It Takes a Village: The Impact of Social Groups on Future Outcomes

Matthew Song

1 Introduction

Experiences during childhood and adolescent years play a crucial role in shaping an individual's behavior and trajectory in later years. Positive experiences are linked to better mental health outcomes as an adult (Chopik and Edelstein, 2018), while negative experiences are linked to increased rates of unemployment and poverty (Zielinski, 2009). Research has shown that youth who experience Adverse Childhood Experiences (ACE) are more likely to drop out from school and live in poverty as an adult (Metzler et al., 2017). Furthermore, youth that experience higher rates of ACE are more likely to report negative health outcomes in later life (Felitti et al., 1998), which not only puts a burden on the healthcare system but systemically exacerbates those aforementioned socioeconomic factors. In addition, poverty and low socioeconomic status are intergenerational, as children who grew up in impoverished households find it difficult to escape poverty as an adult (Wagmiller and Adelman, 2009). All of these examples within the literature demonstrate the importance of understanding the impacts of childhood experiences on future socioeconomic outcomes.

While growing up, children tend to have a few consistent social groups within which they form social bonds and experiences. One partition of groups is as follows: (a) Family (b) Friends/Peers/School (c) Self. The first two are intuitive, while the third is an uncommon concept. Everyone, including children, perceive and assess themselves through self-esteem or goal-setting habits, and those self-assessments may be influential in developing the necessary skills to succeed as an adult.

Thus, this paper seeks to investigate how parental dynamics, influences from friends/peers/school, and self-assessments are associated with a child's future outcomes. Future outcomes are measured by whether an individual graduates from college by age 24, as well as their income-to-poverty ratio at age 24. To account for the binary nature of the former, I employ a Logistic regression setup, while the continuous nature of the latter results in a standard OLS setup. Furthermore, I extend the field of literature by providing a discussion of predictive modeling through the lens of policy analysis. In particular, I consider whether regression-based models can accurately predict at-risk youth. Results from this paper can lead to future work in creating policies that identify and employ early interventions.

2 Data and Methodology

Data on youth characteristics and behavior are obtained from the National Longitudinal Survey of Youth 1997 Cohort (NLSY97), which is a representative longitudinal study of American youth age 12-18 at time of survey in 1997. There were 8,984 individuals (4,599 males and 4,385 females) in the initial round, and over 77% of individuals are still in the sample today. In additional to standard demographic questions, the survey also asks questions covering many facets of life, such as questions relating to employment, education, health, etc. In particular, initial rounds of the survey contained questions that assessed family dynamics, friends/peers/school behavior, and self-reflection assessments. Survey responses to those questions were recorded using the Likert scale (Strongly Agree/Agree/etc.).

In the regression model, I control for a standard set of demographic variables and other fixed effects. I aggregate responses from the Likert scale questions into metrics corresponding to the three categories of interest. Finally, my two main response variables are: whether obtained a Bachelor's degree by age 24 (boolean *college24*), and the (log of the) ratio of household income to poverty level at age

24 (continuous *logPovertyRatio24*). Logarithms are used due to the skewed nature of income. I also consider an indicator of whether an individual is in poverty or not at age 24 as a robustness check (boolean *inPoverty24*). Some variables require interactions, as they only apply to subsets of individuals in the data. For instance, a variable measuring motherly support is only applicable to those individuals who live with their mother. All together, the regression specification has the following form:

$$\begin{aligned} y_i = \alpha_i &+ \beta_{i1}\text{Sex}_i + \beta_{i2}\text{BirthYear}_i + \beta_{i3}\text{maxParentEducation}_i + ... \\ &+ \gamma_{i1}\text{livesWithMom}_i + \gamma_{i2}(\text{livesWithMom} \times \text{momSupport})_i + ... \\ &+ \gamma_{ij}\text{pctPeersPositive}_i + \gamma_{i(j+1)}\text{pctPeersNegative}_i + ... \\ &+ \gamma_{ik}\text{liesCheats}_i + \gamma_{i(k+1)}\text{negativeEmotions}_i + ... + \epsilon_i \end{aligned}$$

The first row contains controls, the second row contains variables related to family dynamics, the third row contains variables related to peers/school, and the fourth row contains variables related to self-assessments. A full list of variables and basic descriptions are provided in the Appendix. y_i is the response variable, which is either *college24*, *logPovertyRatio24*, or *inPoverty24*. I run a logistic regression when using *college24* and *inPoverty24* as my response variables since they are binary, and a standard OLS when using *logPovertyRatio24* as it is continuous.

