MURAJ
Minnesota Undergraduate Research & Academic Journal

z.umn.edu/MURAJ

Reconsidering The American Dream: A Statistical Analysis

By Kevin Jacobson[1]

[1] *Heller-Hurwicz Economics Institute, Department of Economics*

Abstract: The United States has observed appreciable GDP growth since the 1940's, both in aggregate and per-capita terms. Accompanying advancements in technology, productivity, and social justice, one might expect a promising future for the next generation. After all, upward social mobility is the national ethos of the U.S., better known as the American Dream. At a time when wage stagnation and income inequality are becoming forefront concerns for young Americans, this paper seeks to determine if the American Dream is a living promise or a naive ideal. Specifically, this study uses statistical methods and historical data to estimate the effect of childhood socioeconomic status on future wages, conditional on a set of controls. Indeed, we find that parental income is a statistically significant predictor, with a p-value of 0.033 for the full sample and 0.007 for the low socioeconomic status (SES) sample. Our findings suggest that income inequality will continue to increase in the future, raising concerns about sustained GDP growth.

Introduction

The American Dream is the belief that anyone can achieve upward mobility through hard work and perseverance, regardless of where you are born. This hallmark of Western society is meant to inspire the new generation to achieve a higher standard of living than the previous. When examining per-capita GDP growth since the mid-1900's as shown in **Figure 1**, this phenomenon seems entirely realistic. As a country, the U.S. has become wealthier and more productive with each generation since WWII, corresponding to a consistent upward trend in GDP growth. However, this prosperity has not trickled down to the lowest divisions of the income distribution, as upward social mobility is becoming a more prominent challenge for children of poorer families [2]. Over time, such inequality in opportunity seems to have calcified the boundaries between income classes, with measures of income inequality continuing to climb [3]. This prompts a critical question:

As GDP continues to climb in the United States, are children born into high-socioeconomic-status households more likely to outperform their parents in income?

Literature Review

This paper's question relates to the nature of economic mobility, income inequality, and GDP growth. As economist Robert Rycroft does in his book *The Economics of Inequality, Discrimination, Poverty, and Mobility*, we can liken the GDP of the country to a pie and slice it among its constituents [2]. Each year, this pie (GDP) can grow, but the share of the pie may change over time, representing increasing inequality. We can think of economic mobility as the ability to change the size of your slice over a career from the slice your parents had.

In this illustration, human capital theory (HCT) provides some justification for why slices differ in size. After all, abstract belief in HCT provides motivation for post-secondary education. Specifically, HCT helps to explain the role that skill development, education, and merit play in the labor market. As economist Gary Becker discusses in his appropriately titled book, *Human Capital*, HCT is an extension of neoclassical wage determination, wherein workers receive higher pay by way of forming skills, usually through education and training [3]. As such, this study carefully accounts for the effect education of education on future earnings in the model.

While there are several economic theories behind income inequality, the data-driven work of Raj Chetty seems to be the most compelling [4]. His paper demonstrates significant differences in earnings and future earning opportunities between different ethnic groups as well as between men and women. To account for these findings, our main model controls for demographics.

Having considered different factors that influence income inequality, we can now incorporate its relationship to GDP growth. A U.S.-centered understanding of macroeconomic variables may lead one to believe that GDP growth and income inequality necessarily go hand-in-hand. However, research conducted by the Federico Cingan at the Organisation of Economic Development (OECD) tells a more complete story [5]. Cingan analyzes GDP growth against movements in the Gini coefficient, an index that measures inequality across the income distribution, for OECD countries during the timeframe 1980-2010. He finds that inequality has less to do with booming income shares in the top quartile, and more to do with stagnant earnings at the bottom. For example, in times of GDP growth in the U.S., the bottom does not grow as fast as the top and may fall quicker during times of recession. Thus, GDP grows in developed countries not *because* of inequality, but *despite* it. In fact, Cingan finds that income inequality has a statistically significant *negative* effect on growth.

Cingan's econometric findings suggest that income inequality stunts potential growth, and this does have robust backing in the available literature. Harvard economist Robert Borro analyzes the stress inequality places on GDP growth from the lens of four well-established economic theories: credit market imperfections, social unrest, political economy, and savings rates [6]. Further, the empirical work of Persson and Tabellini agrees with Cingan and their growth model confirms GDP tends to slow as wealth accrues to the top of the income distribution [7]. On account of the combined empirical and theoretical frameworks found among these studies, we can predict how future

GDP will respond to inequality at different levels of intensity.

