'Generation Debt' Turns 40: Modeling Millennial Debt Trajectories in a Multilevel Framework

Ethan Weiland

2024-04-28

Introduction

There is widespread public concern regarding the finances of young American adults ("Millennials"). Commentators are especially worried about high levels of debt, with one going so far as to label this current crop of young adults "Generation Debt" (Kamenetz 2006). This public worry has been mirrored in the sociological literature (Kus 2015; Dwyer 2018; Wherry and Chakrabarti 2022). In light of the the new reality of widespread debt, scholars have studied its associations with fertility (Nau, Dwyer, & Hodson 2015), physical health (Lippert, Houle, & Walsemann 2022), home buying (Houle and Berger 2015), and cohabitation (Addo 2014). Houle (2014) empirically confirmed high levels of debt for Millennials in their mid-twenties using the National Longitudinal Survey of Youth - 1997. Specifically, he found that a significantly higher proportion of Millennials have a negative net worth compared to their predecessors (Late Boomers, Early Boomers), as well as significantly greater debt-to-asset and debt-to-income ratios (measures of financial distress). While insightful, sociological research has only studied Millennials in their twenties and has failed to track Millennials as they have aged into their thirties. This research updates the literature by studying debt trajectories of Millennials throughout their twenties and thirties. Specifically, this analysis will use the National Longitudinal Survey of Youth - 1997 and multilevel techniques to answer the following three questions:

1. How much of the variance in debt is attributable to between-individual and within-individual differences?
2. What is the association between debt and time?
3. What is the association between parental net worth and initial debt levels and debt trajectories?

Data & Methods

Data

The National Longitudinal Survey of Youth - 1997 (NLSY97) is a nationally representative panel study of American youth between the ages of 12 and 16 in 1997, administered by the Bureau of Labor Statistics. At the time of writing, respondents are between the ages of 39 and 43. The survey contains a wide variety of questions touching on practically every aspect of the respondents' lives: health, work, family formation, etc. The NLSY97 was conducted annually in its earlier iterations but is now conducted biannually. The most recent round - Round 20 - was conducted in 2021 and released for researcher use in February 2024. The NLSY97 is made up of two independently selected, stratified multistage area probability samples. This sample design is accounted for using the `svydesign` command in the `survey` package in `R` when making differences about debt at different ages (Table 1).

The dependent variable is total debt. Total debt includes student debt, automobile debt, debt owed to family/friends, credit card debt, hospital debt, and other forms of debt. Importantly, total debt does not include housing debt as housing debt is generally seen as a positive wealth-building form of debt. Respondents are asked about their debts every five years in what are called Youth Assest (YAST) modules. Respondents receive the YAST20 module when they reach the age of 20, the YAST25 module when they reach the age of 25, etc. This total debt variable is made available by the Bureau of Labor Statistics.

The key independent variable for Research Question #2 is time. Time is an index variable equal to the YAST module where total debt was measured. Time ranges from 0 at the first YAST module (YAST20) to 5 at the most recent YAST module (YAST 40). The key independent variable for Research Question #3 is parental net worth. Parental net worth is equivalent to the assets minus the debts of the respondent's parents. Parental net worth can be positive or negative and can be though of as capturing the respondent's social class background. This variable was measured in the first round of the survey and does not change over time.

The following within-individual control variables are included: age, education, income, assets, marital status, and number of children. These variables are measured each time debt is measured and thus vary over time. Age is the respondent's age in years and is included on top of the time variable to separate out period effects from age effects. Education is a continuous measure equal to the respondents' years of schooling. Education ranges from 0 years of schooling to 20 years of schooling. Income refers to the respondent's household income and assets refers to the sum of the respondent's financial and non-financial assets. Marital status is an indicator variable equal to "1" if the respondent reports being married and equal to "0" if the respondent reports anything else (single, divorced, widowed, etc.). The number of children is equal to the sum of the respondent's residential and non-residential biological children.

