n'_s = N_1$ the inequality holds:

$$\sum_{n=1}^{N_1} \left(P(t-\tau_s = k+\epsilon|n_s = n) - P(t-\tau_s = k|n_s = n)\right) P(n_s = n) <$$
$$\left(P(t-\tau_s = k+\epsilon|n'_s) - P(t-\tau_s = k|n'_s)\right) \underbrace{(\sum_{n=1}^{N_1} P(n_s = n))}_{=P(n<N_1)}$$

Hence, we can divide the LHS by $P(n < N_1)$ so that

$$\frac{P(n_s = n)}{P(n_s < N_1)} = P(n_s = n|n < N_1)$$

Hence we can rewrite the whole LHS as:

$$P(t-\tau_s = k+\epsilon|n_s < N_1) - P(t-\tau_s = k|n_s < N_1) < P(t-\tau_s = k+\epsilon|n'_s) - P(t-\tau_s = k|n'_s)$$

Now we replicate the same procedure with the RHS, adding up for all $N_1 < n'_s < N_2$ and multiplying by the respective probability both sides. Since the monotonicity assumption holds for any $n'_s > n_s$, the inequality still holds. We replicate the same set of operations and we end up with the following inequality:

$$P(t-\tau_s = k+\epsilon|n_s < N_1) - P(t-\tau_s = k|n_s < N_1) <$$
$$P(t-\tau_s = k+\epsilon|N_1 < n_s < N_2) - P(t-\tau_s = k|N_1 < n_s < N_2)$$

We can further rearrange to get:

$$P(t-\tau_s = k+\epsilon|N_1 < n_s < N_2) - P(t-\tau_s = k+\epsilon|n_s < N_1)$$
$$> P(t-\tau_s = k|N_1 < n_s < N_2) - P(t-\tau_s = k|n_s < N_1)$$

Defining this difference in probabilities as d_k:

$$P(t-\tau_s = k|N_1 < n_s < N_2) - P(t-\tau_s = k|n_s < N_1) = d_k$$

We can determine that monotonicity implies this difference being increasing in k, which represents any possible level of distance to the common ancestor:

$$d_k < d_{k+\epsilon}$$

Hence, we can go back to our original equation and we can rewrite it as:

$$\sum_{k=1}^{K} k d_k > 0$$

Knowing that d_k is increasing and $\sum_{k=1}^{K} d_k = 0$.

10.4.3 Corollary 1

In this section we describe the steps that lead to Corollary 1. Most of the logical steps are similar to the previous proof, hence we will recall most of it from there. We want to prove the following:

$$V(E[f_{ist}|sur]|n_s < N_1) > V(E[f_{ist}|sur]|n_s < N_2) \iff N_1 < N_2$$

First from Proposition 1, we can rewrite the variance as follows:

$$\begin{aligned} V(E[f_{list}|sur]|n_s < N_j) &= V(\lambda_l^{t-\tau_s} f_{lis\tau_s}|n_s < N_j) \\ &= E[\lambda_l^{2(t-\tau_s)} f_{lis\tau_s}^2|n_s < N_j] - E[\lambda_l^{t-\tau_s} f_{lis\tau_s}|n_s < N_j]^2 \end{aligned}$$

Now we assume independence of $f_{lis\tau_s}$ from n_s. This means that surname group size does not influence the characteristics of the common ancestor.

$$\begin{aligned} &E[\lambda_l^{2(t-\tau_s)} f_{lis\tau_s}^2|n_s < N_j] - E[\lambda_l^{t-\tau_s} f_{lis\tau_s}|n_s < N_j]^2 = \\ &E[\lambda_l^{2(t-\tau_s)}|n_s < N_j] \underbrace{E[f_{lis\tau_s}^2]}_{=1} - E[\lambda_l^{t-\tau_s}|n_s < N_j]^2 \underbrace{E[f_{lis\tau_s}]^2}_{=0} \\ &= E[\lambda_l^{2(t-\tau_s)}|n_s < N_j] \end{aligned}$$

Hence, we can rewrite our original equation as follows:

$$E[\lambda_l^{2(t-\tau_s)}|n_s < N_2] - E[\lambda_l^{2(t-\tau_s)}|n_s < N_1] < 0$$

