The Dynamics of the "Great Gatsby Curve", and a look at the curve during the Great Gatsby Era

Diego Battiston [*] Stephan Maurer [†] Andrei Potlogea [‡]
José V. Rodríguez Mora [§]

June 9, 2023

Abstract

We use linked historical US censuses to study the empirical relationship between inequality and intergenerational mobility. We first confirm that the "Great Gatsby Curve" already existed in the early 20th century. We then study a "dynamic" version of the curve that relates changes in equality to changes in intergenerational mobility. Surprisingly, we find that this relationship is unstable over horizons of two decades for income, but not for education. Finally, we propose novel unitless measures of intergenerational mobility and inequality to show that the "Great Gatsby Curve" result re-emerges over the long run, for the period 1920 to 2011.

JEL classification: J62, N12, N52, R11

Keywords: Intergenerational Mobility, Inequality, Great Gatsby Curve.

[*] University of Edinburgh and CEP
[†] University of Konstanz and CEP
[‡] University of Edinburgh
[§] University of Edinburgh and CEPR

1 Introduction

Equality of outcome and equality of opportunity, as well as the relationship between the two, are at the forefront of public policy debates surrounding critical issues such as taxation, redistribution, and public good provision. Motivated by these debates, a sizable academic literature has developed studying the relationship between inequality, a measure of equality of outcomes, and intergenerational socioeconomic mobility, a proxy for equality of opportunity.[1] This literature has found a robust positive association between intergenerational persistence and inequality which has been called "The Great Gatsby Curve" (GGC for short).[2] This positive association has been confirmed in a number of settings, both across countries and across regions within countries.[3] Thus, at least in the first decades of the 21st century, in places where there is high inequality, there seems to be relatively little intergenerational mobility, and vice-versa.

The implications of the GGC are potentially troubling, as it would seem to refute the often repeated "prospect of upward mobility" narrative, particularly popular in the US, by which society would accept high inequality of outcomes in exchange for a more dynamic economy characterized by greater equality of opportunity. Moreover, extrapolating the relationship between inequality and mobility to a dynamic setting, one may infer that in places where inequality *rises* rapidly, intergenerational mobility can be expected to decrease rapidly as well. This type of inference may be particularly concerning given the recent dynamics of US inequality, as already low levels of socioeconomic mobility may be expected to decline further given the recent marked rise in income inequality.[4]

Nevertheless, to the best of our knowledge, the dynamic relationship between intergenerational socieconomic mobility and income inequality has not been directly studied before, which leaves inferences such as the one described above merely in the realm of speculation. This relative dearth of evidence is perhaps best explained by data limitations. It is unusual to have comparable measures of intergenerational mobility for a large panel of societies covering a meaningful time period that spans several generations. Most existing work is only able to measure mobility across relatively short time periods, typically a single generation, and as a result is only able to exploit cross-sectional variation in mobility (across regions or

[1] See for instance, Hassler et al. (2007a); Andrews and Leigh (2009); Bjorklund and Jäntti (2012); Blanden (2013), Corak (2006), Corak (2013); Ermisch et al. (2012); Durlauf and Seshadri (2018); Fan et al. (2021); Güell et al. (2018); DiPrete (2020)

[2] Alan Krueger was the first to refer to the empirical association between inequality and intergenerational income persistence as the "Great Gatsby Curve" referring to the work of Corak.

[3] See an excellent survey of the literature in Durlauf et al. (2022)

[4] For a documentation of this rise in inequality, see for example Piketty and Saez (2003), Autor et al. (2008) and Acemoglu and Autor (2011)

countries) at a certain (and typically recent) point in time. For a similar reason, we do not have any historical perspective on the stability of the Great Gatsby Curve over longer time periods or evidence of whether this regularity was also present in the more distant past.

This paper aims to advance our understanding of the relationship between income inequality and socioeconomic mobility by building and studying a long panel of inequality and mobility that covers multiple generations. To do so, we make use of recently made available linked historical censuses for the US. Specifically, we calculate inequality and several measures of intergenerational mobility (in terms of both income and education) at the level of US counties for a period spanning the years 1880 to 1940.[5] To compute county-level measures of intergenerational persistence we link the censuses 1880-1900, 1900-1920, and 1920-1940 and locate pairs of fathers and sons, observing outcomes of the parents in the older census and that of the sons in the newer census. Intergenerational persistence is then measured as the average correlation of these outcomes at the county level. We measure county-level inequality as the dispersion in individual incomes as reflected by the county's Gini coefficient.

We make use of this newly assembled dataset to make three contributions to the existing literature. First, we check for the presence of a "Great Gatsby Curve" across US counties during the first half of the twentieth century. We document a positive association between inequality among fathers and intergenerational persistence in both income and education, which is robust across different definitions of inequality and intergenerational persistence. This confirms that the "Great Gatsby Curve" was a salient feature of the data at the time when the Great Gatsby would have been alive. We believe that it is not an irrelevant result, as it shows that the hypothetical tradeoff between inequality and mobility not only does not exist today, but it did not exist either in the relatively distant past, not even at a time characterized by massive changes and migration.

Second, we make use of the panel structure of our data to study, we believe for the first time, the relationship between *changes* in inequality and *changes* in socioeconomic mobility across US counties during the first half of the twentieth century. We think of this exercise as an attempt to document a "Dynamic Great Gatsby Curve" across US counties for the first half of the twentieth century. We find that increases in income inequality do not always correlate with decreases in the intergenerational mobility of income, at least over time horizons spanning a generation (i.e. circa 20 years). In other words, the relationship between changes in *income* inequality and changes intergenerational *income* mobility is unstable during our period of analysis. By contrast, the relationship between changes in income inequality and

[5]Thus, had the Great Gatsby been a real-life person, instead of a fictional one, he would be part of our micro-data.

changes in the intergenerational persistence of *educational* attainment is positive and robust over the period covered by our sample, mimicking our findings relating to the more conventional, "static" Great Gatsby Curve (for income and education).