The following between-individual control variables are included: parental education, parental income, sex, and race. All these variables are measured in Round 1 of the NLSY97 and do not vary over time. Parental education is a continuous measure equal to the years of schooling completed by the respondent's parents, whichever is greatest. Parental income is equal to total household income. Sex is an indicator variable with two categories: "Male" (baseline) and "Female". Race is factor variable with four-levels: "Black", "Hispanic", "Multiracial (Non-Hispanic)", and "Non-Black / Non-Hispanic" (baseline category). The inadequacies of the sex and race measures are discussed in the Discussion & Conclusion.

Only respondents reporting debt at least once were included (N = 7523). Missing values on all variables except for debt were imputed using bagging, implemented in the `caret` packages in `R`. Debt was included in the imputation equation. Measurements with a missing value for debt were dropped from the analysis. The final analytic sample size was 29397 measurement occasions. Many of the financial variables were top coded, oftentimes at 2%, by the Bureau of Labor Statistics to ensure privacy. Additionally, all financial measures (debt, income, assets, parental income, parental net worth) were adjusted for inflation. The appropriate conversion rates to adjust for inflation were obtained from the Bureau of Labor Statistics CPI Inflation Calculator (https://www.bls.gov/data/inflation_calculator.htm). All multilevel models were ran using the `lme4` package in `R`.

Methods

To answer the research questions, multilevel models will be run with individuals as the level-2 units and measurement occasions as the level-1 units. A multilevel model, as opposed to a traditional linear model, is necessary because measurements within individuals are clustered, violating the assumption of independent samples. Multilevel modeling also offers other advantages in the context of longitudinal data like the ability for heterogeneous growth rates through parameterizing time as a random slope as well as the ability to determine the association between initial levels and growth rates.

The first research question is: "How much of the variance in debt is attributable to between-individual and within-individual differences"? To answer this, an intercept-only model will be implemented and then the residual intraclass correlation coefficient will be calculated. The second research question is "What is the association between debt and time?" and the third research question is "What is the association between parental net worth and initial debt levels and debt trajectories?". To answer these questions, the following model will be run:

Level 1:

$$Debt_{ij} = \beta_{0j} + \beta_{1j}Time_{ij} + \beta_2 Time_{ij}^2 + \beta_3 Age_{ij} + \beta_4 Education_{ij} + \beta_5 Income_{ij} + \beta_6 Assets_{ij} + \beta_7 Married_{ij} + \beta_8 NumChildren_{ij} + \beta_9 HasDebt_{ij} + r_{ij}$$

Level 2:

$\beta_{0j} = \gamma_{00} + \gamma_{01} ParentNetWorth_j + \gamma_{02} ParentEducation_j + \gamma_{03} ParentIncome_j + \gamma_{04} Sex_j + \gamma_{05-07} Race_j + u_{0j}$

$\beta_{1j} = \gamma_{10} + \gamma_{11} ParentNetWorth_j$

Where

$r_{ij} \sim \text{N}(0,\sigma^2)$ and $\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim MVN\ (\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_0^2 & \tau_{01} \\ \tau_{10} & \tau_1^2 \end{pmatrix})$

The model includes a random intercept and a random slope for time. On top of the controls discussed above, the following variables are included: $Time_{ij}^2$ and $HasDebt_{ij}$. The former allows for quadratic growth of debt, granting more flexibility. This term is included as a fixed effect due to convergence issues when trying to model it as a random effect. The latter term is an indicator variable equal to "1" if respondent$_j$ has debt at measurement occasion$_i$ and "0" if they have do not. This variable is necessary because respondents that don't have debt at a certain measurement differ from those that do. Not including this variable results in a bimodal distribution of the level-2 residuals. Answering the second research question amounts to testing τ_1^2 and β_2 for significance and answering the third research question amounts to testing γ_{01} and γ_{11} for significance.