As in the previous proof we can rewrite this equation as:

$$E[\lambda_l^{2(t-\tau_s)}|N_1 < n_s < N_2] - E[\lambda_l^{2(t-\tau_s)}|n_s < N_1] < 0$$

Then, similarly to before we open the expectation and we gather on k.

$$\sum_{k=1}^{K} \lambda_l^{2k} \left(\underbrace{P(t-\tau_s = k|N_1 < n_s < N_2) - P(t-\tau_s = k|n_s < N_1)}_{=d_k} \right) < 0$$

By the assumption of monotonicity, and operating specularly to the previous proof, we know that d_k is increasing in k. Hence we find a similar problem to Proposition 2.

$$\sum_{k=1}^{K} \lambda_l^{2k} d_k < 0$$

Knowing that d_k is increasing and $\sum_{k=1}^{K} d_k = 0$ and $0 < \lambda_l < 1$.

10.4.4 Corollary 2

In this section we describe the steps to Corollary 2. We want to prove the following: $\forall\ N_1 < N_2\ \wedge\ \lambda_j > \lambda_l$

$V(E[f^j_{ist}|sur]|n_s < N_1) - V(E[f^j_{ist}|sur]|n_s < N_2) < V(E[f^l_{ist}|sur]|n_s < N_1) - V(E[f^l_{ist}|sur]|n_s < N_2)$

By Corollary 1 we know that we can rewrite each side of the inequality as :

$$\sum_{k=1}^{K} \lambda_j^{2k} d_k < \sum_{k=1}^{K} \lambda_l^{2k} d_k$$

Then we can gather in the LHS and obtain:

$$\sum_{k=1}^{K} (\lambda_j^{2k} - \lambda_l^{2k}) d_k < 0$$

where we know that $\lambda_j^{2k} - \lambda_l^{2k} > 0$ for any $k > 0$. Moreover, from monotonicity we know that d_k is increasing and being d_k a difference of probabilities then $\sum_{k=1}^{K} d_k = 0$.

10.4.5 ICS

For the proof of equation 2, first we can write the surname level average as follows:

$$\begin{aligned} E_n[y_{it}|sur = j] =& \rho_1 \lambda_1^{t-\tau} f_{1i\tau} + \rho_2 \lambda_2^{t-\tau} f_{1i\tau} + \rho_1 \sum_{j=0}^{\tau-1} \lambda_1^j E_n[\epsilon_{1i,t-j}|sur = j] \\ &+ \rho_2 \sum_{j=0}^{\tau-1} \lambda_2^j E_n[\epsilon_{2i,t-j}|sur = j] + E_n[u_{it}|sur = j] \end{aligned}$$

This means that the surname average is given by the values of the factors of the **Common Ancestor** ($f_{ki\tau}$), living in generation τ. Furthermore the shock to the transmission process matter as long as the surname group size is small enough so that the empirical average does not converge to the population average of the shock which is identically zero.

Now, we know that the ICS varies over surname size. We define Surname Size as an indicator between two actual bound ($n1$,$n2$) and for brevity we will define it as follows.

$$Sz(n1, n2) = 1(n_1 < Size < n_2)$$

Consequently, we consider the $V(E_n[y_{it}|sur]|Sz(n_1, n_2 = 1)$. This is the variance conditional on a given surname size included between n_1 and n_2. For further simplify, we will always only indicate conditional on sz, although it is an actual number and not a function.

$$\begin{aligned}V(E_n[y_{it}|sur]|sz) =&\rho_1^2 V(\lambda_1^{t-\tau} f_{1i\tau}|sz) + \rho_2^2 V(\lambda_2^{t-\tau} f_{2i\tau}|sz) + \rho_1^2 V(\sum_{j=0}^{\tau-1} \lambda_1^j E_n[\epsilon_{1i,t-j}|sur]|sz)\\ &+ \rho_2^2 V(\sum_{j=0}^{\tau-1} \lambda_2^j E_n[\epsilon_{2i,t-j}|sur]|sz) + V(E_n[u_{it}|sur]|sz)\end{aligned}$$