3 Results

Full table results are in the Appendix; a selection of interesting results are discussed here. I do not put much stock into interpreting exact coefficients, especially since logistic regression utilizes odds, and so probabilities will change depending on the fixed values for each variable. Rather, from a policy standpoint, I am mostly interested in which variables are significant, and what their direction (sign) is.

For Regression 1 in Table 1 (Bachelor's Degree by Age 24), many variables are significant. Because Regression 1 is a logistic regression, coefficients are interpreted as the impact on the log-odds of receiving a Bachelor's degree by age 24. Most control variables are consistent with the literature and are highly significant (e.g. females more likely to get a college degree; Black/Hispanic less likely compared to the baseline (non-Black/Hispanic); higher parental education more likely). This is a good start, as it serves as a check that the data does conform to expectations known in the literature.

Interaction terms should be interpreted as average treatment effects. Consider *livesWithMom* which has a coefficient of 0.333, while *livesWithMom:momSupport* has a coefficient of 0.196. The baseline is someone who does not live with their Mom (all values 0). Then, on average, someone who lives with their Mom who is unsupportive would have a 0.333 modifier, but someone who lives with their Mom who is supportive would have a $0.333 + 0.196 = 0.529$ modifier. Notably, it seems like Mom and Dad support is highly indicative of future college outcomes, while whether Mom/Dad are strict is non-significant.

Next, all of the peer coefficients and most school coefficients have the "correct" sign and are highly significant. For instance, an individual with more peers engaging in positive behavior is more likely to graduate from college, agreeing with the intuition that a rising tide lifts all boats. Finally, in terms of self-assessment, only *positiveOutlook* is significant, which can lead us to think about a "manifesting" behavioral effect, in which students who believe the future is bright are more likely to achieve that bright future. Thus, one takeaway could be the emphasis of establishing mental health resources for kids to ensure that they remain optimistic and determined about their future.

To further interpret these results quantitatively, consider this baseline person: White Male, born in 1980, lives with both parents who graduated from high school but are unsupportive, and the remaining variables are 0. Then the probability that this individual obtains a Bachelor's by age 24 would be:

$$\frac{e^{-6.059+12\cdot 0.214+0.333-0.352}}{1+e^{-6.059+12\cdot 0.214+0.333-0.352}} \approx 2.99\%$$

However, all else equal, if both parents were actually supportive, then the probability is:

$$\frac{e^{-6.059+12\cdot 0.214+(0.333+0.196)+(-0.352+0.236)}}{1+e^{-6.059+12\cdot 0.214+(0.333+0.196)+(-0.352+0.236)}} \approx 4.40\%$$

This may not seem like a drastic change nominally, but by simply having supportive parents, the probability of graduating college for this individual has increased by about 50% (net +1.41%), which is a large boost! Other variable coefficients are interpreted in a similar manner, with the direct interpretation being that e.g. engaging in violence at school leads to a 0.463 reduction in the log-odds ratio (of graduating with a Bachelor's degree), or alternatively, $e^{-0.463} = 63\%$ lower odds-ratio.

Turning towards Regression 1 in Table 2 (Ratio of Household Income to Poverty at Age 24), fewer coefficients are marked as significant, but most coefficients still have the expected sign. As this table is from a standard OLS regression, coefficients are interpreted in the usual log-log/linear manner (e.g. *livesWithMom* implies an 0.117 increase in the log poverty ratio at age 24, or a $e^{0.117} = 1.124 = 12.4\%$ increase in the ratio). It is interesting to note that in both regressions, *schoolAbsent* is highly significant. So, the fact that school absence is highly significant in both regressions could be an avenue for future research, and it makes sense since students who skip school are not investing in their future.