Results

Table 1: Total Debt Inferences Across Age, N=(7523)

Age	Mean Debt	Mean Debt SE	Debt SD	Debt SD SE
20	7,045.960	178.210	13,742.370	3,676.970
25	22,180.640	580.040	35,056.440	11,505.220
30	32,916.560	840.200	52,840.610	14,573.350
35	34,727.430	927.140	55,636.320	15,016.360
40	34,776.810	1,249.310	55,647.820	16,264.550

Source: NLSY97. Among respondents reporting debt at least once. Adjusted for survey design and inflation.

To begin, Table 1 lists the inferences of total debt for Millennials at different ages. Since this sample was restricted to respondents reporting debt at least once, these inferences are population estimates of debt at different ages for Millennials who will hold debt at some point. These inferences are made accounting for clustering in the multistage sampling design. Average debt starts off relatively low at age 20, equal to $7045.96. This is unsurprising as the typical American is still enrolled in schooling or just beginning their labor market career. Many have not made a major purchase nor even have a credit card. Average debt then increases until age 30, where it stays relatively constant. This provides evidence for quadratic growth in total debt over time. The population estimate for standard deviation in debt also follows this pattern. The most recent estimate for total (non-housing) debt for Millennials is $34,776.81.

Table 2: Intercept Only Model for Total Debt

	Dependent variable:
	Total Debt (Logged)
Constant	6.66
	(0.03)
Level-Two Variation	
τ_0 Intercept SD	1.9
Level-One Variation	
σ Residual SD	4.16
Number of Groups (Individuals)	7523
Number of Observations (Measurements)	29397
Deviance	171642.67
Note:	Source: NLSY97

Table 2 provides the results of the intercept only model for total debt. As the name implies, this model contains a random intercept and no predictors. This model is useful because it decomposes the total variance before additional predictors (i.e., random slopes) make interpreting this decomposition near impossible. The residual intraclass correlation coefficient is 0.17. In other words, 17% of the variation in total debt is attributable to between-individual differences. While this is not the majority of variation in debt, this is still enough variation to make sociological explanations of between-individual differences meaningful. But, a full explanation of total debt will need to consider where an individual is at in their own life.

Table 3 lists the means, standard deviations (where appropriate), minimums, and maximums of the variables in the model described in the Methods section. Table 4 displays the results. In the model debt is logged, so the coefficients are interpreted as percent changes in total debt rather than additive changes. The coefficient on time (γ_{10}) is equal to 0.31288 ($p < 0.001$). Conditioning on other variables, the average growth rate of total debt across each measurement occasion (five years) is 31%. The coefficient on time^2 (β_2) is -0.0623 ($p < 0.001$). On average, the growth rate of total debt is *decelerating*. This aligns this the results from the basic inferences across different ages in Table 1.

The model was re-calculated without the random slope for time and compared to the model with the random slope for time. A Likelihood Ratio Test ($p < 0.001$), AIC, and BIC all prefer the model with the random slope for time. This indicates that there is significant variation in the the growth rate of debt. The parameter for the random slope τ_1 is equal to 0.19. Quantifying this heterogeneity: For 95% of individuals, the growth rate of debt varies by $\tau_1 * 4 = 76\%$.

The correlation between initial debt levels and the growth rate of debt is negligible: $\tau_{01} = 0.00375$. This is a surprising result, but one that makes more sense when considering that

Table 3: Descriptive Statistics, N=(7523)