For what concerns the $V(E_n[y_{it}|fakesur]|sz)$, we take an assumption to simplify computations. Since we can go back to infinity in principle as there is no common ancestor, we stop for each fake surname at the same generation of the common ancestor for each actual surname.

$$\begin{aligned}V(E_n[y_{it}|fakesur]|sz) =&\rho_1^2 V(\lambda_1^{t-\tau} E_n[f_{1i\tau}|fakesur]|sz) + \rho_2^2 V(\lambda_2^{t-\tau} E_n[f_{2i\tau}|fakesur]|sz)\\ &+ \rho_1^2 V(\sum_{j=0}^{\tau-1} \lambda_1^j E_n[\epsilon_{1i,t-j}|fakesur]|sz) + \rho_2^2 V(\sum_{j=0}^{\tau-1} \lambda_2^j E_n[\epsilon_{2i,t-j}|fakesur]|sz)\\ &+ V(E_n[u_{it}|fakesur]|sz)\end{aligned}$$

Now, since we know that the distribution of the factors is independent of the distance to the common ancestor (or to the generation of the common ancestor τ), we can write as follows:

$$\begin{aligned}V(\lambda_k^{t-\tau} f_{ki\tau}|sz) &= E[\lambda_k^{2(t-\tau)} f_{ki\tau}^2|sz] - E[\lambda_k^{t-\tau} f_{ki\tau}|sz]^2\\ &= E[\lambda_k^{2(t-\tau)}|sz]E[f_{ki\tau}^2] - (E[\lambda_k^{t-\tau}|sz]E[f_{ki\tau}])^2\\ &= E[\lambda_k^{2(t-\tau)}|sz]\end{aligned}$$

since $E[f_{ki\tau}^2] = 1$ and $E[f_{ki\tau}] = 0$. Furthermore we can partially extend this result to the case of fake surnames. Due to the random nature of the grouping, we know that:

$$E[E_n[f_{ki\tau}|fakesur]|sz] = 0$$

From this and the previous proof, we can then write:

$$V(\lambda_k^{t-\tau} E_n[f_{ki\tau}|fakesur]|sz) = E[\lambda_k^{2(t-\tau)} \frac{V(f_{ki\tau})}{n_j}|sz]$$

where n_j is the surname size depending on the interval surname group size interval. Finally, we consider the shocks. Since they are iid, it does not make a difference whether we average over real surnames or fake surnames given that they have the same distribution.

$$V(\sum_{j=0}^{\tau-1}\lambda_k^j E_n[\epsilon_{ki,t-j}|fakesur]|sz) = V(\sum_{j=0}^{\tau-1}\lambda_k^j E_n[\epsilon_{ki,t-j}|sur]|sz)$$
$$V(E_n[u_{it}|fakesur]|sz) = V(E_n[u_{it}|sur]|sz)$$

Now we have all the elements to write the ICS.

$$ICS_{n_1,n_2} = \frac{V(E_n[y_{it}|sur]|sz) - V(E_n[y_{it}|fakesur]|sz)}{V(y_{it})}$$
$$= \frac{\rho_1^2(E[\lambda_1^{2(t-\tau)}|sz] - E[\frac{\lambda_1^{2(t-\tau)}}{n_j}|sz]) + \rho_2^2(E[\lambda_2^{2(t-\tau)}|sz] - E[\frac{\lambda_2^{2(t-\tau)}}{n_j}|sz])}{\rho_1^2 + \rho_2^2 + V(u_{it})}$$

10.4.6 Ratio - Controlling for covariates

First, accounting for geography in the grouping regression amounts to perform the pseudo-IGE on the residuals of both the outcome and the regressor onto the controlling characteristic at the father level. We take three factors as the third can also be considered as the error term restricting its relevance parameter to one and its transmission parameter to 0

$$y_{it} = \rho_1 f_{1it} + \rho_2 f_{2it} + \rho_3 f_{3it}$$
$$E[y_{it}|f_{1it-1}] = \lambda_1\rho_1 f_{1it}$$
$$E[\overline{y}_{it-1}|f_{it-1}] = \rho_1\overline{f}_{1it}$$

