Using Qualitative Data to Inform Structural Modeling: An Application to Post-Secondary Education Investments†

Stefanie DeLuca[1,2,3], Nicholas W. Papageorge[2,4], Andrew Gray[5], Seth Gershenson[6], Joseph Boselovic[1,2], and Jasmine Sausedo[1,2]

[1] *Dept. of Sociology, Johns Hopkins University*
[2] *The Poverty and Inequality Research Lab, Johns Hopkins University*
[3] *School of Government and Policy, Johns Hopkins University*
[4] *Dept. of Economics, Johns Hopkins University, IZA and NBER*
[5] *Olin Business School, Washington University in St. Louis*
[6] *School of Public Affairs, American University and IZA*

September 15, 2025

Abstract: Policy conclusions from structural models hinge on model specification. Yet, multiple models could rationalize the same empirical patterns. To aid in model specification, we introduce a new, interdisciplinary mixed-methods approach that relies on qualitative data from systematically analyzed open-ended interviews. The idea is straightforward: asking decision-makers how they make decisions can help the researcher to design better structural models of decision-making—or rule out classes of models that are less useful. We suggest three steps: (i) collecting qualitative data and analyzing it using rigorous qualitative methods to generate insights for modeling decision-making; (ii) corroborating these insights using large-N quantitative datasets like the National Longitudinal Survey of Youth 1997 (NLSY97); and (iii) using these insights to guide specification of a tractable structural econometric model. We apply our approach to post-secondary education choices, where investment levels are puzzlingly low among resource-constrained students. Findings suggest that students facing poverty and instability anticipate a low probability of completing longer educational programs and thus rationally choose shorter, more achievable alternatives. Estimated utility parameters, policy conclusions, and welfare calculations hinge on what we assume individuals believe about the likelihood of completing school. Our results underscore the value of using qualitative data and analytical methods to gather insights from decision-makers to build better models of decision-making.

Keywords: Mixed Methods Research, Qualitative Data, Structural Modeling, Post-Secondary Education

JEL Classification: I2, Z1.

---

†First draft: September 1, 2020. Current draft: September 15, 2025. The authors note that a related but substantially different paper was circulated as a working paper under the title, "When Anything Can Happen": Anticipated Adversity and Postsecondary Decision-Making." The authors also wish to thank Kathryn Edin and Susan Clampet-Lundquist, Co-PIs with DeLuca for the MTO Q10 Transition to Adulthood Study, which provided the qualitative data used in this paper. We are also grateful for: support from the Russell Sage and William T. Grant Foundations; comments from Peter Arcidiacono, Lawrence Katz, Michael Keane, Robert Moffitt, Philip Oreopoulos, and Kevin Thom; and research support from Paige Ackman, Olivia Cigarroa, Jamie Chan, Courtney Colwell, Olivia Morse, Kiara Nerenberg, Lauren Ricci, Oscar Volpe and Allison Young. Email [corresponding author]: `papageorge@jhu.edu`.

1 Introduction

A fundamental task in economic research is to construct structural models of behavior for use in counterfactual policy analysis. Policy conclusions drawn using such models hinge on their appropriate specification. Historically, there are many ways to build such models, including reliance on economic theory, descriptive data analyses, and contextual knowledge. Model misspecification is thus of particular concern when behavior seems misaligned to standard theory, data are scarce, and the researcher's understanding of the context is limited. One example of such a situation is the question of post-secondary educational investments among low-income youth, which seem puzzlingly low given that human capital accumulation is typically viewed as the best way to escape poverty. Analyses of qualitative data—for instance, answers to open-ended interview questions—offer a path toward building more useful structural models. The idea is straightforward: when constructing an appropriate model of decision-making, asking decision-makers how they made their choices could offer valuable insights, allowing researchers to consider or rule out various classes of models and home in on key channels that drive behavior. However, while the use of qualitative data is not unknown in economics (Bewley, 1995), it is rarely applied.