4 Conclusion

Overall, the general regression results are fairly consistent regardless of which specification is used. However, the significance level of some variables change. This might happen because although college degrees and poverty are correlated, they are not completely indicative of one another. It is possible (and fairly common) to not have a Bachelor's degree yet remain above the poverty line. For instance, completing trade school (electrician, plumber, etc.) would not count as having a Bachelor's degree, yet trade school provides opportunities for well-paying jobs and a secure financial future. Furthermore, the poverty threshold itself is outdated (discussed in the Economics literature), and so the difference in significance between college degrees and poverty ratios may be due in part to using a bad classifier. Adjustments to the poverty threshold will lead to better disambiguation of the people who are actually living in poverty, and then we may start to see more significant results for that regression.

To directly answer the original research question, it seems that there are important conclusions to draw from each considered category. Having parental support is strongly correlated with better future outcomes. However, perhaps counter-intuitively, parental strictness is not significant, which could be because teenagers will rebel regardless of strictness. Furthermore, all peer effects (percentage of peers engaging in positive/negative behavior) are extremely significant, which could lead to support in building initiatives that surround children with positive influences and avoid allowing children to engage in bad behavior like underage drinking/smoking. In particular, this result could lead to policymakers providing additional funding for mentoring programs such as Big Brothers Big Sisters of America. Finally, the data supports the idea that having a more positive outlook on the future sets children up with a winning mindset to turn that goal into reality.

In addition to previously mentioned caveats, there are other limitations to this study. For one, the data is primarily from the early 2000s. Though some time-invariant trends may still hold, other effects may be lessened with the advancement of technology. Furthermore, some variables were aggregated from a Likert Scale, which implies that a 1-unit increase in the total score would have a constant effect. This is not necessarily true, nor is it a robust way to measure responses. It will be necessary to develop more refined methods of dealing with those responses in the future.

Still, this research at its core has not only confirmed expected results from the literature, but has also added new avenues to consider regarding the impacts of childhood surroundings on future outcomes. Policymakers and researchers would be well-advised to consider these topics moving forward.

5 Extensions

I provide some additional robustness checks, interpretation, and discussion. I also introduce and discuss the predictive portion of the project.

5.1 Robustness Checks

First, in both Table 1 and Table 2, I consider exchanging out the *logAvgPovertyRatio18* (log of the average poverty ratio prior to age 18, used as a metric of how well-off an individual's family was during their childhood) with the binary *inPoverty18*, seen in Column 2 of each table. In each case, the significance does not change (still highly significant) and the coefficient is in the correct direction: positive when using the ratio, and negative when using the binary indicator.

I also consider using the binary variable *inPoverty24* instead of *logPovertyRatio24* as my response variable in Table 2, with results shown in Columns 3 and 4. As discussed in the prior section, when using the poverty ratio as a response, many variables are no longer significant (compared to when using college graduation as the response), and my hypothesis from domain knowledge is that the poverty ratio is messy, and thus the income-ratio may not be as informative as one would expect. The usage of *inPoverty24* is mostly as a robustness check, considering whether effects hold under a simpler metric of whether someone is in-poverty or not, compared to using their income ratio. Comfortingly, most of the significant results still hold, and are also still in the correct direction.

5.2 Additional Interpretation and Discussion

One thing to note is that *sex* becomes significantly positive when using *inPoverty24*. This seems to contradict the results found in Table 1, which shows that *sex* is highly significant and positive towards getting a college degree. According to these two results, females are more likely to graduate with a Bachelor's degree by age 24, but are also more likely to be in poverty by age 24. However, intuition and literature say that obtaining a college degree is strongly negatively correlated with being in poverty, so there is a discrepancy. One possible explanation (outside of prior discussion regarding the poverty threshold) might be pregnancies. For instance, it might be the case that early to mid-20s females are more likely to graduate with a degree, but if females take maternity leave and do not work for some time then they may be classified as being in-poverty, albeit temporarily, due to a lack of income. Further investigation into the data would have to be done in order to determine whether this hypothesis could explain this phenomenon.