Data

This study incorporates two primary data sources from the National Center of Educational Statistics: the Educational Longitudinal Study (**ELS**) and the High School Longitudinal Study (**HSLS**)[1][8][9]. These large-sample questionnaires follow different cohorts of high school students through graduation and into early adulthood. Each dataset contains detailed information about academic performance, demographics, parental income, and personal career outcomes.

The ELS respondents were first surveyed in 2002 as high school sophomores, with biennial follow-up questionnaires until 2012 - eight years after high school. The following variables were extracted for this study: ethnicity, gender, parental income in 2002, parent's highest degree of education, personal income in 2012, personal educational attainment, and GPA from the respondent's most recent institution. This dataset is used to estimate the coefficients of the model.

The HSLS began in 2009 for high school freshmen. Since this survey has a similar structure to ELS, the same variables listed above were ascertained. Unlike ELS, HSLS is ongoing, but has not released public data since 2017. Thus, the purpose of this second dataset is not to fit coefficients, but to apply them. Specifically, the ELS-fitted coefficients are used to estimate the probability of upward mobility by the year 2021 for different income classes within the HSLS sample. Once the 2021 HSLS data is released, it would be interesting to see how these predictions compare to the outcomes.

Data Manipulation

To ensure the coefficients of the model retain meaning between the ELS and HSLS cohorts, a few key modifications to the datasets were made. First, all income values were converted to 2012 dollars. Then, each demographic and background variable were transformed into a series of dummy variables, according to some baseline. Ethnicity, for example, consists of two binary variables: "Black" and "Hispanic," and a white respondent will have Black = 0 and Hispanic = 0. Parental and personal educational variables function much in the same way, both of which have high school graduate as their baseline value. Finally, observations with missing responses were dropped to prevent unwanted imputation.

Theory

From the literature discussed in **Section 2**, there are two crucial factors to consider if we wish to determine the likelihood of upward mobility. To remain in lockstep with Chetty's work, we need to account for race and gender demographics [4]. Second, we need to establish the role of human capital in determining professional earnings and consider what variables may influence human capital development.

To address this first concern, we placed the string of binary demographic variables described in **Section 3**. While a series of binary controls is more demanding for regression analysis compared to one or two nominal

1 For data summary tables, reference the **Appendix**

variables, the number of observations of the ELS sample is not nearly small enough for us to worry about potential divergence. Further, binary controls have more easily interpreted coefficients.

The second concern requires us to identify potential mechanisms by which human capital investment affects future income. There is the obvious pathway of improved career prospects and annual compensation after completing a degree or certification. Furthermore, it is arguable that the parent's income and education level affect the child's education level. Higher income families have greater means to invest in early human capital development, such as pre-K programs that can cultivate important cognitive and social skills at an early age. Additionally, the parent's level of education may inform a set of expectations they have for their child. These assertions are empirically supported by the existing literature [10][11]. Because of these associations, it would be ill-informed to say that these variables are all pairwise independent. As such, the main model is constructed to limit the number of unexpected interactions between related variables, as illustrated in **Figure 2**.

This paper will use a logit regression with binary random variable $\mathrm{Y_i}$ for our dataset of N individuals, where

$$\mathrm{Y_i} = \begin{cases} 0, & \text{if individual } i \text{ earns more than parent} \\ 1, & \text{if individual } i \text{ earns less than parent} \end{cases}$$

and $\mathrm{i} \in \{1,2, \dots, N\}$. In addition, we have

$$\ell_\mathrm{i} = \ln\left(\frac{\mathrm{p_i}}{1-\mathrm{p_i}}\right) = \beta_0 + \beta_1 \mathrm{x}_{\mathrm{pi_i}} + \beta_2 \mathrm{x}_{\mathrm{GPA_i}} + \boldsymbol{\beta}\mathbf{X_i} \ (1)$$

where ℓ_i is person *i*'s log-likelihood propensity to earn more than their parents (i.e., for event $\mathrm{Y_i} = 1$), $\mathrm{p_i}$ is their probability of upward mobility[2], $\mathrm{x_{pi}}$ is parental income, $\mathrm{x_{GPA}}$ is the student's GPA from their most recent institution and $\mathbf{X_i}$ is a vector of controls, with corresponding linear parameters stored in $\boldsymbol{\beta}$. To solve for the probability of upward mobility:

$$\mathrm{p_i} = \frac{1}{1+e^{-\ell_\mathrm{i}}} = \frac{1}{1+e^{-(\beta_0+\beta_1 \mathrm{x}_{\mathrm{pi_i}}+\beta_2 \mathrm{x}_{\mathrm{GPA_i}}+\boldsymbol{\beta}\mathbf{X_i})}} \ (2)$$