This paper presents a method for using qualitative data analysis to inform structural modeling. The method consists of three steps. First, we collect and analyze qualitative data from open-ended interviews in which individuals discuss their decision-making, which we use to develop novel hypotheses about decision-making. Second, we use large-N data sets to corroborate whether ideas that emerged (or hypotheses rejected) in step one are borne out in larger and nationally-representative samples. Third, and based on what we learned in the first step and verified in the second, we build a structural model of dynamic decision-making that we use to analyze policy and welfare. Research by economists has used qualitative data to aid the design of quantitative data collection (Kling et al., 2005) and, more recently, to understand the mechanisms underlying experimental effects (Bergman et al., 2024; DeLuca et al., 2023). However, we do not know of other research that systematically analyzes qualitative data with the explicit purpose of building a structural econometric model of behavior to be parameterized, estimated, and used in counterfactual policy analysis.

How, precisely, we collect and then transform qualitative data into specific hypotheses and—when appropriate—additions to structural models is described in detail when we introduce our qualitative data so that other researchers can follow or employ our methods. The approach is inductive and includes reading through entire interviews to identify the range of responses related to decision-making about an outcome—in this case, what to do after high school—and observe whether there are new or surprising insights that challenge or depart from existing models that might merit modifications to existing models. To be clear, we do not have reason to believe respondents' narratives *must* diverge from existing ones; we merely allow for the possibility that the qualitative data may offer novel information about decisions that guide us to improve upon existing models. We also read full interviews to understand the context in which the decision-maker is operating, so that we can understand the constraints and options they face when making choices. As we discuss in more detail below, when reviewing related literature, our approach is complementary to recent scholarship on the use of automated tools, e.g., Large Language Models to generate hypotheses (see, e.g,. Ludwig and Mullainathan (2024)).

We apply our approach to the problem of understanding post-secondary education decisions in the U.S. Even though four-year degrees have high relative returns, the modal outcome in the U.S. is "some college," which comprises shorter degree programs and unfinished coursework.[1] Common explanations of this behavior can be subsumed into a "deficit model," e.g., insufficiencies in academic preparation, information about the returns to a bachelor's degree, or knowledge of financing options. Other narratives suggest impatience, a distaste for schooling, or feelings of despair. In contrast, our analysis leads to a model that allows individuals to make insightful and rational decisions based on information they *do* have. In particular, shorter degree programs may be optimal for students who face a high risk of not completing a four-year degree. This risk could be particularly high for students with unstable home lives or living in neighborhoods that are violent or under-resourced. These students have often experienced adverse shocks and may anticipate future adverse shocks that could derail their education. A rational response may be to choose shorter, albeit less lucrative, educational programs that they anticipate being able to complete. The logic is analogous to that of a rational, risk-averse investor eschewing high-risk, high-return assets in favor of safer assets with lower expected returns, though the deficit narrative is rarely applied to such investment choices. Moreover, recent literature seems to suggest that more certainty about financial aid does induce low-income students to start longer and more lucrative degree programs (Burland et al., 2023; Dynarski et al., 2021), though finances are far from the only relevant uncertainty facing such students.

Our hypothesized link between anticipated shocks and educational investments emerged from the first step of our approach: analysis of rich qualitative data. Data are from systematically sampled interviews with youth in high-poverty areas in Baltimore. Nearly all respondents reported having experienced adverse events (which we henceforth refer to as

[1]Research has increasingly shown this educational outcome has some benefits (Darolia et al., 2023; Liu et al., 2015; Jepsen et al., 2014), but is still not as valuable as a four-year degree (Webber, 2016; Hillman, 2014; Oreopoulos and Petronijevic, 2013), and those with sub-baccalaureate credentials have more difficulty paying back student loans (Brown et al., 2019).