Variable	Mean	StdDev	Min.	Max.
Debt 20 ($)	6781.54	14002.61	0	348000
Debt 25 ($)	20790.9	33268.37	0	562400
Debt 30 ($)	30661.7	50337.31	0	525400
Debt 35 ($)	32762.7	52892.59	0	488400
Debt 40 ($)	34523.43	55955.56	0	436600
Parent Net Worth ($1000s)	178.56	244.72	-1814.39	1164
Parent Income ($1000s)	88.58	72.86	0	478.16
Parent Education	13.15	2.98	1	20
Sex (1 = Female)	0.51		0	1
Non-Black / Non-Hispanic	0.54		0	1
Black	0.24		0	1
Hispanic	0.22		0	1
Multiracial (Non-Hispanic)	0.01		0	1
Age 20	20.12	0.63	19	24
Age 25	25.12	0.59	19	29
Age 30	30.52	0.73	30	34
Age 35	35.58	0.79	35	39
Age 40	40.47	0.5	40	41
Education 20	12.34	1.53	0	17
Education 25	13.39	2.52	1	20
Education 30	13.82	2.89	0	20
Education 35	14.07	3.01	0	20
Education 40	14.11	3.01	0	20
Income ($1000s) 20	93.48	97.63	0	703.68
Income ($1000s) 25	88.51	81.06	0	557.23
Income ($1000s) 30	94.33	77.9	0	596.36
Income ($1000s) 35	113.8	101.18	0	707.68
Income ($1000s) 40	129.44	128.26	0	707.68
Assets ($1000s) 20	31.16	60.68	0	1451.12
Assets ($1000s) 25	51.69	82.68	0	1109.6
Assets ($1000s) 30	77.32	115.12	0	1260
Assets ($1000s) 35	121.82	171.01	0	1188
Assets ($1000s) 40	176.68	214.08	0	1062
Married 20	0.08		0	1
Married 25	0.27		0	1
Married 30	0.43		0	1
Married 35	0.48		0	1
Married 40	0.51		0	1
# Children 20	0.24	0.56	0	5
# Children 25	0.69	1	0	8
# Children 30	1.24	1.29	0	9
# Children 35	1.61	1.39	0	12
# Children 40	1.38	1.23	0	7
Has Debt 20	0.52		0	1
Has Debt 25	0.78		0	1
Has Debt 30	0.76		0	1
Has Debt 35	0.74		0	1
Has Debt 40	0.69		0	1

Source: NLSY97. Among respondents reporting debt at least once. Adjusted for inflation.

Table 4: Longitudinal Random Slope Model for Total Debt (Logged)

	Dependent variable:
	Total Debt (Logged)
γ_{10} Time	0.31^{***}
	(0.05)
β_2 Time^2	-0.06^{***}
	(0.005)
γ_{01} Parent Net Worth ($1000s)	-0.0002^{**}
	(0.0001)
γ_{02} Parent Education	-0.003
	(0.004)
γ_{03} Parent Income ($1000s)	0.0003
	(0.0002)
γ_{04} Sex (Female=1)	0.02
	(0.02)
γ_{05} Black	-0.08^{***}
	(0.02)
γ_{06} Hispanic	-0.09^{***}
	(0.02)
γ_{07} Multiracial (Non-Hispanic)	0.06
	(0.09)
β_3 Age	0.01
	(0.01)
β_4 Education	0.13^{***}
	(0.004)
β_5 Income ($1000s)	0.0004^{***}
	(0.0001)
β_6 Assets ($1000s)	0.001^{***}
	(0.0001)
β_7 Married	0.31^{***}
	(0.02)
β_8 Number Children	-0.03^{***}
	(0.01)
β_9 Has Debt	9.06^{***}
	(0.02)
γ_{11} Time:Parent Net Worth	-0.0001^{***}
	(0.0000)
Constant	-2.14^{***}
	(0.20)
Level-Two Variation	
τ_0 Intercept SD	0.4
τ_1 Slope SD	0.19
τ_{01} Intercept Slope Correlation	0
Level-One Variation	
σ Residual SD	1.06
Number of Groups (Individuals)	7523
Number of Observations (Measurements)	29397
Deviance	92600.4
Note:	*p<0.05; **p<0.01; ***p<0.001
	Source: NLSY97

the initial measurement was taken at age 20. As discussed above, individuals are too young at this age to accumulate that much debt. The model was recalculated with the YAST20 data removed, making the initial measurement at age 25. The correlation in this model is -0.52. This moderate negative correlation means that as debt increases at age 25, the rate of debt growth decreases. Potential explanations of this is that the individuals with high debt in their twenties aren't the same as individuals with high debt in their thirties. For example, a respondent may have a lot of student debt in their twenties to attend medical school, but in their thirties they are graduated and working a high-paying occupation financed by their debt and don't need to take out further debt. Whatever the mechanism, this finding provides evidence against the idea of a "debt trap" where individuals with high debt experience a spiral of increasing debt.