Results

Fitting our main model (**Equation 2**) with the ELS dataset, and iterating these estimations for a high socioeconomic status (above mean parental income) and a low socioeconomic status sample, we obtain the following[3]:

All GPA coefficients in **Table 1** have statistical and practical significance. For each income group, a unit increase in GPA predicts an increase in the predicted likelihood of upward mobility between 0.176 and 0.181.

Parental income is significant at the full sample and low-SES sample, which is in agreement with the existing literature. However, it is not significant at the high-SES level. That is, the association between parental income and the probability of upward mobility may be zero at the high-income population level. Given the model's construction, this is reasonable.

To illustrate the potential insignificance of parental income within the high-SES

2 $\mathrm{p_i}$ can also be thought of as the expected value of $\mathrm{Y_i}$.

3 For the observed control coefficients, reference **Table 5** in the **Appendix**

subsample, imagine two theoretical children in the sample: one in a top 10% income-earning household, the other in the 1%. Assume the parents from both houses invested in a quality pre-K program for their child. While the 1% household can afford more, their child will also have to earn much more than the 10% for there to be upward mobility. As such, the sign and significance of parental income at the high-SES level is ambiguous.

Using these coefficients[4], we can estimate the likelihood of mobility for the HSLS sample, where post-college income has not yet been observed. To offer more precision, we can break up these estimates by income decile.

Table 2 estimates the probability of HSLS respondents earning more than parents and earning above average (**P(Upward Mobility)** and **P(Earning Above Average),** respectively). These outcomes have been observed for the ELS survey and are used to train **Eqn. (2)** for each decile of parental income in the HSLS data.

For the bottom three deciles, this study estimates a high probability of upward mobility but low probability of earning income above average. This is intuitive, as the salary they require to earn more than their parents is relatively low, but they are still more likely than not to earn salaries below the mean. The opposite relationship is seen for the top two deciles. The median respondents (50th percentile) have less than .5 probability of upward mobility and earning above average. This is significant because it indicates the median person is unlikely to observe upward mobility or achieve an above average salary.

While the probability of earning above-mean income is strictly increasing by HSLS decile, it does not exceed 0.5 until the 70th percentile. That is, individuals in the 60th percentile and below are more likely than not to earn less than the average salary, indicating the persistence of increasing income inequality.

Conclusion

This paper explored the relationship between GDP and inequality based on childhood socioeconomic status. Based on the results of this paper, childhood socioeconomic status greatly determines future success and upward mobility. Furthermore, GPA matters less for the high-SES subsample, indicating perhaps a greater pool of connections or resources to gain more lucrative employment. In the end, the surprisingly low likelihood for individuals to earning above-mean income (below 50% for all but the top three income deciles) in the HSLS forecasting suggests that inequality may continue to increase, and as mean income increases with per-capita GDP, this could spell threats to large and sustained levels of growth in the future. Perhaps the American Dream is not dead, but it is certainly not a birthright. Hopefully, faith in this ideal helps us to realize this dream rather than discourage criticism over the extent of its promises.

4 As with the ELS, this study partitions the HSLS respondents into high-SES and low-SES samples, using the corresponding coefficients from **Table 1** to estimate probabilities.