"shocks" or "adverse shocks"), such as housing instability, incarceration of a family member, or violence. Many also anticipated future shocks that could interrupt their educational path. Based on some of their accounts, it is clear that this anticipation caused them to make *safer* investments in their human capital. For example, Rhiannon, 22, articulated this concern by stating: "I had never really been outside of Baltimore, and I was just *afraid that something would happen while I was away* [at school]." Consider the implications of Rhiannon's thought process. First, she recognized that her studies could be interrupted by events at home, independent of her own behavior, even while away at school. Second, this recognition influenced Rhiannon's decision-making: she attended post-secondary school, but closer to home than she otherwise would have. Third, this dimension of her decision-making had little to nothing to do with her academic plans, the quality of the college she was considering, or her beliefs about schooling in general. This quote, like many others in the interview dataset, points to non-school factors and events that have not even happened yet that impinge on the decision-making process. The connection between these other contexts and educational decision-making would likely not have been made so easily using survey data. At the same time, Rhiannon's statement could be construed in different ways—as reflecting an information deficit, despair, etc. This highlights the importance of conducting the type of systematic and careful qualitative analysis we outline in the following section to generate insights that can be operationalized (rather than over-analyzing or cherry-picking a few quotations). It also motivates our consideration of large-N data to corroborate the validity and generalizability of what we learn from the qualitative data.

Indeed, the second step of our approach draws on large-N quantitative data from publicly available, nationally representative longitudinal surveys to corroborate a link between adverse household shocks and educational decision-making. This step is in part a validation exercise to assess whether what we learn from the qualitative data improves our understanding of outcomes in large-N data sets relative to existing narratives, which in turn would provide further rationale for developing new models. In our case, we focus on the National Longitudinal Survey of Youth 1997 (NLSY97).[2] Students across demographic groups report perceived non-completion risk. However, the perceived risk is higher among respondents who have experienced adverse shocks in the past. As we show, lower expectations with regard to educational outcomes are associated with increased fears about future adverse events, as well as with enrollment in two-year educational programs (rather than four-year bachelor's degree programs).[3] Unfortunately, the variables available in the NLSY do not allow us to

[2] The Educational Longitudinal Study (ELS) and the High School Longitudinal Study (HSLS) can also be used to investigate some of the issues we examine related to education decisions. We analyze these data sets to corroborate our main findings using the NLSY97 where possible and provide evidence in appendices.

[3] We focus on two-year versus four-year degree programs in our main analyses even though there are other

analyze the anticipation of shocks as fully or as deeply as is possible in the smaller qualitative data set; thus, our proposed approach could also inform future large-scale quantitative data collection and survey design.

The third step of our approach is to formalize our qualitative conclusions about the effects of anticipated shocks on educational pathways by developing a dynamic structural model of postsecondary decision-making. The model envisions students maximizing their lifetime utility, which is affected by the opportunity cost and consumption value of schooling as well as future income. A longer degree program entails higher upfront financial sacrifices but is more lucrative. Anticipated shocks are incorporated as a positive probability that students who enroll in an educational program do not progress in it, delaying or precluding degree completion and thereby lowering returns to educational investments. The model thus captures the idea from our narrative interview analysis that students may anticipate the possibility of adverse shocks that derail long-run education plans, and in turn may opt for shorter degree programs they believe they can finish (or for no degree at all) even if doing so reduces lifetime earnings. This model of educational decision-making serves several purposes. First, it formalizes the way an *anticipation* of potential future shocks—not just a realized shock—affects students' choices. The model can thus shed light on how instability and poverty can reverberate through optimal choices and expectations even if anticipated future shocks are never realized. Second, the model allows us to explore the consequences of misspecification, i.e., how a failure to incorporate anticipated shocks (e.g., relying only on a deficit model of decision-making) can generate incorrect conclusions about the utility cost of schooling and thus lead to misleading conclusions about policy and welfare.

Results from our structural model show that the disutility of schooling (and thus the welfare value of student subsidization and retention policies) varies widely depending on students' perceived probabilities of shock-related derailment. Among high-achieving, underrepresented minority students from low-income families and adverse-shock-prone childhoods, the estimated disutility of a year in a bachelor's degree program is \$2,081 (here and elsewhere in 2013 dollars) when we assume they rationally account for noncompletion chances, which we calculate indirectly using the NLSY97 data, but rises to \$11,897 when we omit noncompletion rates from the model. Moreover, we simulate that a 10-percentage-point reduction in the probability of educational noncompletion yields the lifetime welfare value equivalent of an annual subsidy of \$2,412 to all students in four-year programs, with greater values accruing to more disadvantaged students who exhibit higher rates of noncompletion.