Overall, I re-emphasize caution against trying to interpret the exact values of the coefficients, especially those that correspond to variables on the Likert Scale. For instance, *positiveOutlook* is composed of questions which ask for self-reflection ratings on: expecting good things in uncertain times; rarely expecting good things to happen to them (reverse-coded); optimistic about the future; rarely expecting things to go their way (reverse-coded). So, a one-point increase in any of these questions would have the same effect, but perhaps more weight should be put on e.g., optimism about the future, in which case the coefficient values may change and the current interpretation no longer holds.

So, I chose to focus on the direction and significance of the variables, and urge a similar viewpoint. After all, from a social policy standpoint, it may not really matter whether the coefficient of *pctPeersPositive* is 0.77 or 0.87 or 0.67, but the fact that it is strongly positive and highly significant is important in providing arguments as to why policymakers should focus on instituting policies to help encourage that behavior, such as providing more funding for extracurricular activities or bringing college admission experts to schools to talk about the importance of applying for college.

5.3 Prediction

Now, I consider the predictive element of the project. The main research objective here is to compare the differences in performance between OLS, Ridge, and LASSO models on predicting college outcomes (*college24*). I split my data into a train and test set and run 10-fold cross validation on my training data with the following logistic regression specification (analogous to the prior regressions) to tune hyperparameters for Ridge and LASSO:

$$\begin{aligned}\text{college24}_i = \alpha_i &+ \beta_{i1}\text{Sex}_i + \beta_{i2}\text{BirthYear}_i + \beta_{i3}\text{maxParentEducation}_i + \beta_{i4}\text{inPoverty18} + ...\\ &+ \gamma_{i1}\text{livesWithMom}_i + \gamma_{i2}(\text{livesWithMom} \times \text{momSupport})_i + ...\\ &+ \gamma_{ij}\text{pctPeersPositive}_i + \gamma_{i(j+1)}\text{pctPeersNegative}_i + ...\\ &+ \gamma_{ik}\text{liesCheats}_i + \gamma_{i(k+1)}\text{negativeEmotions}_i + ... + \epsilon_i\end{aligned}$$

Note that I use *inPoverty18* to measure family income prior to age 18. I also ran these models using the log-ratio with similar results. After running cross-validation to obtain my best model, I run predictions using the test data and compare certain accuracy metrics with the true test results.

In my test data, the "No-Information Rate" was 79.58%, meaning that if we naively predicted the majority class (not having a Bachelor's degree) for each observation, we would get 79.58% of the predictions correct. That is our baseline model, and we hope that our models will end up beating that value by a significant amount. If so, that means our regression-based models perform better than the baseline and can potentially be utilized in future work in preventative intervention.

5.3.1 Prediction Results

OLS:

```
## Confusion Matrix and Statistics
##
##                    Reference
## Prediction          noBachelorDegree hasBachelorDegree
##   noBachelorDegree               877               147
##   hasBachelorDegree               39                88
##
##                Accuracy : 0.8384
##                  95% CI : (0.8158, 0.8592)
##     No Information Rate : 0.7958
##     P-Value [Acc > NIR] : 0.0001365
##
##                   Kappa : 0.4003
##
##  Mcnemar's Test P-Value : 4.308e-15
##
##             Sensitivity : 0.37447
##             Specificity : 0.95742
##          Pos Pred Value : 0.69291
##          Neg Pred Value : 0.85645
##               Precision : 0.69291
##                  Recall : 0.37447
##                      F1 : 0.48619
##              Prevalence : 0.20417
##          Detection Rate : 0.07646
##    Detection Prevalence : 0.11034
##       Balanced Accuracy : 0.66595
##
##        'Positive' Class : hasBachelorDegree
##
```

From our OLS results, we see that the accuracy of our model reaches 83.84%, and the p-value of the model accuracy being above the No-Info Rate is very low (<0.001). So, we can be confident that the OLS model passes the baseline check and does indeed achieve a significantly higher performance in predicting youth outcomes than naive methods.

From a policy standpoint however, we are likely more interested in how well the model assigns negative values (not graduating from college) conditional on the individual actually being negative. This corresponds to the Specificity rate, which is 95.742% in this model.