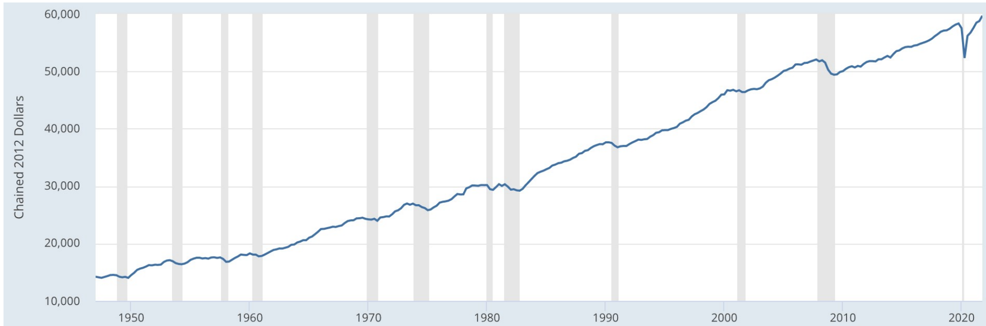

***Figure 1*. Time Series of GDP per-capita in the United States [1]**

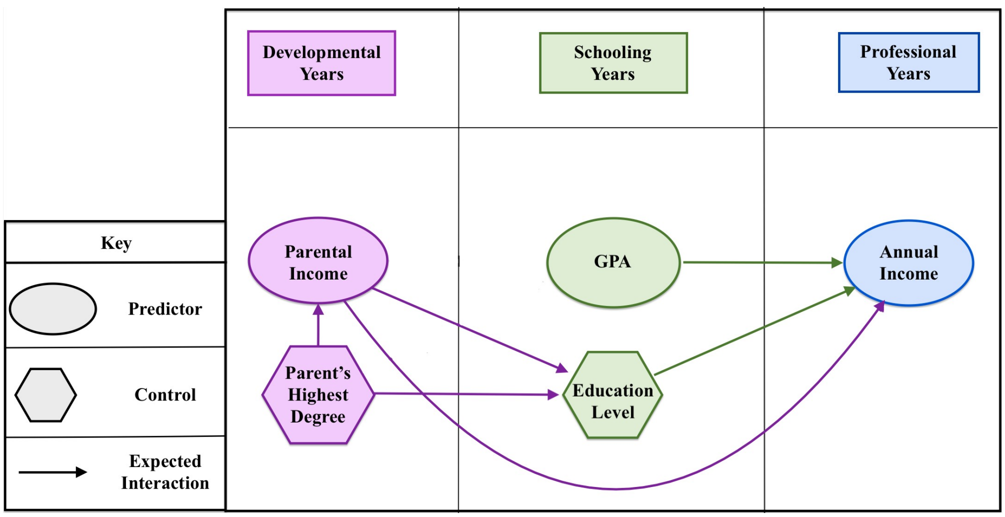

***Figure 2*. Diagram of Main Model**

Table 1. **Fitted Coefficients for Explanatory Variables**

	Full Sample	Low-SES	High-SES
Parental Income	0.104**	0.097***	0.0187
	(0.049)	*(0.034)*	*(0.052)*
GPA	0.180**	0.181**	0.176*
	(0.098)	*(0.083)*	*(0.091)*
Intercept	2.17***	1.52***	3.16***
	(0.307)	*(0.452)*	*(0.580)*
N	5326	3036	2290

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Note: parenthetical values represent standard error.

Table 2. **Likelihood of Earnings Outcomes for 2013 Graduates**

Parent's Income Level	**P(Upward Mobility)**	**P(Earning Above Average)**
10*th* **Percentile**	0.841	0.289
20*th* **Percentile**	0.794	0.304
30*th* **Percentile**	0.728	0.342
40*th* **Percentile**	0.612	0.420
50*th* **Percentile**	0.423	0.482
60*th* **Percentile**	0.537	0.492
70*th* **Percentile**	0.520	0.627
80*th* **Percentile**	0.364	0.783
90*th* **Percentile**	0.232	0.879

References

[1] *Real gross domestic product per capita*. FRED. (2022, January 27). Retrieved February 22, 2022, from https://fred.stlouisfed.org/ series/A939RX0Q048SBEA.

[2] Rycroft, R. S. (2018). *The Economics of Inequality, Discrimination, Poverty, and Mobility*. Routledge.

[3] Becker, G. S. (2008). *Human capital: A theoretical and empirical analysis, with special reference to Education.* The University of Chicago Press.

[4] Chetty, R. et. al. (2019). *Race and Economic Opportunity in the United States*. Opportunity Insights. Retrieved 23 February 22 2022, from https://opportunityinsights.org/wp-content/uploads /2018/04/race paper.pdf.

[5] Cingan, Federico (2014). "Trends in Income Inequality and its Impact on Economic Growth." *OECD Economics Department Working Papers*, OECD Publishing. Accessed 20 March 2022.

[6] Barro, R. J. (2000). "Inequality and Growth in a Panel of Countries." *Journal of Economic Growth*, 5(1), 5–32. http://www.jstor.org/ stable/40216021.

[7] Persson, Torsten, and Guido Tabellini. "Is Inequality Harmful for Growth?" *The American Economic Review*, vol. 84, no. 3, 1994, pp. 600–621. JSTOR, http://www.jstor.org.ezp1.lib.umn.edu/ stable/2118070. Accessed 22 Feb. 2022.