This study contributes to several strands of existing research on factors affecting educational decisions and educational attainment gaps. Increasingly, studies in economics examine

options, including certificate or training programs that are often shorter.

returns to sub-baccalaureate pathways (often subsumed into the category "some college"). Findings typically suggest that benefits to postsecondary coursework outweigh costs (Liu et al., 2015; Darolia et al., 2023; Jepsen et al., 2014; Lovenheim and Smith, 2023). Still, along multiple dimensions and outcomes, individuals with "some college" but no degree better resemble people with no college versus bachelor's degree holders (Hillman, 2014), suggesting a need to further our understanding of why students opt for these pathways.

Explanations of educational choices often emphasize information, family background, and resource constraints and generally focus on four-year degrees (Dynarski et al., 2021; Hoxby and Turner, 2015; Attewell et al., 2011; Keane and Wolpin, 2001). Additional research from across the social sciences considers other factors contributing to the *social context* within which students make decisions, including: family background, neighborhood quality (Wodtke et al., 2011; Sharkey, 2010), school-based inputs like guidance counselors (Bettinger and Evans, 2019; Castleman and Goodman, 2018), inadequate housing, exposure to violence, food insecurity, or teacher supportiveness (Jack, 2019; Chyn, 2018; Hart, 2018; Cox, 2016; Goldrick-Rab, 2016; Roderick et al., 2011; Jones, 2010).[4] These factors may create insurmountable educational barriers to disadvantaged students.[5] However, this research does not acknowledge a role for noncompletion risk, which is an emergent theme from our qualitative data and the focus of our analysis. While novel, this idea is related to work suggesting that for-profit institutions can attract students from low-income backgrounds who may perceive a higher likelihood of success at such schools versus at traditional colleges, *precisely because* they offer shorter programs (Iloh, 2018; Holland and DeLuca, 2016; Iloh and Tierney, 2014).

Our work also links to research on the role of beliefs and expectations in decision-making processes regarding education and work (Belzil and Leonardi, 2013; Wiswall and Zafar, 2015; van der Klaauw, 2012; Bozick et al., 2010). Raley et al. (2012) show an association in the NLSY97 between higher subjective probabilities of young pregnancy and lower postsecondary enrollment and persistence rates. Indeed, going back to the 1970s, researchers have noted that parents' and students' own educational expectations are a strong predictor of final attainment, even conditional on many background characteristics (Jacob and Wilder, 2010; Alexander et al., 1994). Moreover, Papageorge et al. (2020) show that high school teachers' expectations affect youths' attainment. However, the exact mechanisms through which these effects operate remain unknown. We examine how the anticipation of adverse shocks that derail education affects educational investments.

[4] Also relevant are the less easily measured factors that comprise *social capital*, including social ties and networks as well as social norms surrounding education (Chetty et al., 2022).

[5] A field experiment carried out at a large community college campus in Texas found that access to individual case managers who helped students navigate "life barriers" significantly increased persistence and degree completion among female students, while financial assistance alone had no effect (Evans et al., 2020).

Finally, a wealth of prior work explores mixed-methods research in the social sciences (Pearce and Hardie, 2024; Hesse-Biber and Johnson, 2015). Economists and others have periodically called for complementing analyses of large data sets with qualitative methods (Grigoropoulou and Small, 2022; Akerlof, 2020; Moffitt, 2000; Bewley, 1995). There is growth in acceptance of this approach (DeLuca et al., 2024; Bergman et al., 2024), but economics lags behind other fields (Thelwall and Nevill, 2021). Moreover, how best to use qualitative methods in economics remains an open question. Our paper provides a methodology that integrates qualitative data and analysis with econometric analysis and economic modeling. The approach that most closely resembles ours is the use of qualitative data to generate a theory or formulate a question that is then combined with population-representative quantitative data (e.g., Myers and Oetzel (2003)).