The final research question revolves around the association between parental net worth and initial debt levels and debt trajectories. The fixed effect of parental net worth (γ_{01}) is -1.6×10^{-4} ($p < 0.01$). An increase of \$1,000 in parental net worth decreases initial debt levels by 0.02%. Or, an increase of \$100,000 in parental net worth decreases initial debt levels by 1.63%. The interaction between time and parental net worth (γ_{11}) is -7.8×10^{-5} ($p < 0.001$). An increase in parental net worth significantly decreases the growth rate of debt. While statistically significant, the magnitude of these associations are rather small. This leads to the conclusion that parental net worth does not *meaningfully* impact the amount of initial debt nor the growth rate of debt.

Discussion and Conclusion

This study used the National Longitudinal Survey of Youth - 1997 and multilevel modeling techniques to analyze the dynamics of debt for Millenials adults as they age through their twenties and thirties. There are five key takeaways. First, most of the variation in debt comes from within individuals rather than between individuals. Second, on average, debt increases over time but this increase diminishes. Third, individuals experience varied trajectories of debt growth. Fourth, there is no association between debt levels at age 20 and the rate of debt growth and there is a negative association between debt levels at age 25 and the rate of debt growth. Finally, parental net worth does not meaningfully impact debt.

Like all research, this study faced several limitations. First, the analytic sample was restricted to respondents with debt at at least one measurement occasion and respondents with no debt at a certain measurement differ from those that do. This problem was handled by incorporating an indicator variable, but an even stronger technique would be to use a mixture model (e.g., a zero-inflated model). This model would first model the probability that a respondent has debt at a certain time, and then examine the associations between debt and any variables of interest among respondents that are predicted to have debt. Since most mixture models are used for count data, researchers should work on developing a multilevel zero-inflated model for continuous outcomes. Second, the model only included two levels but the data actually consists of four levels with the additional two reflecting the sampling design. Model convergence issues

prevented these additional levels from being estimated. The result is that the parameter estimates may be less efficient than reported in Table 4. However, due to the large sample size (29397), this is unlikely to alter substantive findings.

Third, the gender and race measures made available by the NLSY97 are inadequate. The gender measure only contains two categories: "Male" and "Female" and is only measured in the initial round. Since this aligns more with sociological understandings of sex, I refer to this variable as "sex" not "gender" throughout this paper. The race measure contains four categories: "Black", "Hispanic", "Multiracial (Non-Hispanic)", and "Non-Black / Non-Hispanic". An obvious problem is that this groups White and Asian respondents into a single category. Since race and gender are major predictors of social outcomes, the Bureau of Labor Statistics should do a better job of measuring these characteristics. Finally, this analysis only focuses on a specific subset of Millenials: those between the ages of 12-16 in 1997. Thus, it is unclear whether these results hold for all members of the Millennial generation. Policymakers should be especially careful in extrapolating these findings to even younger generations.

Navigating debt - whether it be taking out student loans to pay for higher education or using a credit card to pay for essential goods during a period of unemployment - is a crucial skill needed for contemporary social mobility. This analysis adds to the literature by studying debt longitudinally, revealing the dynamic features of debt. Debt is - and will continue to be - an important aspect of life for contemporary young adults in the United States.