Last but not least, we study the long-run relationship between income inequality and socioeconomic mobility, by relating changes in inequality to changes in socioeconomic persistence over the period 1920 to 2011. In performing this exercise we face two key challenges. First, we do not have access to contemporary data allowing us to measure current mobility and inequality in the same manner as we do for the Great Gatsby Era. Second, while other researchers with access to restricted use data have been kind enough to make measures of inequality and socioeconomic mobility for US counties during the recent past publicly available (see Chetty et al. (2014)), these measures are different from ours and impossible to replicate with our data.

To overcome these challenges and study the relationship between inequality and mobility over the long-run, we propose alternative *unitless* measures of inequality and mobility: the county ranks. For each time period, we rank all US counties by the available measures of inequality and socieconomic mobility, and think of these ranks as novel measures of inequality and mobility at the county level. These measures capture informative variation in the relative position of counties in terms of both inequality and socieconomic mobility. Moreover, changes in county ranks are arguably a useful measure of relative changes at the county level for both inequality and socioeconomic mobility, even if the ranks are based on different underlying measures of inequality and mobility at different points in time.

We then employ these county ranks to study the association between inequality and socioeconomic mobility across US counties, both statically and dynamically. We first validate our new measures by showing that they produce qualitatively similar results to our previous analysis using more established measures with respect to the "static" and "short-run dynamic" GGC: we find a positive association between inequality and socioeconomic mobility (in terms of both income and education) during the first half of the twentieth century, as well as an unstable relationship between changes in inequality and changes in the intergenerational persistence of income over 20 year horizons during the same time period. We also confirm the robust positive association between changes in inequality and changes in the intergenerational persistence of education during the first half of the twentieth century, as well as the GGC pattern for the modern era documented by Chetty et al. (2014). Finally, we calculate the long-run dynamic Great Gatsby Curve across US counties by correlating the changes in county ranks in terms of inequality with changes in county ranks in terms of socioeconomic mobility, using our data to construct the ranks for the early time period and the Chetty et al. (2014) data to construct ranks for the later time periods. We find that in the long run (i.e. over the period 1920 to 2011) the positive relationship between inequality and intergenerational income persistence reappears.

Our findings suggest a complex process governs the joint determination of inequality and socioeconomic mobility. For the relationship between inequality and intergenerational income mobility, we observe a positive cross-sectional association, as well as a positive dynamic relationship over the long-run, while the relationship between changes in inequality and changes in intergenerational persistence is less stable over shorter time horizons of two decades. By contrast, for the relationship between income inequality and intergenerational educational persistence we find a highly robust positive association in all of the settings we were able to study.

Our interpretation of these results is that structural factors generate a correlation between changes in inequality and changes in intergenerational *educational* persistence, as suggested for instance in Hassler et al. (2007b)[6] This relationship translates into a positive relationship between inequality and intergenerational income mobility in the "long-run" or "steady state". However, given that the income process is noisy and subject to large shocks, the positive relationship between changes in inequality and the intergenerational persistence of income may fail to hold over relatively short time periods, as we observe in our data for the period 1920 to 1940. This interpretation seems particularly plausible given that the period for which we fail to identify the positive relationship found elsewhere, namely the period 1920 to 1940 is characterized by particularly large shocks such as the Great Depression, while when referring to educational attainment the relationship remains unaltered.

2 Data

To measure our key variables of interest, income inequality and intergenerational socioeconomic mobility, we link men across censuses using the full count censuses available for the late 19th and early 20th centuries from IPUMS (Ruggles et al. (2021)). We draw on two different methods for linking people across censuses developed in the literature. The first, developed by Abramitzky et al. (2020) uses a fully automated approach that creates links based on standardized first and last names, as well as age. The second method, developed by Helgertz et al. (2023) uses a probabilistic approach that employs machine learning techniques and also incorporates information on birthplace and family/household characteristics.

[6]They present a model where differences in accessibility to public education generate a Great Gatsby curve, while differences in labor market environments would generate a negative relationship

Perhaps surprisingly, the overlap between the links identified by the two approaches is not large. Because of this, using matches created via both procedures results in a much larger sample size. It should also be noted that both linking approaches that we employ use the last name as key information. Because of this, we omit females from our analysis as their surnames typically changed upon marriage, making it difficult to link across censuses.

We create links across three 20-year intervals: 1880 to 1900, 1900 to 1920, and 1920 to 1940. In each interval, we link sons to their fathers. We obtain a dataset capturing the outcomes of fathers in the earlier census period and the corresponding outcomes of their sons 20 years later. We always restrict the sample to parents and children aged 25-50 in their respective adult census observations.

A key challenge we face in conducting our analysis is that US censuses before 1940 do not report any information on income, with the 1940 census being the only one in our data for which wage income information is available. To overcome this challenge, we employ an imputation procedure developed by Collins and Wanamaker (2022)[7], to construct an individual-level income measure. We regress 1940 (log-) wage income on dummies that interact occupation with state of residence, race[8] and age (measured in 5-year bins), and use the estimated relationship to predict wages for our whole sample based on their state of residence, race, occupation, and age. This gives us a granular measure of predicted wage income for most of our sample. However, this approach of predicting wage income is not a good income measure for farmers, who derive most of their income from non-wage sources and also represent a substantial fraction of our sample. We address this concern by following an imputation procedure for farmers' incomes that also draws on Collins and Wanamaker (2022) and Abramitzky et al. (2021). We use data on the total income of farmers and farm laborers for the year 1960 and calculate the average ratio of farmers' total income to farm laborers' total income within every product of agebin, race and census region cell.[9] We then predict farmer total income by multiplying farm laborers' predicted wage income as calculated above by their cell-specific total income ratio. In the Appendix, we show that our results are robust to also including industry in our wage predictions.