To elaborate, we can consider the Type I/Type II error framework. Consider that a hypothetical policymaker instituted policies to provide aid for students considered "at-risk" of not graduating with a degree in the future. If an individual actually would end up with a Bachelor's degree, and the model predicts that they will not obtain a Bachelor's degree, the worst case scenario is that policymakers may be devoting extra funding to someone who does not necessarily require it. But in this case, the funding is not wasted, as it is likely that the additional support granted to the individual may still help them. Perhaps the student does better in school with the funding, so that they can now attend a better college.

On the contrary, the opposite scenario is devastating. If someone who is at-risk of not graduating is erroneously overlooked for aid, then their future outcomes are severely hampered. That is, it is more costly from a public policy standpoint in failing to provide help to someone who needs it (false negative; Type II error) than it is to provide help to someone who does not need it (false positive; Type I error). So, yet again we see that our model is indeed useful, as it provides better overall accuracy than baseline in addition to having high specificity.

As a further extension, in part based on the above discussion of Type I/Type II errors, I consider tuning the threshold for classifying predictions. By default, after obtaining the model and its predictions on the test data, the canonical method is to set a threshold of 0.5 and assign predicted values ≥ 0.5 to 1, and predicted values < 0.5 to 0. However, we can run a grid search to determine the optimal threshold for prediction purposes. Results are shown below:

```
## Confusion Matrix and Statistics
##
##                    Reference
## Prediction          noBachelorDegree hasBachelorDegree
##   noBachelorDegree               864               130
##   hasBachelorDegree               52               105
##
##                Accuracy : 0.8419
##                  95% CI : (0.8195, 0.8625)
##     No Information Rate : 0.7958
##     P-Value [Acc > NIR] : 3.898e-05
##
##                   Kappa : 0.4449
##
##  Mcnemar's Test P-Value : 1.146e-08
##
##             Sensitivity : 0.44681
##             Specificity : 0.94323
##          Pos Pred Value : 0.66879
##          Neg Pred Value : 0.86922
##               Precision : 0.66879
##                  Recall : 0.44681
##                      F1 : 0.53571
##              Prevalence : 0.20417
##          Detection Rate : 0.09123
##    Detection Prevalence : 0.13640
##       Balanced Accuracy : 0.69502
##
##        'Positive' Class : hasBachelorDegree
```

```
##
```

These results are interesting, because we see that adjusting the threshold has both helped and harmed our model (no free lunch). In particular, the accuracy on the test data has risen slightly, from 83.84% to 84.19%. However, our Specificity has dropped from 95.74% to 94.32%. Looking at the 2×2 confusion matrix, we notice that this updated-threshold model has achieved better accuracy in classifying people who do have Bachelor's degrees, but as a consequence, has gotten slightly worse at classifying people without Bachelor's degrees. Comparing these two models, and given the above discussion, I would argue that the first model with slightly lower accuracy is better, since it has better Specificity. This is a trade-off I would be willing to make.

5.3.2 Ridge

I run the same processes as above, utilizing Ridge Regression for prediction. I omit the "threshold tuning" discussion, as generally the results are quite similar to the default threshold.

```
## Confusion Matrix and Statistics
##
##                    Reference
## Prediction          noBachelorDegree hasBachelorDegree
##   noBachelorDegree               883               157
##   hasBachelorDegree               33                78
##
##                Accuracy : 0.8349
##                  95% CI : (0.8122, 0.8559)
##     No Information Rate : 0.7958
##     P-Value [Acc > NIR] : 0.0004335
##
##                   Kappa : 0.3681
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.33191
##             Specificity : 0.96397
##          Pos Pred Value : 0.70270
##          Neg Pred Value : 0.84904
##               Precision : 0.70270
##                  Recall : 0.33191
##                      F1 : 0.45087
##              Prevalence : 0.20417
##          Detection Rate : 0.06777
##    Detection Prevalence : 0.09644
##       Balanced Accuracy : 0.64794
##
##        'Positive' Class : hasBachelorDegree
##
```

We see that our two main metrics, Accuracy and Specificity, are 83.49% and 96.40%, respectively. Compared to OLS (with values of 83.84% and 95.74%, respectively), we see that Ridge Regression has a slightly lower overall Accuracy but slightly higher Specificity. However, these comparisons are all within 1% of each other, so it is not overwhelmingly clear that one model performs significantly better than the other. Yet again though, I would err on the side of having higher specificity, and so I would prefer to use the Ridge model for prediction purposes.