This point relates broadly to the recent work suggesting machine learning as a potential solution to find patterns in data that could aid in economic research (Ludwig and Mullainathan, 2024). The hypothesis-generation problem that work identifies is similar to the one we do, as is the solution: using different, novel tools to improve the kinds of questions we ask, the classes of models we can reject or develop, and the outcomes we could consider in the social sciences. An algorithm can provide the econometrician with unique insights from otherwise opaque data sets, but so too can careful analysis of themes emerging from open-ended interviews by social scientists. Experienced social scientists are—as of today—relevantly more expert on such topics than, e.g., LLMs, given their full access to the literature (which has not been granted to existing generative AI) and the fact that AI can often miss subtleties in human expression, meaning, and interpretation (e.g., Rhue (2018)). We propose that our approach is particularly warranted given our goal to model behavior because structural models are not simply econometric, but representative of qualitative processes on the part of decision-makers and qualitative reasoning on the part of researchers. Structural modeling is, therefore, best grounded in rigorous analysis by the methods of researchers trained in qualitative approaches and might be considered an ideal locus for interdisciplinary collaboration. It is a conceptual leaping-off point toward the answers to difficult questions via methodological synthesis across the social sciences.

This is a manifestation of the insight of Moffitt (2000) that theoretical economic rationales are necessarily qualitative. If a structural econometric model is an attempt to quantify a qualitative narrative, then that narrative itself can (and in some cases ought to) emerge from qualitative data and analysis. Moreover, the descriptions decision-makers provide of their own thought processes, and insights gleaned therefrom by qualitative social science methods, can provide a uniquely rich informational foundation on which to build an economic *model*—

not just a research question.[6]

Sections 2 and 3 present our data analysis and findings using small-N qualitative data and large-N quantitative data, respectively. Section 4 presents a dynamic structural model of educational decision-making based on this data analysis, model estimates, and counterfactual policy simulations. Section 5 concludes.

2 Qualitative Interview Data Analysis

2.1 The Collection of Qualitative Interview Data

The relationship between anticipation of negative shocks and postsecondary decision-making is an inductive finding that emerges from the stories of youth who participated in in-depth, semi-structured interviews as part of an evaluation of the impacts of the Moving to Opportunity (MTO) housing mobility experiment in Baltimore, Maryland. We use data from a mixed-methods study of families and children in the Baltimore site of MTO. A total of 636 families in Baltimore participated in this program, all of whom were Black.[7] In 2010, a qualitative interview component was added to the study to examine the transition to adulthood for the MTO participants and for disadvantaged youth more broadly. A stratified random sample of 200 youth (ages 15 to 24) was chosen from the Baltimore MTO sample, and 75 percent of these youth agreed to participate in the qualitative portion of the study (N=150).[8]

We draw on qualitative interview data from these 150 low-income youth and young adults, all of whom initially lived in Baltimore's highest-poverty neighborhoods, concentrated mostly in high- and mid-rise public housing developments. Although the interviews focused on youths' experiences changing neighborhoods as a result of MTO, they covered a wide range of topics concerning respondents' transition into adulthood.[9] These semi-structured, in-depth interviews covered open-ended questions about employment, education, neighborhoods, friends and family, risky behavior, and mental health. Youth were asked about their college and career preparation, postsecondary decision-making process, and, for

[6] It is also true that in-depth narrative accounts of decision-making from systematically sampled groups like this are not at present available online for generative AI.

[7] A potential limitation of these data is that the sample consists only of Black youth. The addition of a sample of disadvantaged non-Black students would help us understand whether some of the patterns in these data generalize to other racial or ethnic groups. Indeed, DeLuca and Burland (2023) and DeLuca et al. (2025) have asked samples of mostly white young adults, who varied in their incomes, about their postsecondary decisions and exposure to shocks and have found similar patterns to those we find here, albeit not as extreme, given the extent of the racial and economic disadvantage of this Baltimore sample. Our quantitative analyses reported in Section 3, using nationally representative data (which include respondents of different races) show that main patterns generalize to other populations.

[8] Kathryn Edin and Susan Clampet-Lundquist were Co-PIs with DeLuca for the MTO Q10 Transition to Adulthood Study.

[9] See DeLuca et al. (2016) and Boyd and DeLuca (2017) for more information on sample and interview design.

youth who were interviewed after high school, their experiences in postsecondary institutions and at work. Most interviews were conducted in respondents' homes, and 96% of respondents still lived in the Baltimore area at the time of the interview. All names used in this paper have been changed to pseudonyms chosen by the respondents themselves.