For education, the 1940 census contains an almost ideal measure: years of schooling. We use this to create a restricted sample of father-son-pairs for the intervals 1900 to 1920, and

[7]Similar imputations based on occupation and other variables have also been used by Abramitzky et al. (2021) and Tan (2023). Authors sometimes also adjust income for self-employment in a way similar to our farmer adjustment. We do not do this, as data on self-employment is only available from the 1920 census onward.

[8]Black, White, and Other

[9]We use census regions here, since for 1960 we only have a 5% of the sample and want to avoid having many small cells.

1920 to 1940, where both the father and son are matched to their 1940 outcome. For a 1920-1940 matched pair, this requires that the father is still alive and matched to the 1940 census. For a 1900-1920 pair, it requires that both father and son are still alive and matched in 1940.[10] Additionally, we need to assume that schooling does not change anymore after men enter the labor market. We find this a plausible assumption in our period of analysis.

For modern outcomes, we draw on measures of intergenerational persistence and inequality calculated by Chetty et al. (2014). Using federal income tax data of parents over the years 1996-2002 and their children in 2011 and 2012, Chetty et al. (2014) compute county-specific rank-rank correlations between father and sons, as well as Gini coefficients for the fathers.

3 Methods

3.1 Measuring county mobility and inequality.

Using the measures of income and education described above we proceed to construct measures of intergenerational socioeconomic mobility at the county×interval level (where intervals represent the 1880-1900, 1900-1920 and 1920-1940 periods for income mobility, and the 1900-1920 and 1920-1940 periods for educational mobility). When constructing our measures of intergenerational mobility we assign each father-son pair to the county where they were initially observed sharing the same household (i.e. the county where the sons "grew up"). We then compute our county-level measures of socioeconomic mobility by correlating, for each county, fathers' outcomes in the early year of each interval with sons' outcomes in the late year, separately for income and education.

$$y^s_{ict} = \beta_c + \rho^{IGE}_{ct} \times y^f_{ic,t-20} + \epsilon_{ict} \tag{1}$$

where i denotes the father-son pair, y^s_{ict} is the log of income of the son and $y^f_{ic,t-20}$ is the log of the income of the father twenty years before (in the previous census). ρ_{ct} is the intergenerational elasticity of income of county c at time $t \in \{1900, 1920, 1940\}$.In addition, following Chetty et al. (2014), we also estimate rank-rank correlations. We calculate the father's rank in the national income distribution in his cohort, do the same for the son in his cohort and then correlate the ranks of sons and fathers within a country. Akin to the county-time specific intergenerational elasticity of income, this approach produces a county-

[10]We do not extend this to the 1880-1900 period, as it would give us only those few fathers that were adults in 1880 and still alive in 1940.

time specific rank-rank correlation ρ_{ct}^{RR}. For education, we compute the correlation between the education of parents and sons. In the Appendix, we show that our results are robust to controlling for fathers' and sons' ages in the regressions that produce the county-level persistence measures.

To construct our novel county-level measures of intergenerational mobility, that are suitable for long-run analysis, we note that any period t can be characterised by a distribution of the persistence coefficients ρ across counties. We rank counties by their persistence coefficients for each time period to construct a unitless and time-varying index of socioeconomic persistence. We denote by $rank\rho_{ct} \in [0, 100]$ the county-rank of county c by each of the measures of persistence (ρ_{ct}) at time t. This county-ranking procedure is applied to the intergenerational correlation of income, the intergenerational elasticity of education, and the rank-rank correlation.

To measure inequality, we calculate the (income) Gini coefficient among fathers for each county c and each time t. We also calculate an analogous unitless measure of inequality for the purposes of long-run analysis: the rank of counties by inequality at each point in time. We denote this inequality measure $rankI_{ct} \in [0, 100]$.[11]

3.2 Measures of the GGC

We empirically verify the presence of the GGC by measuring the cross-county correlation between measures of intergenerational persistence at time t and measures of inequality twenty years prior, i.e. among the fathers. We denote the regression coefficient between these moments γ_t. The GGC relationship is then estimated via

$$\rho_{ct} = \alpha + \gamma_t \times I_{c,t-20} + \epsilon_{ct} \tag{2}$$

We also run the same regression with the county-rank measures of persistence $rank\rho_{ct}$ and inequality $rankI_{c,t-20}$ instead of their levels. Likewise, we compute it with our measure of the intergenerational persistence of educational attainment as the dependent variable. We run all these regressions separately for each interval.

[11] In the Appendix, we show robustness to using other measures of inequality such as the variance of log income or the difference between the log of the 90th percentile and the log of the 10th percentile.

3.3 Dynamic GGC

To study how the relationship between inequality and socioeconomic mobility evolves over time we calculate the changes in inequality and persistence for each county at each point in time, as well as the correlation between these changes (which we call the Dynamic GGC):

We define the change in intergenerational persistence and inequality in county c as $\Delta\rho_{ct} = \rho_{ct} - \rho_{c,t-20}$ and $\Delta I_{ct} = I_{ct} - I_{c,t-20}$ respectively; and we run:

$$\Delta\rho_{ct} = \alpha_t + \delta_t \Delta I_{c,t-20} + \epsilon_{ct} \tag{3}$$

where the variable of interest is δ_t, capturing the dynamic relationship between inequality and intergenerational persistence.

Notice that 3 does not allow us to measure the correlation between changes in inequality and changes in persistence if we have different measures of inequality and persistence for the two periods under consideration. Because of this, it is not possible to combine our measures for the 1900-1940 period with modern measures of inequality and mobility, such as those provided by Chetty et al. (2014). We thus employ an alternative methodology to study the long-run relationship between changes in inequality and changes in intergenerational mobility that makes use of our novel country-rank-based measures of inequality and mobility.