GitHub Link

https://github.com/ethanphilipweiland/millennial-debt

References

1. Addo, Fenaba R. 2014. "Debt, Cohabitation, and Marriage in Young Adulthood." Demography 51(5):1677–1701. doi: 10.1007/s13524-014-0333-6.
2. Dwyer, Rachel E. 2018. "Credit, Debt, and Inequality." Annual Review of Sociology 44:237–61. doi: 10.1146/annurev- soc- 060116- 053420.
3. Houle, Jason N. 2014. "A Generation Indebted: Young Adult Debt across Three Cohorts." Social Problems 61(3):448–65. doi: 10.1525/sp.2014.12110.
4. Houle, Jason N., and Lawrence Berger. 2015. "Is Student Loan Debt Discouraging Homeownership among Young Adults?" Social Service Review 89(4):589–621. doi: 10.1086/684587.
5. Kamenetz, Anya. 2006. Generation Debt: How Our Future Was Sold Out for Student Loans, Bad Jobs, No Benefits, and Tax Cuts for Rich Geezers–And How to Fight Back. New York, New York: Riverhead Books.
6. Kus, Basak. 2015. "Sociology of Debt: States, Credit Markets, and Indebted Citizens." Sociology Compass 9(3):212–23. doi: 10.1111/soc4.12247.

7. Lippert, Adam M., Jason N. Houle, and Katrina M. Walsemann. 2022. “Student Debt and Cardiovascular Disease Risk Among U.S. Adults in Early Mid-Life.” American Journal of Preventive Medicine 63(2):151–59. doi: 10.1016/j.amepre.2022.02.002.
8. Nau, Michael, Rachel E. Dwyer, and Randy Hodson. 2015. “Can’t Afford a Baby? Debt and Young Americans.” Research in Social Stratification and Mobility 42:114–22. doi: 10.1016/j.rssm.2015.05.003.
9. Wherry, Frederick F., and Parijat Chakrabarti. 2022. “Accounting for Credit.” Annual Review of Sociology 48:131–47. doi: 10.1146/annurev-soc-030320-114444.

Appendix

Figure A1. Distribution of U_0

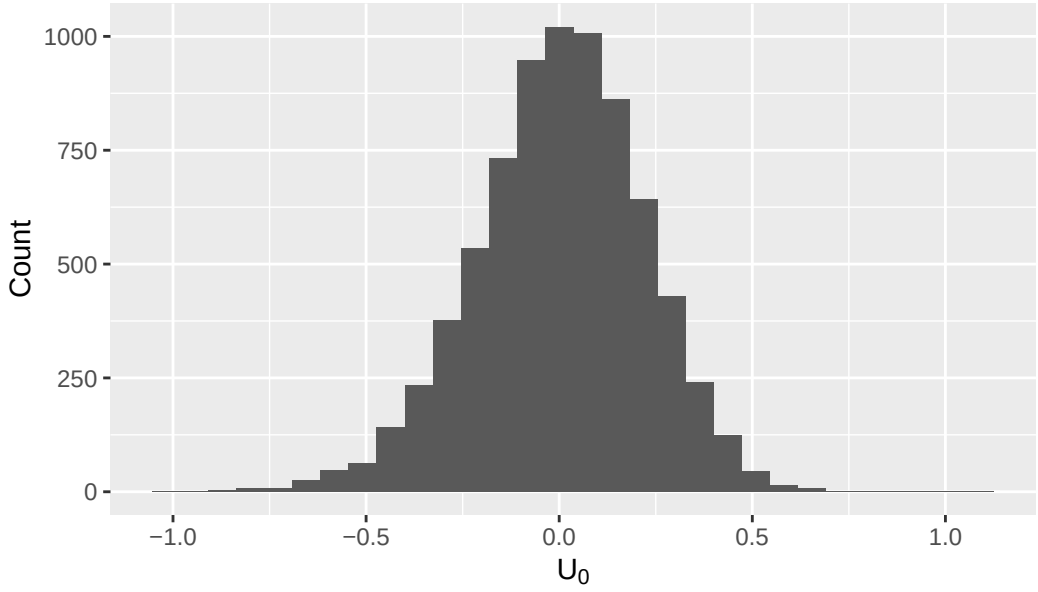

Figure A2. Distribution of U_1