5.3.3 LASSO

Finally, I consider the predictive performance of the LASSO model. Again, I do not discuss the "adjusted-threshold" results as they are similar to the default threshold results.

```
## Confusion Matrix and Statistics
##
##                    Reference
## Prediction          noBachelorDegree hasBachelorDegree
##   noBachelorDegree               882               157
##   hasBachelorDegree               34                78
##
##                Accuracy : 0.8341
##                  95% CI : (0.8113, 0.8551)
##     No Information Rate : 0.7958
##     P-Value [Acc > NIR] : 0.0005701
##
##                   Kappa : 0.366
##
##  Mcnemar's Test P-Value : < 2.2e-16
##
##             Sensitivity : 0.33191
##             Specificity : 0.96288
##          Pos Pred Value : 0.69643
##          Neg Pred Value : 0.84889
##               Precision : 0.69643
##                  Recall : 0.33191
##                      F1 : 0.44957
##              Prevalence : 0.20417
##          Detection Rate : 0.06777
##    Detection Prevalence : 0.09731
##       Balanced Accuracy : 0.64740
##
##        'Positive' Class : hasBachelorDegree
##
```

We see that our two main metrics, Accuracy and Specificity, are 83.41% and 96.29%, respectively. These values are both lower than the Ridge results, so in a vacuum, we would always prefer the Ridge model. However, as the theme goes, these values are extremely close, and perhaps a different test/train split could lead to different conclusions by similar razor-thin margins.

5.3.4 Prediction Conclusion

Overall, this exercise in prediction shows that all of the models have strong predictive power and perform similarly for this setup. Indeed, OLS/Ridge/Lasso do perform significantly better than the baseline naive model. The accuracy increase may not seem significant, as the OLS model with an 83.84% accuracy rate compared to the baseline of 79.58% is only a raw 4.26% increase in accuracy. Still, an additional 1 in 25 youth can be properly identified, which affects many youth when aggregated across the entire population. From another point of view, the OLS model explains $\frac{4.26}{1-79.58} = 20.86\%$ of the remaining error, which is a non-trivial improvement.

In any case, it is clear that from a policy perspective, building and utilizing predictive models may be a promising avenue to identify at-risk youth. Coupled with knowledge of what factors highly influence future success from the first part of this paper, policymakers can turn towards combining these two elements and construct programs that help out the next generation of students.

6 Appendix

6.1 Variables

The following list contains a brief summary of each variable.

- **college24:** Whether individual obtained Bachelor's degree by age 24.
- **inPoverty24:** Whether individual in poverty at age 24.
- **logPovertyRatio24:** log of Poverty Ratio at age 24.
- **sex:** 0 = Male, 1 = Female.
- **birthYear:** 0-4, with 0 = 1980.
- **isLessAge14R1:** Whether individual is $\leq$ age 14 in Round 1 (identifies who can answer self-assessment questions).
- **isBlack/isHispanic:** Whether individual is Black or is Hispanic.
- **hasSibling:** Whether individual has a sibling.
- **maxParentEducation:** Maximum number of years of schooling either parent achieved.
- **logAvgPovertyRatio18:** log of the Average Poverty Ratio prior to age 18.
- **inPoverty18:** Whether considered in Poverty at age 18 (robustness check).
- **livesWithMom/livesWithDad:** Boolean of whether lives with Mom/Dad (prior to age 18).
- **safeEnv:** Whether home environment growing up was considered safe.
- **pctPeersPositive:** Percentage of peers engaging in positive behaviors.
- **pctPeersNegative:** Percentage of peers engaging in negative behaviors.
- **schoolAbsent:** Number of unexcused absent days in recent school year.
- **schoolLate:** Number of days late to school in recent school year.
- **schoolEnvironment:** Score that measures school environment; higher scores are better.
- **schoolViolence:** Whether there is violent behavior at school.
- **teacherQuality:** Score that measures teacher quality; higher scores are better.
- **momSupport/dadSupport:** Whether Mom/Dad supports the individual; only applicable if individual lives with Mom/Dad (hence the interaction term).
- **momStrict/dadStrict:** Whether the individual considers their Mom/Dad to be strict; only applicable if individual lives with Mom/Dad (hence the interaction term).
- **liesCheats:** Score for whether individual tends to lie/cheat; higher scores are more frequent lie/cheat behavior.
- **negativeEmotions:** Score for whether individual experiences negative emotions; higher scores are more frequent negative emotions.
- **positiveOutlook:** Score for whether individual has positive (future) outlook; higher scores are more frequent positive outlook.