This methodology is described formally below:

$$\Delta rank\rho^s_{ct} = rank\rho_{ct} - rank\rho_{cs}; \qquad \Delta rankI^s_{ct} = rankI_{c,t-20} - rankI_{c,s-20}$$
$$\Delta rank\rho^s_{ct} = \alpha^s_t + \delta^s_t \Delta rankI^s_{ct} + \epsilon_{ct} \tag{4}$$

where t and $s > t$ are two periods over which we evaluate the change in persistence and inequality in each county by computing the change in the respective county ranks ($\Delta rank\rho^s_{ct}$ and $\Delta rankI^s_{ct}$ respectively for each county c). The variable capturing the slope of the dynamic "Great Gatsby Curve" over the period being δ^s_t.

To get an intuitive sense of how this methodology allows us to get around the issue of the comparability of measures over time, consider a case in which the available measures of mobility at county level differ across time (for example if our measure of mobility at time t is the IGE while our measure of mobility for time $s > t$ is the rank-rank correlation employed by Chetty et al. (2014)). While these measures are not directly comparable, insofar as they both reflect the same underlying concept of social mobility, we expect that county ranks constructed on the basis of each measure result in similar rankings when derived from data resulting from the same data generating process.[12]

In other words, we expect that, were the high-quality tax data used by Chetty et al. (2014) already available for the early 20th century, we would obtain county rank measures similar to the ones that we calculate based on the census income data. Using this logic, changes in the county level rankings, even if based on different underlying measures of persistence at different points in time, should reflect the relative change in the position of a county's mobility in the overall distribution of US counties. A similar argument holds for our county-rank measures of inequality, which suggests that correlating the changes of county ranks of intergenerational persistence with changes in the county ranks of inequality, is likely to be a valid way of testing for the presence of a long-run dynamic GGC.

4 Results

In this section, we methodically present and demonstrate our main results. Here we refer only to the baseline sample and definitions, but we want to point out that in the Appendix we reproduce the same qualitative results in a multitude of robustness checks with alternative definitions, and methodologies.

Result 1 *The intergenerational persistence of both income and education is positively correlated with inequality across US counties during the period 1900-1940, as it is nowadays.*

Table 1 and Figure 1 outline our findings concerning the cross-sectional relationship between intergenerational socieconomic persistence and income inequality at the level of US counties and at different points in time. The first line of Table 1 employs standard measures of socioeconomic mobility (the IGE and the rank-rank coefficient for income mobility, the intergenerational persistence of educational attainment for education) and inequality (the income Gini coefficient), while the second row employs our novel measures of socioeconomic persistence (the county ranks of persistence coefficients and Gini indices). Moreover, the first seven columns outline our results concerning the relationship between the intergenerational persistence of income and income inequality, while the last two columns present our findings regarding the relationship between intergenerational educational persistence and inequality.

[12] In the `online appendix`, we show that the county rank measure strongly correlates with the underlying measure of persistence in 1900, 1920, and 1940. It thus captures the same variation as the underlying measurement, but allows us to generate measures that are comparable over time even when the underyling measures are not.

Across all specifications, we recover the positive association between intergenerational socioeconomic persistence and inequality that has been documented elsewhere for more recent time periods. As it is the case nowadays, one hundred years ago more unequal US counties tended to display more intergenerational persistence in terms of both income and education, irrespective of the measures of intergenerational persistence we use. Indeed, the strength of the association seems if anything stronger for the first half of the twentieth century than for the more recent past[13], though caution is advised when interpreting coefficient magnitudes given the limitations of our historical income measures.

We believe that this is the first documentation for the GGC in an historical setting, and indicates that the relationship is extremely robust to the passage of time. Thus, the hypothesis that upward mobility correlates with inequality is not only rejected today, this relationship was already not holding during the Great Gatsby era.

Result 2 *The static relationship between inequality and socioeconomic mobility is qualitatively similar when employing our novel measures of inequality and mobility using county ranks.*

Reassuringly, our results concerning the Great Gatsby Curve are qualitatively similar when we employ our novel methodology based on county ranks (see second row of Table 1). Both in the early twentieth century and closer to the present, counties that ranked highly in terms of inequality also tended to rank highly in terms of intergenerational socioeconomic persistence, irrespective of the underlying measures of persistence used to construct the ranking. The positive association between inequality and intergenerational persistence identified using the new measures applies for both income and education. As in the analysis with more established measures, it seems to be stronger in the first half of the twentieth century than in the more recent past. Overall, we interpret these results as evidence supporting the validity of our novel measures for the analysis of the relationship between inequality and socioeconomic mobility.

Result 3 *For the first half of the twentieth century, the correlation between changes in inequality and changes in the intergenerational persistence of income across US counties is not always positive (i.e. the "Dynamic" GGC is unstable over this period). By contrast, the correlation between changes in inequality and changes in the intergenerational persistence of education remain positive.*

[13]This can be seen by comparing coefficient magnitudes in columns 4 to 6 in Table 1 to those in column 7

Table 2 outlines our results regarding the relationship between *changes* in inequality and *changes* in intergenerational socioeconomic persistence (which we call the "Dynamic" GGC). Similarly to our discussion of the "Static" GGC, the first row of Table 2 presents our results employing established measures of income inequality (i.e. the Gini Coefficient) and socieconomic mobility (the IGE and the rank-rank correlations between fathers and sons for the intergenerational persistence of income, and the father-son correlation in educational attainment for the intergenerational persistence of education), while the second row of the table presents results employing our novel measures of inequality and intergenerational socioeconomic mobility using county ranks. Moreover, columns 1 to 6 of Table 2 present our findings concerning the dynamic relationship between inequality and the intergenerational persistence of income, while column 7 documents our findings for the relationship between inequality and the intergenerational persistence of education.