[8] National Center for Education Statistics. "Education Longitudinal Study of 2002 (ELS:2002)" [Data set]. Available: https://nces.ed.gov/surveys/els2002/. Accessed 22 February 2022.

[9] National Center for Education Statistics. "High School Longitudinal Study of 2009 (HSLS:09)" [Data set]. Available: https://nces.ed.gov/surveys/hsls09/. Accessed 22 February 2022.

[10] Guryan, J., Hurst, E., & Kearney, M. (2008). "Parental education and parental time with children." *Journal of Economic Perspectives*, 22(3), 23–46. https://doi.org/10.1257/jep.22.3.23

[11] Heckman, J. J. (2008). "Role of income and family influence on child outcomes." *Annals of the New York Academy of Sciences*, 1136(1), 307–323. https://doi.org/10.17226/25246

[12] Bolyard, Atticus and Peter Savelyev "Understanding the Educational Attainment Polygenic Score and Its Interactions with SES in Determining Health in Young Adulthood." *Human Capital Economic Opportunity Working Group*, The University of Chicago. Accessed 20 March 2022.

[13] Bureau of Labor Statistics. "CPI Seasonal Adjustment Tables" [Data set]. Available: https://www.bls.gov/cpi/tables/seasonaladjustment/home.htm. Accessed 20 March 2022

Appendix

Table 3. **Summary Statistics - (ELS)**

	N	Mean	Min	Max
Explanatory Variables				
Parental Income	5326	$42831	$10000	$200000
GPA	5326	2.78	0.2	4.00
Parent Degree				
< GED	5326	0.061	0	1
Associate	5326	0.104	0	1
Bachelor	5326	0.226	0	1
Master	5326	0.117	0	1
Doctoral/Professional	5326	0.069	0	1
Own Degree				
Associate	5326	0.140	0	1
Bachelor	5326	0.301	0	1
Master	5326	0.121	0	1
Doctoral/Professional	5326	0.044	0	1
Demographic Controls				
Woman	5326	0.503	0	1
Black	5326	0.133	0	1
Hispanic	5326	0.137	0	1

Table 4. **Summary Statistics - (HSLS)**

	N	Mean	Min	Max
Explanatory Variables				
Parental Income	16788	$51756	$14050	$234160
GPA	16788	2.80	0.20	4.00
Parent Degree				
< GED	16788	0.061	0	1
Associate	16788	1.456	0	1
Bachelor	16788	0.260	0	1
Master	16788	0.134	0	1
Doctoral/Professional	16788	0.062	0	1
Own Degree				
Associate	16788	0.060	0	1
Bachelor	16788	0.347	0	1
Master	-	-	-	-
Doctoral/Professional	-	-	-	-
Demographic Controls				
Woman	16788	0.490	0	1
Black	16788	0.146	0	1
Hispanic	16788	0.201	0	1

Table 5. **Estimated Coefficients of Control Variables**

	Full Sample	Low-SES	High-SES
Parent Degree			
< GED	0.376	0.364	0.014
	(0.347)	*(0.356)*	*(0.350)*
Associate	0.225	0.220	0.260
	(0.214)	*(0.292)*	*(0.262)*
Bachelor	-0.123	-0.195	-0.025
	(0.181)	*(0.228)*	*(0.253)*
Master	-0.276	-0.375	-0.176
	(0.162)	*(0.277)*	*(0.193)*
Doctoral/Professional	-0.550	-0.244	-0.493**
	(0.180)	*(0.350)*	*(0.210)*
Own Degree			
Associate	-0.101	0.123	-1.31***
	(0.205)	*(0.254)*	*(0.361)*
Bachelor	-0.148	-0.032	-0.148
	(0.139)	*(0.211)*	*(0.169)*
Master	0.317	0.479	0.250
	(0.233)	*(0.340)*	*(0.313)*
Doctoral/Professional	-0.140	0.092	-0.376
	(0.229)	*(0.104)*	*(0.287)*
Demographic Controls			
Woman	0.115	-0.285*	0.014
	(0.121)	*(0.168)*	*(0.104)*
Black	-0.123	-0.195	-0.025
	(0.181)	*(0.228)*	*(0.253)*
Hispanic	-0.181	-0.139	-0.186
	(0.167)	*(0.224)*	*(0.224)*
N	5326	3036	2290