6.2 Regression Tables

Table 1: Bachelor's Degree by Age 24

	college24 *logistic* (1)	college24 *logistic* (2)
Constant	-6.059^{***} (0.450)	-6.136^{***} (0.451)
sex	0.696^{***} (0.080)	0.666^{***} (0.079)
birthYear	-0.137^{**} (0.056)	-0.139^{**} (0.055)
isLessAge14R1	-0.318 (0.273)	-0.366 (0.271)
isBlack	-0.236^{**} (0.106)	-0.337^{***} (0.104)
isHispanic	-0.410^{***} (0.120)	-0.521^{***} (0.118)
hasSibling	-0.015 (0.098)	-0.169^{*} (0.095)
maxParentEducation	0.214^{***} (0.016)	0.258^{***} (0.015)
logAvgPovertyRatio18	0.601^{***} (0.057)	
inPoverty18		-0.863^{***} (0.146)
livesWithMom	0.333 (0.332)	0.440 (0.330)
livesWithDad	-0.352^{*} (0.194)	-0.204 (0.191)
safeEnv	0.188^{**} (0.085)	0.196^{**} (0.084)
pctPeersPositive	0.770^{***} (0.259)	0.920^{***} (0.255)
pctPeersNegative	-0.876^{***} (0.245)	-0.854^{***} (0.242)
schoolAbsent	-0.085^{***} (0.012)	-0.088^{***} (0.012)
schoolLate	-0.022^{**} (0.010)	-0.016^{*} (0.010)
schoolEnvironment	0.031 (0.021)	0.033 (0.021)
schoolViolence	-0.463^{***} (0.085)	-0.453^{***} (0.084)
teacherQuality	0.104^{**} (0.044)	0.104^{**} (0.043)

livesWithMom:momSupport	0.196**	0.188**
	(0.096)	(0.095)
livesWithMom:momStrict	0.137	0.116
	(0.084)	(0.083)
livesWithDad:dadSupport	0.236***	0.248***
	(0.089)	(0.088)
livesWithDad:dadStrict	0.146	0.152*
	(0.093)	(0.092)
isLessAge14R1:liesCheats	−0.049	−0.032
	(0.089)	(0.088)
isLessAge14R1:negativeEmotions	0.040	0.020
	(0.084)	(0.084)
isLessAge14R1:positiveOutlook	0.056**	0.062**
	(0.026)	(0.026)
Observations	5,753	5,753
Log Likelihood	-2,187.116	-2,235.599
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 2: Ratio of Household Income to Poverty at Age 24