The interview approach used to collect these data builds on long-standing methods in urban sociology used to observe social life and the ways in which individuals make meaning of their everyday routines (Edin and Lein, 1997; Anderson, 1990; Becker et al., 1961). Specifically, the data were collected using *narrative interviewing*, a semi-structured approach that employs open-ended questions. This approach allows for a wide range of responses to emerge along with targeted follow-up probes, which help to ensure all interviews cover the same material. Interviews conducted with this approach tend to create natural, in-depth conversations, rather than a clinical series of short questions and answers. Interviewers focus on empathetic, non-judgmental listening in order to signal to study participants that they, not the research team, are the experts on the topics of the study and invite them to tell their own stories within semi-structured question modules. When successful, this interviewing method invites descriptive narratives about social processes, such as educational decision-making, to emerge naturally.

On this project, the research team developed a detailed interview guide designed to gather data on all domains of young adult life, which included open-ended "main" questions (e.g., "tell me how you ended up at the community college?"). Reliance on open-ended questions is preferable to highly structured questions, which run the risk of leading respondents or closing off unanticipated themes (Becker, 1990). Respondents therefore answered at length with detailed and often unsolicited information, in an order that made sense to them. This approach generated insights that informed both pre-existing research aims and topics unanticipated by the researchers, many of which would likely have been missed by standard survey questions with pre-ordained, "forced choice" answer sets.

Once respondents had a chance to reply to open-ended prompts, interviewers followed with more detailed "probes" to support less talkative respondents and ensure systematic topic coverage across respondents, including allowing them to touch upon issues discussed by other respondents, thereby increasing the comparability of data collected in varied interviews. Probes were typically questions about "how" rather than "why" events happened and included specific topics we wanted to ask about—such as financial aid—that might not come up unsolicited from the respondent. Alternating between open-ended questions and detailed probes, interviews took on the form of extended conversations, usually lasting two or more hours. Interviewers largely memorized interviews (aided by a notepad when needed) so that the interviews flowed more like a conversation than a clinical back-and-forth. They also used verbal cues, eye contact, and body language to signal interest without disrupting the flow of respondents' stories. All these techniques allowed interviewers to collect narrative data on decisions and descriptions of events at length and avoid the impression that respondents' answers would be judged.

This interview style is based upon the idea that decision-makers themselves may have necessary, untapped insight about how decisions are made. The question is not whether young people make decisions *for some specific reason or not*, with the reason determined *a priori*. Instead, we want to know how the decision-making process works, broadly speaking. To paint the complete picture, we need to acknowledge that many of the reasons students choose certain educational paths might not readily occur to any one researcher (or indeed any eight researchers). In particular, the experiences of the youth in Baltimore shed light on something that is more difficult to imagine for people who have not faced the same circumstances and have not experienced the same histories, families, institutions, and communities: youths' beliefs about the costs and benefits of the alternatives they perceive as available to them. Our goal is to take lessons from narrative interview data to inform further large-N data analysis as well as structural modeling, both addressed in later sections of the paper.

2.2 Hypothesis Formation and Preliminary Analysis

Once interview data were collected, they were analyzed in search of patterns that suggested cohesive hypotheses about how individuals made choices. Alongside full transcript readings, we systematically coded every interview by manually tagging text for topics in all domains relevant to decision-making based on initial research aims and the topics that emerged in the course of the fieldwork. We then pulled segments related to education, work, and career and identified the presence or absence of data on a range of postsecondary options (college, work, military, trade school). We also identified text related to respondents' rationales for their choices of postsecondary path, as well as specific language about the perceived tradeoffs with each option (e.g., why going to the military was better than college).

We note that this initial step could potentially occur with a computer algorithm, a reasonable approach that some researchers have begun to explore with qualitative data (Geiecke and Jaravel, 2025). A complementary approach, and our chosen one, is for an analyst to read the interviews with care—admittedly a time-consuming endeavor—not in search of certain words or phrases, but to understand how respondents describe the entire decision-making processes. This process can offer unique clues about how appropriate conceptual and econometric models could be built. The aim is to listen to the whole decision-making process, in the context of a life history with neighborhoods, schools, family, and peer influences. These interviews contain information about expected utility, beliefs about payoffs, and perceived constraints in a holistic and relevant context. In particular, we read transcripts not only for explanations consistent with existing literature and narratives, but also for those that upend or expand them—novel factors we can add to a model.