Then, the residuals will be as follows:

$$y_{it} = \rho_1 f_{1it} + \rho_2 f_{2it} + \rho_3 f_{3it}$$
$$\tilde{y}_{it} = y_{it} - E[y_{it}|f_{1it-1}] = \rho_2 f_{2it} + \rho_3 f_{3it} + \rho_1\epsilon_{1it}$$
$$\tilde{\overline{y}}_{it-1} = \overline{y}_{it-1} - E[\overline{y}_{it-1}||f_{1it-1}] = \rho_2\overline{f}_{2it-1} + \rho_3\overline{f}_{3it-1}$$

Consequently, we perform the estimation:

$$\beta_{n,g} = \frac{Cov(\tilde{y}_{it};\tilde{\overline{y}}_{it-1})}{Var(\tilde{\overline{y}}_{it-1})} = \frac{\lambda_2\rho_2^2 V(\overline{f}_{2it-1}) + \lambda_3\rho_3^2 V(\overline{f}_{3it-1})}{\rho_2^2 V(\overline{f}_{2it-1}) + \rho_3^2 V(\overline{f}_{3it-1})}$$

Now we take the ratio between the standard grouping estimator and the grouping controlled for geography. Recall that we indicate the estimate with subindex n as they are retrieved using surnames smaller than size n. Also we will indicate for brevity $V_j = \rho_j^2 V(\overline{f}_{jit-1})$

$$\begin{aligned}
\frac{\beta_n}{\beta_{n,g}} &= \left(\frac{V_2 + V_3}{V_1 + V_2 + V_3}\right)\left(\frac{\lambda_1 V_1 + \lambda_2 V_2 + \lambda_3 V_3}{\lambda_2 V_2 + \lambda_3 V_3}\right) \\
&= \left(\frac{\lambda_1(V_1V_2 + V_1V_3) + \lambda_2(V_2^2 + V_2V_3) + \lambda_3(V_2V_3 + V_3^2)}{\lambda_2 V_2(V_1 + V_2 + V_3) + \lambda_3 V_3(V_1 + V_2 + V_3)}\right) \\
&= \left(\frac{\lambda_1(V_1V_2 + V_1V_3) + \lambda_2(V_2^2 + V_2V_3) + \lambda_3(V_2V_3 + V_3^2)}{\lambda_2(V_2^2 + V_2V_3) + \lambda_3(V_2V_3 + V_3^2) + \lambda_2 V_1 V_2 + \lambda_3 V_1 V_3}\right)
\end{aligned}$$

Now we sum and subtract at the denominator $\lambda_1(V_1V_2 + V_1V_3)$ and we regroup the numerator as k We obtain the following:

$$\frac{\beta_n}{\beta_{n,g}} = \frac{k}{k + (\lambda_2 - \lambda_1)V_1V_2 + (\lambda_3 - \lambda_1)V_1V_3}$$

We compute the adjusted ratio by subtracting one and we get:

$$\frac{\beta_n}{\beta_{n,g}} - 1 = \frac{(\lambda_1 - \lambda_2)V_1V_2 + (\lambda_1 - \lambda_3)V_1V_3}{k + (\lambda_2 - \lambda_1)V_1V_2 + (\lambda_3 - \lambda_1)V_1V_3}$$

Finally, we rearrange back the denominator as above for interpretability, recognizing that $(V_1 + V_2 + V_3) = (V_1 + V_2 + V_3) = V(\overline{y}_{it-1})$

$$\begin{aligned}
R^*_{n,g} = \frac{\beta_n}{\beta_{n,g}} - 1 &= \frac{(\lambda_1 - \lambda_2)V_1V_2 + (\lambda_1 - \lambda_3)V_1V_3}{\lambda_2 V_2(V_1 + V_2 + V_3) + \lambda_3 V_3(V_1 + V_2 + V_3)} \\
&= \left(\frac{1}{V(\overline{y}_{it-1})}\right)\left(\frac{(\lambda_1 - \lambda_2)V_1V_2 + (\lambda_1 - \lambda_3)V_1V_3}{\lambda_2 V_2 + \lambda_3 V_3}\right)
\end{aligned}$$

Now, we can see that the first component is always positive and common to every characteristic, so it is invariant with respect to which factor you are controlling for. The second instead is quite interesting as we can interpret both the sign and the magintude depending on the two defining parameters of a characteristic: relevance and persistence.