Perhaps our most striking finding is that the Dynamic Great Gatsby Curve for income is unstable across periods lasting two decades, as can be seen in columns 1 to 4 and in figures 2a, 2b. Over the period 1900 to 1920, changes in inequality at the level of US counties, correlate positively, albeit more weakly in magnitude, with changes in intergenerational income persistence. This pattern holds whether the intergenerational income persistence is measured by the IGE or by father-son rank-rank correlations. In other words, over this period, the Dynamic GGC is upward sloping and hence qualitatively similar (albeit flatter) to the Static GGC that was documented above and elsewhere in the literature. By contrast, this familiar pattern is absent for the period 1920 to 1940. During this period, we find that there is no relationship between changes in inequality and changes in the intergenerational persistence of income across US counties when we measure the intergenerational persistence of income by the rank-rank correlations. We find that the relationship is even negative when we measure the intergenerational persistence of income by the IGE. These findings suggest that the dynamic relationship between inequality and intergenerational persistence can deviate from the static one even for relatively long time periods on the order of two decades.

For the relationship between changes in inequality and changes in the intergenerational persistence of education, data limitations restrict our analysis to the period 1920 to 1940. For this period we find that changes in inequality correlate positively with changes in the intergenerational persistence of educational attainment. In other words, the Dynamic GGC for education mimics the Static GGC and is robust even during time periods where the positive relationship between changes in inequality and changes in the intergenerational persistence of income is absent in the data. In addition to Table 2 this can be seen in Figure

2c

Result 4 *Evaluating the correlation between changes in inequality and changes in socioeconomic persistence using our novel measures based on county ranks produces the same results.*

The results in the second row of Table 2 (that can be visualized comparing figures 2a and 2d) provide further reassurance regarding the reliability of our novel measures in inequality and socioeconomic mobility based on county ranks. Our findings using these measures closely track our previous results using more established measures of inequality and mobility. We confirm the instability of the Dynamic GGC during the first half of the twentieth century. The correlation between changes in the county ranks in terms of inequality and changes in the county ranks in terms of intergenerational income persistence is found to be positive (albeit weaker than in the case of the Static GGC) for the period 1900-1920, and insignificant or even negative, depending on the underlying measure of intergenerational persistence employed, for the period 1920 to 1940. Moreover, we also confirm an upward sloping Dynamic GGC for education over the period 1920 to 1940, in line with our previous findings. Overall, we interpret this pattern of findings as further validation of our novel measures of inequality and intergenerational socioeconomic mobility, as they are able to capture the same patterns in the data as more established measures even during time periods of instability in the empirical relationship between our moments of interest.

Result 5 *Changes in county ranks of intergenerational income persistence correlate positively with changes in the county ranks of inequality over the periods 1920-2011 and 1940-2011*

Lastly, we use our novel measures of inequality and socieconomic mobility to study the long-run association between changes in inequality and changes in the intergenerational persistence of income (we deem this relationship the "Long-run Dynamic GGC"). Our results covering the periods 1920 to 2011 and 1940 to 2011, respectively, are presented in columns 5 and 6 of table 2 and Figure 3. For both columns our underlying measure of the intergenerational persistence of income is the father-son rank-rank correlation.

For both periods of interest we find that changes in a county's rank in terms of inequality correlate positively with changes in the county's rank in terms of the intergenerational persistence of income. Quantitatively, the association seems a bit stronger over the longer period 1920 to 2011, where the slope of the Dynamic GGC is about two thirds that of the Static GGC identified with the same measures of inequality and mobility.

All in all, over this long time periods we recover the familiar upward sloping GGC that has been documented in a variety of cross-sectional settings.

This finding suggests a complex relationship between inequality and intergenerational persistence: the relationship seems to be robustly positive over long time periods (indeed we can also think of the Static GGC as a long-run or "steady-state" type relationship) but can break down over shorter time periods on the order of a couple of decades.

5 Conclusion

In this paper we have used recently made available linked census data for the US to revisit the relationship between inequality and socioeconomic mobility - the so-called "Great Gatsby Curve". Our findings point towards a subtle relationship between inequality and mobility. While the positive association between inequality and the intergenerational persistence of income documented elsewhere is also a robust regularity in our data, the relationship between *changes* in inequality and *changes* in the intergenerational persistence of income is more complex: the correlation is positive for the period 1900 to 1920, zero or even negative for the period 1920 to 1940, and positive again for the long periods 1920 to 2011 and 1940 to 2011. By contrast, the association between inequality and the intergenerational persistence of educational attainment is positive in both levels and changes for all the periods that we were able to study.

We interpret our findings as consistent with mechanisms that yield a positive association between inequality and the intergenerational persistence of socioeconomic status over the long-run (i.e. we interpret the static GGC as a long-run or steady-state relationship). However, over the medium-run these mechanisms may be muted by the presence of large shocks such as the Great Depression or the World Wars. Moreover, it is reasonable to argue that the relationship between inequality and the intergenerational persistence of educational attainment, which is very robust, may itself be driving the long-run positive association between inequality and the intergenerational persistence of income. Thus, the serendipity of individual income determination would explain the unstable short-run dynamics, while the process of human capital accumulation would generate the stable long-run relationship. However, more research is required both to verify the robustness of our findings in other settings and to uncover the specific economic mechanisms connecting human capital accumulation with the dynamics of the Great Gatsby Curve.

- An online appendix can be found at: `This link`.

References

Abramitzky, R., Boustan, L. P., Eriksson, K., Pérez, S., and Rashid, M. (2020). Census linking project.

Abramitzky, R., Boustan, L. P., Jácome, E., and Pérez, S. (2021). Intergenerational Mobility of Immigrants in the US over Two Centuries. *The American Economic Review*, 111(2):580–608.