This approach is consistent with widely accepted sociological methods for generating and handling surprising or novel findings. These methods fundamentally involve being steeped in the data on the ground, creating the opportunity for discovery but requiring discipline to align unexpected results with existing theory. In seminal work, Merton (1945) wrote about "chance" discovery of "valid results which were not sought" (p. 469). Later, Merton (1968) elaborated on this idea of "serendipity" as a process of "observing an unanticipated, anomalous and strategic datum which becomes the occasion for developing a new theory or for extending an existing theory" (p.158). More recent work concerning "abductive" analysis by Timmermans and Tavory (2012) defines some qualitative analysis as a "process aimed at producing new hypotheses and theories based on surprising research evidence" (p. 180). Our approach follows and builds on this methodological tradition, bridging from each new hypothesis thus produced to econometric analysis that stays in harmony with it.

Through our analysis, it became apparent that two important factors youth considered when weighing their perceived postsecondary education options were, first, whether they would be likely to complete the proposed programs and, second, whether they believed they would be better off or more economically stable if they made the attempt. These two factors were bound together more tightly than existing narratives tend to acknowledge. In one form or another, nearly all respondents mentioned concerns and uncertainty about whether or not educational investments would be worthwhile. However, the reason was not a lack of future orientation or a large discounting of future well-being, hopelessness in the face of adversity, or confusion about the financial payoffs to college.[10] Rather, respondents, having faced household-level instability in a number of domains (housing, parental employment, the deaths of loved ones), seemed primarily to doubt whether they would be able to complete investments and make it to a future where payoffs would be realized. This suggested that consideration of past household-level shocks was leading to anticipation of future instability, which would make potential educational investments less likely to be completed and therefore less lucrative to start. A novel explanation for educational decisions—shocks and the anticipation of shocks—thus emerged. We began to conceive of adverse household shock anticipation as a way to rationalize how these youth "downgraded" their educational investment plans. It was only at this point that we turned to secondary, nationally representative data to explore whether this was indeed a more general phenomenon.

For the present analysis, interview transcripts were systematically coded for the following:

[10]In Section 2.4, we systematically analyze our qualitative data to asses alternative hypotheses.

negative shocks; anticipation of negative shocks; postsecondary plans (including four-year, two-year, and for-profit school enrollment; employment; military enlistment; and illicit activity); rationales for postsecondary plans; discussion of the costs, benefits, and tradeoffs of postsecondary plans; and beliefs about the future.[11] The research team read full interview transcripts and field notes for all respondents in the sample and focused on coded segments concerning postsecondary decision-making, adverse life events, and any specific language around respondents' anticipation of future negative shocks.

In the next few sections, we leverage the interviews to describe: the prevalence of adverse shocks and how they derail educational pathways; the anticipation of future shocks and their connection to youths' expectations regarding their future outcomes; and how a few youths explicitly connected their anticipation of shocks to their postsecondary decisions, especially about sub-baccalaureate programs. We have more data on how youth discuss postsecondary education and adverse events in general and less on the specific connections between the anticipation of shocks and postsecondary choices and outcomes. However, that evidence is at least suggestive of our story and warrants the further exploration we embark on in the subsequent sections.

2.3 Findings in the Qualitative Data

Adverse Shocks: Experiences and Derailments: To begin, we describe respondents' experiences of adverse shocks, as well as instances in which those shocks led to *actual* disruption of their long-term educational paths.[12] Accounts of unexpected adverse life events were pervasive among the 150 youths who were interviewed. Table 1 reports experiences of significant shocks in childhood and adolescence. Each row of Table 1 is a specific type of shock or adverse event. To understand the table, consider "Arrested/incarcerated (self)" referring to the respondent being arrested or incarcerated. Column 1 shows how many individuals reported experiencing this shock at least once (40 people); column 2 divides this number by 150 to show the proportion of the sample who experienced this shock at least once (26.7%); and column 3 shows the total number of instances reported throughout the interviews (57), allowing for multiple instances per individual.