Regarding the sign, we observe that the sign of the ratio clearly depends on the numerator as the denominator is always positive. In the numerator, if we control for a characteristic more persistent than others, i.e. $\lambda_1 > \lambda_2 \ \ \& \ \ \lambda_1 > \lambda_3$, the adjusted ratio is going to be positive. Otherwise, the ratio is going to be negative. Regarding the magnitude, we see that V_1 which we recall contains the relevance parameter only appears at the numerator. Hence, while the sign is solely determined by the persistence parameter, the magnitude on the ratio depends on both the deviation of the persistence of the characteristic with respect to the others and the relevance of the characteristic itself.

Now, we evaluate what it means to control for characteristics in the setting of the simple latent factor model devoid of the multiplicity factor. We assume two independent factors and an error term but we constrain the persistence parameter of both factors. We make a similar argument with respect to the Multiple Latent Factor Model, only fixing one factor to be an iid error term and the other two to have the same persistence parameter. Consequently, the estimator for the unconditional surname-level IGE is the following:

$$\beta_n = \frac{\lambda(V_1 + V_2)}{V_1 + V_2 + V_u}$$

For the Frisch-Wald-Lowell theorem, then the conditional one will just exclude the first factor.

$$\beta_{n,1} = \frac{\lambda(V_2)}{V_2 + V_u}$$

Hence, taking the ratio:

$$\begin{aligned}\frac{\beta_n}{\beta_{n,1}} &= \left(\frac{\lambda(V_1 + V_2)}{V_1 + V_2 + V_u}\right)\left(\frac{V_2 + V_u}{\lambda(V_2)}\right) \\ &= \frac{\lambda(V_1V_2 + V_2^2 + V_2V_u + V_1V_u)}{\lambda(V_1V_2 + V_2^2 + V_2V_u)}\end{aligned}$$

We already notice that this ratio is always larger than one, but now we consider the adjusted version and it is going to be clearer that this is always positive.

$$\frac{\beta_n}{\beta_{n,1}} - 1 = \frac{V_1V_u}{V_1V_2 + V_2^2 + V_2V_u}$$

In simple terms, this positive deviation is generated by the fact that the attenuation bias of the error term is smaller when you do not control as you would capture a larger portion of the latent factor. Moreover, since this is true, as we include larger and larger surname groups the impact of the attenuation bias shrinks toward zero. Formally V_u goes to zero as surname group size increases faster than V_1 or V_2. This means that we expect this ratio to converge toward zero as surname size increases.

10.4.7 Heterogeneity across families

We define a naive version of persistence inside one's family as the percentage deviation of the child outcome from the parent outcome in absolute value.

$$\begin{aligned}\phi_s &= E\left[\left(\frac{y_{ist} - y_{ist-1}}{y_{ist-1}}\right)^2 | f_{1ist-1}, f_{2ist-1}\right] \\ &= \frac{1}{y_{ist-1}^2} E\left[\left(\rho_1\left((\lambda_1 - 1)f_{1ist-1} + \epsilon_{1ist-1}\right) + \rho_2\left((\lambda_2 - 1)f_{2ist-1} + \epsilon_{2ist-1}\right)\right)^2 | f_{1ist-1}, f_{2ist-1}\right]\end{aligned}$$

By independence between factors and errors, we can get rid of any cross product generated by the square. Hence, we can rewrite

$$\begin{aligned}
\phi_s &= \frac{1}{y_{ist-1}^2} E\left[\rho_1^2\left((\lambda_1-1)f_{1ist-1}+\epsilon_{1ist-1}\right)^2+\rho_2^2\left((\lambda_2-1)f_{2ist-1}+\epsilon_{2ist-1}\right)^2|f_{1ist-1},f_{2ist-1}\right] \\
&= \frac{1}{y_{ist-1}^2} E\left[\rho_1^2(\lambda_1-1)^2f_{1ist-1}^2+\rho_1^2\epsilon_{1ist-1}^2+\rho_2^2(\lambda_2-1)^2f_{2ist-1}^2+\rho_2^2\epsilon_{2ist-1}^2|f_{1ist-1},f_{2ist-1}\right]
\end{aligned}$$