Acemoglu, D. and Autor, D. H. (2011). Skills, Tasks and Technologies: Implications for Employment and Earnings. In Card, D. and Ashenfelter, O., editors, *Handbook of Labor Economics*, volume 4B, pages 1043–1171. Elsevier.

Andrews, D. and Leigh, A. (2009). More Inequality, less Social Mobility. *The Review of Economic Studies*, 16(15).

Autor, D. H., Katz, L. F., and Kearney, M. (2008). Trends in U.S. Wage Inequality: Revising the Revisionists. *Review of Economics and Statistics*, 90(2):300–323.

Bjorklund, A. and Jäntti, M. (2012). Intergenerational Income Mobility and the Role of Family Background. *Oxford Handbook of Economic Inequality.*

Blanden, J. (2013). Cross-Country Rankings in Intergenerational Mobility: A Comparison of Approaches from Economics and Sociology. *Journal of Economic Surveys*, 27(1):38–73.

Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of Opportunity? The Geography of Intergenerational Mobility in the United States *. *The Quarterly Journal of Economics*, 129(4):1553–1623.

Collins, W. J. and Wanamaker, M. H. (2022). African American Intergenerational Economic Mobility Since 1880. *American Economic Journal: Applied Economics*, 14(3):84–117.

Corak, M. (2006). Do Poor Children Become Poor Adults? Lessons from a Cross-Country Comparison of Generational Earnings Mobility. In Creedy, J. and Kalb, G., editors, *Dynamics of Inequality and Poverty*, volume 13 of *Research on Economic Inequality*, pages 143–188. Emerald Group Publishing Limited.

Corak, M. (2013). Income Inequality, Equality of Opportunity, and Intergenerational Mobility. *Journal of Economic Perspectives*, 27(3):79–102.

DiPrete, T. A. (2020). The impact of inequality on intergenerational mobility. *Annual Review of Sociology*, 46(1):379–398.

Durlauf, S. N., Kourtellos, A., and Tan, C. M. (2022). The Great Gatsby Curve. *Annual Review of Economics*, 14(1):571–605.

Durlauf, S. N. and Seshadri, A. (2018). Understanding the Great Gatsby Curve. *NBER Macroeconomics Annual*, 32:333–393.

Ermisch, J., Jäntti, M., and Smeeding, T., editors (2012). *From Parents to Children: The Intergenerational Transmission of Advantage*. Russell Sage Foundation.

Fan, Y., Yi, J., and Zhang, J. (2021). Rising Intergenerational Income Persistence in China. *American Economic Journal: Economic Policy*, 13(1):202–230.

Güell, M., Pellizzari, M., Pica, G., and Rodríguez Mora, J. V. (2018). Correlating Social Mobility and Economic Outcomes. *The Economic Journal*, 128(612):F353–F403.

Hassler, J., Mora, J. V. R., and Zeira, J. (2007a). Inequality and Mobility. *Journal of Economic Growth*, 12(3):235–259.

Hassler, J., Mora, J. V. R., and Zeira, J. (2007b). Inequality and Mobility. *Journal of Economic Growth*, 12(3):235–259.

Helgertz, J., Ruggles, S., Warren, J. R., Fitch, C. A., Hacker, J. D., Nelson, M. A., Price, J. P., Roberts, E., and Sobek, M. (2023). IPUMS Multigenerational Longitudinal Panel: Version 1.1 [dataset].

Piketty, T. and Saez, E. (2003). Income Inequality in the United States, 1913-1998. *The Quarterly Journal of Economics*, 118(1):1–39.

Ruggles, S., Fitch, C. A., Goeken, R., Hacker, J. D., Nelson, M. A., Roberts, E., Schouweiler, M., and Sobek, M. (2021). IPUMS Ancestry Full Count Data: Version 3.0 [dataset].

Tan, H. R. (2023). A different land of opportunity: The geography of intergenerational mobility in the early twentieth-century united states. *Journal of Labor Economics*, 41(1):77–102.

Tables and Figures

	(1) IGE	(2) IGE	(3) IGE	(4) r-r	(5) r-r	(6) r-r	(7) r-r	(8) Edu	(9) Edu
County Level	0.82***	0.80***	0.68***	0.86***	0.72***	0.57***	0.24***	0.57***	0.65***
	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.03)	(0.02)	(0.08)	(0.03)
County Rank	0.47***	0.40***	0.43***	0.48***	0.39***	0.36***	0.33***	0.14***	0.45***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Num. Counties	2,275	2,661	3,006	2,275	2,661	3,006	2,769	2,439	3,055
Year	1900	1920	1940	1900	1920	1940	2010	1920	1940

Robust std errors in parethenses.