	logPovertyRatio24		inPoverty24	
	OLS		*logistic*	
	(1)	(2)	(3)	(4)
Constant	0.192	0.362^{**}	-1.145^{***}	-0.847^{***}
	(0.153)	(0.155)	(0.291)	(0.287)
sex	0.034	0.039	0.127^{*}	0.138^{**}
	(0.035)	(0.035)	(0.070)	(0.070)
birthYear	0.004	-0.001	-0.003	-0.011
	(0.025)	(0.025)	(0.049)	(0.049)
isLessAge14R1	-0.155	-0.138	0.138	0.186
	(0.121)	(0.121)	(0.238)	(0.237)
isBlack	-0.408^{***}	-0.408^{***}	0.710^{***}	0.728^{***}
	(0.045)	(0.045)	(0.084)	(0.083)
isHispanic	0.006	-0.014	-0.035	-0.049
	(0.048)	(0.048)	(0.098)	(0.099)
hasSibling	-0.082^{*}	-0.101^{**}	0.160^{*}	0.157^{*}
	(0.046)	(0.046)	(0.095)	(0.095)
maxParentEducation	0.023^{***}	0.026^{***}	-0.027^{***}	-0.027^{***}
	(0.005)	(0.005)	(0.010)	(0.010)
logAvgPovertyRatio18	0.192^{***}			-0.276^{***}
	(0.017)			(0.029)
inPoverty18		-0.535^{***}	0.837^{***}	
		(0.046)	(0.080)	
livesWithMom	0.117	0.121	-0.271	-0.263
	(0.112)	(0.112)	(0.206)	(0.205)
livesWithDad	0.124	0.132^{*}	-0.243	-0.236
	(0.079)	(0.079)	(0.154)	(0.153)
safeEnv	0.012	0.013	-0.024	-0.025
	(0.037)	(0.037)	(0.072)	(0.072)
pctPeersPositive	0.095	0.146	-0.080	-0.019
	(0.111)	(0.111)	(0.214)	(0.214)
pctPeersNegative	0.033	0.023	-0.312	-0.319
	(0.102)	(0.102)	(0.198)	(0.197)
schoolAbsent	-0.007^{***}	-0.007^{***}	0.012^{***}	0.012^{***}
	(0.002)	(0.002)	(0.004)	(0.004)
schoolLate	-0.001	0.0004	0.005	0.006
	(0.003)	(0.003)	(0.005)	(0.005)
schoolEnvironment	-0.016^{*}	-0.015^{*}	0.026	0.028
	(0.009)	(0.009)	(0.018)	(0.018)
schoolViolence	-0.087^{**}	-0.087^{**}	0.198^{***}	0.203^{***}
	(0.037)	(0.037)	(0.072)	(0.072)
teacherQuality	0.025	0.025	-0.032	-0.033
	(0.018)	(0.018)	(0.035)	(0.035)
livesWithMom:momSupport	0.017	0.018	-0.046	-0.042

	(0.039)	(0.039)	(0.074)	(0.074)
livesWithMom:momStrict	0.040	0.035	−0.063	−0.072
	(0.038)	(0.038)	(0.075)	(0.075)
livesWithDad:dadSupport	−0.003	0.001	−0.053	−0.050
	(0.039)	(0.039)	(0.077)	(0.077)
livesWithDad:dadStrict	0.030	0.026	0.070	0.055
	(0.044)	(0.044)	(0.093)	(0.092)
isLessAge14R1:liesCheats	−0.001	0.001	0.022	0.020
	(0.039)	(0.039)	(0.076)	(0.076)
isLessAge14R1:negativeEmotions	−0.057	−0.056	0.033	0.036
	(0.036)	(0.036)	(0.069)	(0.069)
isLessAge14R1:positiveOutlook	0.014	0.013	−0.014	−0.018
	(0.012)	(0.012)	(0.024)	(0.024)
Observations	5,753	5,753	5,753	5,753
R^2	0.098	0.098		
Adjusted R^2	0.094	0.094		
Note:		*p<0.1; **p<0.05; ***p<0.01		

6.3 Sources

Works Cited:

- Chopik and Edelstein (2018), *"Retrospective Memories of Parental Care and Health from Mid to Late Life,"* Health Psychology, Vol. 38, No. 1, pp. 84-93.
- Felitti et al. (1998), *"Relationship of Childhood Abuse and Household Dysfunction to Many of the Leading Causes of Death in Adults,"* American Journal of Preventive Medicine, Vol 14, Issue 4, pp. 245-258
- Metzler et al. (2017), *"Adverse Childhood Experiences and Life Opportunities: Shifting the Narrative,"* Children and Youth Services Review, Vol. 72, pp. 141-149.
- Wagmiller and Adelman (2009), *"Childhood and Intergenerational Poverty: The Long-Term Consequences of Growing Up Poor,"* National Center for Children in Poverty, Columbia University.
- Zielinski (2009), *"Child Maltreatment and Adult Socioeconomic Well-Being,"* Child Abuse and Neglect, Vol. 33, Issue 10, pp. 666-678.

Data Sources:

- National Longitudinal Survey of Youth, 1997 Cohort