Since we are conditioning on the factors and errors are independent instead we can rewrite as:

$$\phi_s = \frac{1}{y_{ist-1}^2}\rho_1^2(\lambda_1-1)^2f_{1ist-1}^2+\rho_1^2(1-\lambda_1^2)+\rho_2^2(\lambda_2-1)^2f_{2ist-1}^2+\rho_2^2(1-\lambda_2^2)$$

We group whatever does not depend on the factors, as it is constant over them and we define it as κ

$$\phi_s = \kappa+\frac{\rho_1^2(\lambda_1-1)^2f_{1ist-1}^2}{y_{ist-1}^2}+\frac{\rho_2^2(\lambda_2-1)^2f_{2ist-1}^2}{y_{ist-1}^2}$$

As before, by independence of the factors we can rewrite the square of the father's outcome as the square of each component. Then, consistently with previous discussions, we defined the weights:

$$\omega_j = \frac{\rho_j^2 f_{jist-1}^2}{y_{ist-1}^2} = \frac{\rho_j^2 f_{jist-1}^2}{\rho_1^2 f_{1ist-1}^2+\rho_2^2 f_{2ist-1}^2}$$

Hence, we can rewrite our naive measure of family-specific persistence as:

$$\phi_s = \kappa+(\lambda_1-1)^2\omega_1+(\lambda_2-1)^2\omega_2$$

Given that the weights sum up to 1 we can rearrange as:

$$\phi_s = \kappa+(\lambda_1-1)^2+\left((\lambda_2-1)^2-(\lambda_1-1)^2\right)\omega_2$$

Then, we take derivative of the family specific measure of intergenerational persistence with respect to the weight of the persistent factor.

$$\frac{\partial\phi_s}{\partial\omega_2} = (\lambda_2-1)^2-(\lambda_1-1)^2$$

Under the assumption that the persistent factor is the second, hence $\lambda_2 > \lambda_1$:

$$\begin{aligned}
(\lambda_2-1)^2-(\lambda_1-1)^2 &< 0 \\
(\lambda_2-1)^2 &< (\lambda_1-1)^2 \\
|(\lambda_2-1)| &< |(\lambda_1-1)|
\end{aligned}$$

Given that we know that these quantities are negative because λ is negative:

$$|(\lambda_2 - 1)| < |(\lambda_1 - 1)|$$
$$1 - \lambda_2 < 1 - \lambda_1$$
$$\lambda_2 > \lambda_1$$

Hence, we prove that as the share of the persistent factor grows the expected distance of the outcome of the child with respect to the outcome of the father falls.

10.4.8 Ar2

We prove the behaviour of the grouping estimator under any stationary Ar(2) statistical process. The first-order autocovariance of any stationary $Ar(2)$ process can be written as follows:

$$Cov(\overline{y}_{it}, \overline{y}_{it-1}|Sz(n)) = \left(\frac{\rho_1}{1-\rho_2}\right) V(\overline{y}_{it}|Sz(n))$$

Consequently, the estimator is going to capture the following ratio:

$$\beta_n^G = \frac{Cov(\overline{y}_{it}, \overline{y}_{it-1}|Sz(n))}{V(\overline{y}_{it}|Sz(n))} = \frac{\left(\frac{\rho_1}{1-\rho_2}\right) V(\overline{y}_{it}|Sz(n))}{V(\overline{y}_{it}|Sz(n))} = \left(\frac{\rho_1}{1-\rho_2}\right)$$

Universidad
Carlos III de Madrid
www.uc3m.es

Departamento de Economía

C/ Madrid, 126 • Edificio López Aranguren
28903 Getafe • Madrid

Jan Stuhler
Associate Professor
Universidad Carlos III de Madrid
Department of Economics
Email: jan.stuhler@uc3m.es
Office: +34 91 624 9678

Monday, June 13, 2024

Dear committee of the Spanish Economic Association,

I hereby confirm that Andrea Del Pizzo, born on March 25th, 1997 in Pordenone (Italy), is a PhD student at the Universidad Carlos III under my supervision since March, 2023. If you have any questions, please feel free to contact me directly.

Best wishes,
Jan Stuhler