Table 1: Static Great Gatsby Curves

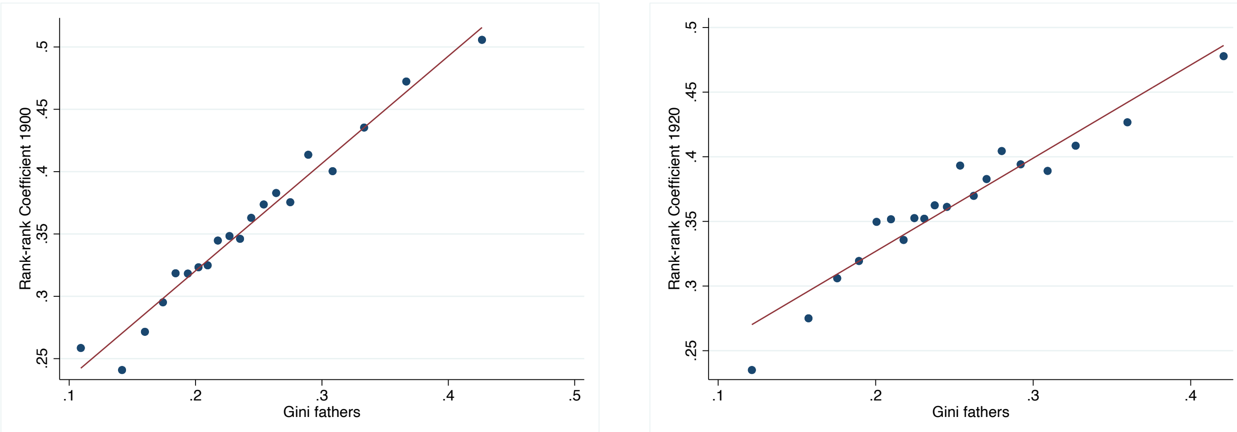
Each coefficient comes from a separate regression. The first row correlates the county persistence and the county inequality levels. Inequality is always measured as the Gini index of the income of the parents (i.e., measured in the census 20 years prior to the measurement of the son's income). Data for 1900-1940 are based on our income imputations, data for 2011 are from Chetty et al. (2014) and refer to parents' income measured in 1996-2002 and children's in 2011/12. Persistence is measured in each county either by IGE (columns 1, 2, 3), the correlation of the rank of the father with the rank of the son (4, 5, 6, 7), or the correlation of years of education of fathers and sons (8,9).
The second row correlates the rank of the county in the persistence distribution with the county rank in the inequality distribution. The measures of persistence and inequality are the same in the previous raw.
In the `online appendix` we show binned scatter plots and scatter plots with the row data corresponding to each entry in the table.

	(1) ΔIGE	(2) ΔIGE	(3) Δr-r	(4) Δr-r	(5) Δr-r	(6) Δr-r	(7) ΔEducation
Δ County Levels	-0.23***	0.11*	-0.11	0.26***			0.64***
	(0.06)	(0.06)	(0.10)	(0.10)			(0.14)
Δ County Rank	-0.07**	0.04	-0.05	0.12***	0.28***	0.18***	0.27***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)	(0.04)
Observations	2,657	2,271	2,657	2,271	2,448	2,688	2,437
Year	1940/1920	1920/1900	1940/1920	1920/1900	2010/1920	2010/1940	1940/1920

Robust std errors

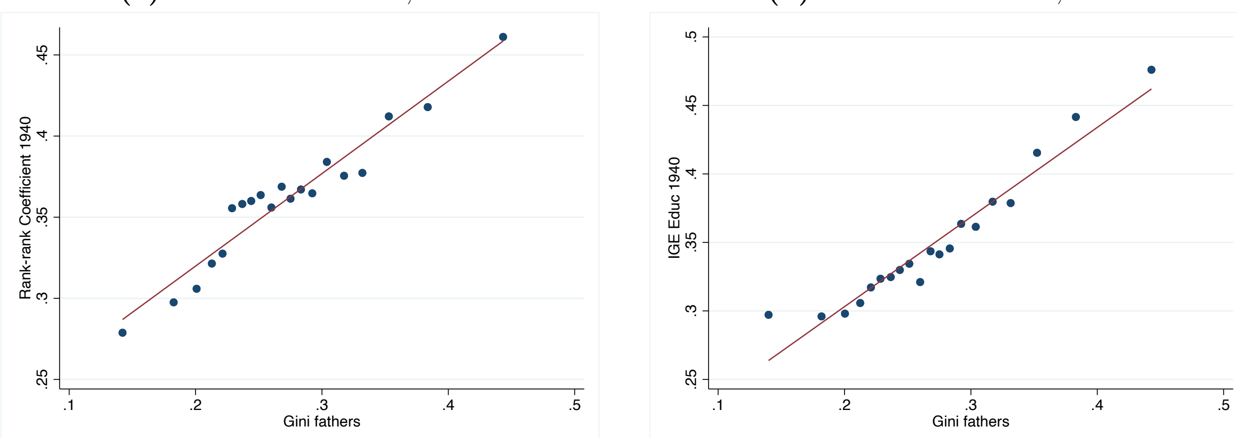
Table 2: Dynamic Great Gatsby Curves.

Each coefficient comes from a separate regression. The first row correlates changes in a county's persistence with changes in a county's inequality. Inequality is always measured as the Gini index of the income of the parents (i.e., measured in the census 20 years prior to the measurement of the son's income). Data for 1900-1940 are based on our income imputations, data for 2011 are from Chetty et al. (2014) and refer to parents' income measured in 1996-2002 and children's in 2011/12. Persistence is measured in each county either by IGE (columns 1, 2), the correlation of the rank of the father with the rank of the son (3, 4, 5, 6, 7), or the correlation of years of education of fathers and sons (7).
The second row correlates the rank of the county in the persistence distribution with the county rank in the inequality distribution. The measures of persistence and inequality are the same in the previous row
In the `online appendix` we show binned scatter plots and scatter plots with the row data corresponding to each entry in the table.



(a) Levels. r-r vs Gini, 1900

(b) Levels. r-r vs Gini, 1920

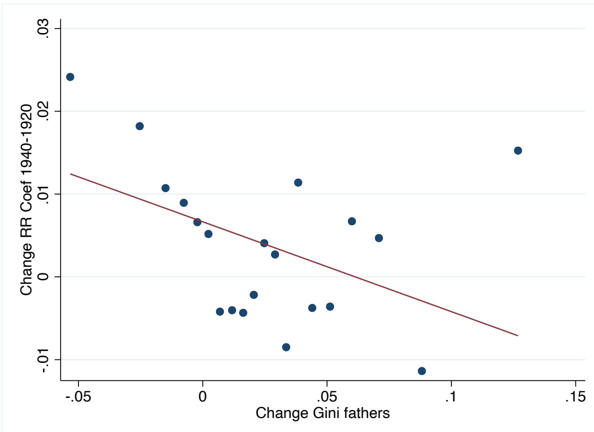


(c) Levels. r-r vs Gini, 1940

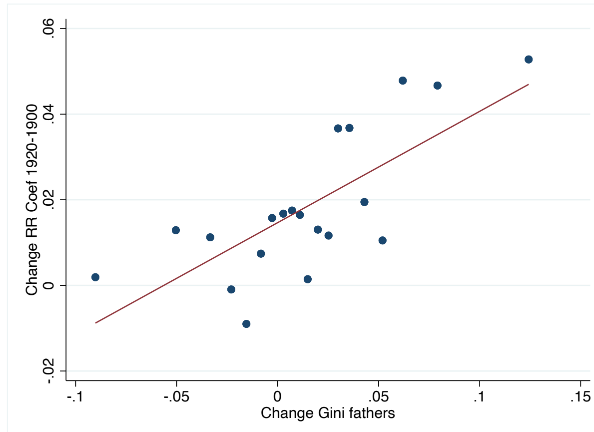
(d) Levels. Education vs Gini, 1940

Figure 1: Static Great Gatsby Curve 1900-1940

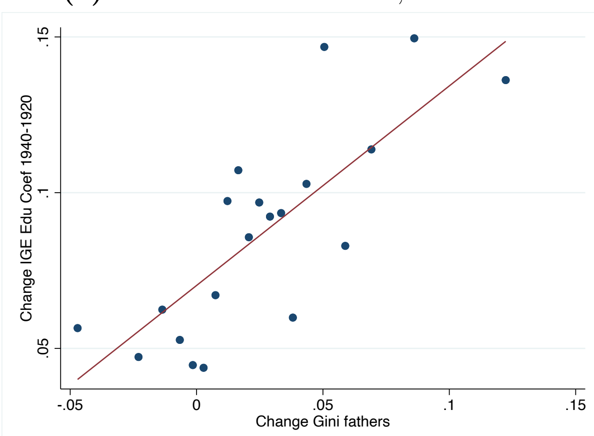
Binned scattered plots. In each, we divide the counties in vintiles of the level of the Gini Index and plot the average measure of intergenerational persistence of the counties in the vintile.



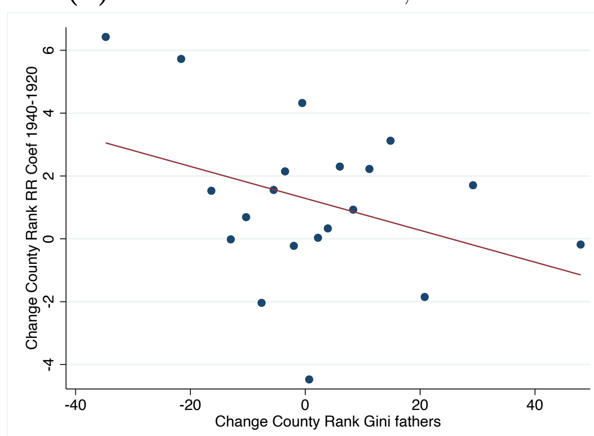
(a) Levels. Δr-r vs ΔGini, 1920-1940



(b) Levels. Δr-r vs ΔGini, 1900-1920



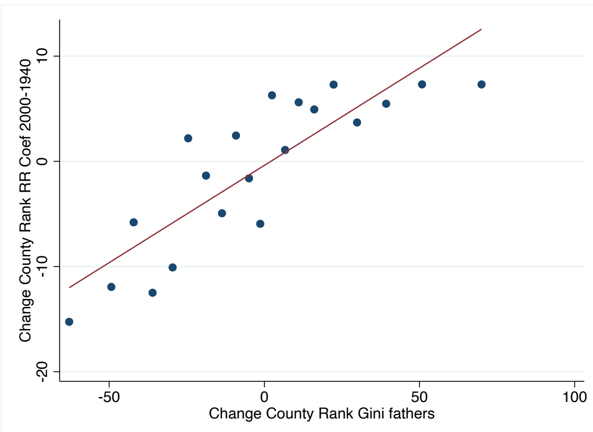
(c) Levels. Δ Education vs ΔGini, 1900-1920



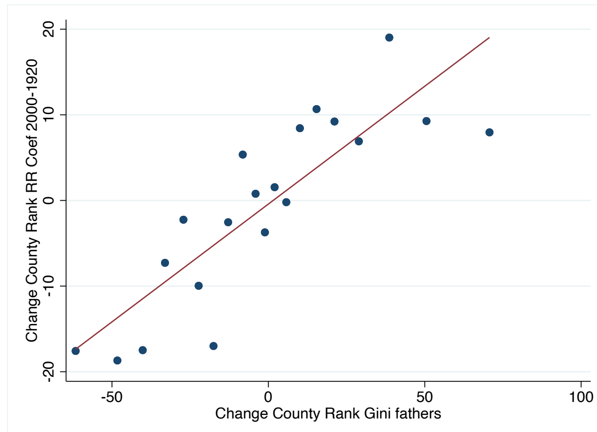
(d) County Ranks. Δr-r vs ΔGini, 1920-1940

Figure 2: Dynamic GGC.

Binned scattered plots. In each, we divide the counties in vintiles of the growth of the county rank of income inequality and plot the average growth of the county rank of persistence of the counties in the vintile.



(a) County Ranks. Δr-r vs ΔGini, 1920-2011



(b) County Ranks. Δr-r vs ΔGini, 1940-2011

Figure 3: Dynamic GGC, r-r County Ranks.

Binned scattered plots. In each, we divide the counties in vintiles of the growth of the county rank of income inequality and plot the average growth of the county rank of persistence of the counties in the vintile.