Looking at Table 1, only one individual out of the sample of 150 reported no adverse shock or event. More than half (54.7%) described having a parent or friend who had been

[11] This process was carried out by research assistants asked simply to look for conversation on certain topics, rather than by the researchers ourselves, so that no knowledge of the particular research questions or hypotheses could bias the process and lead to "finding just what we were looking for."

[12] The link between observed past and expected future shocks is not limited to the sample of 150 respondents discussed here. As we will detail in Section 3, many students in the NLSY97 who anticipated adverse events in their lives but pursued a bachelor's degree anyway ended up unable to achieve their goal, suggesting that anticipation of derailing shocks is a rational expectation for many students.

Table 1: Categories of Adverse Shocks in Youth Interview Sample

Adverse Shock	(1) Indiv.	(2) Pct.	(3) Instances
Neighborhood violence/drug activity/gang presence	91	60.7	107
Arrested/Incarcerated (family/friend)	82	54.7	117
Death of family/friend	79	52.7	112
Substance abuse (family/friend)	55	36.7	66
Police interaction	52	34.7	57
Unplanned pregnancy	49	32.7	49
Absent parent	46	30.7	48
Arrested/incarcerated (self)	40	26.7	47
School violence (respondent not involved)	40	26.7	40
Housing instability (forced/reactive move, eviction)	39	26.0	44
Physical/mental health issue (family/friend)	35	23.3	35
Physical/mental health issue (self)	34	22.7	49
Domestic violence/abuse	31	20.7	37
Experience of violence (self)	16	10.7	16
Witnessing violence	13	8.7	14
Experience of non-violent crime	11	7.3	11
Parental separation/divorce	11	7.3	11
Experience of violence (family/friend)	10	6.7	12
Forced school transfer	10	6.7	10
Experience of bullying (self)	8	5.3	8
Substance abuse (self)	7	4.7	7
Family estrangement	5	3.3	6
School disorder	5	3.3	5
Other moves	5	3.3	5
Lost job (self)	4	2.7	4
Foster care	2	1.3	2
No adverse shocks	1	0.7	–
Observations	150	100.0	919

Table adapted from DeLuca et al. (2024). Each row reports the frequency of different categories of adversity in three ways. Column 1 shows the number of people who report an experience at least once in their interview. Column 2 divides the number in column 1 by 150 to obtain the proportion of the sample that reports experiencing that adversity at least once. Column 3 reports how many times specific instances of this category of adversity are mentioned over the entirety of the 150 interviews; this differs from column 1 because some people report multiple instances. Individuals reporting multiple adverse shocks, both across and within categories, is common with a per-individual rate of 6.13 adversities; over half the sample (52 percent) reported between two and eight such events.

incarcerated or was involved in illegal activity, often drug sales or other illicit employment. About 30 percent of respondents had experienced a period of parental absence during their childhood, and a little over a quarter (26.7%) of respondents had been arrested or were put in jail or on probation. More than half of youths interviewed had experienced the death of someone close to them (52.7%), including parents, siblings, cousins, friends, and other family members.[13] Housing shocks—unplanned moves resulting from events such as evictions, foreclosures, fire, housing voucher inspection failure, and financial insecurity—were also experienced by almost 30 percent of respondents. While only about 10 percent of youths reported being victims of violent crime themselves, over a quarter reported seeing violence at school (26.7%), about 20 percent reported domestic violence in their household (20.7%), and over eight percent (8.7%) witnessed violence elsewhere. Importantly, these percentages likely all undercount the prevalence of these adverse events, as we did not ask about more of them directly.

Analysis of the same data set in DeLuca et al. (2024) shows that, while just one individual reported no adversity or shocks, several report 15 or more distinct adverse events. Over half of the sample, 52 percent, reported between two and eight adverse shocks. In other words, such events were commonplace in the lives of our respondents. As twenty-one-year-old Karen told interviewers, "it's like, when you wanna move one step forward, it's like you getting knocked back five more steps, and it's driving me insane." Given the prevalence of adverse shocks among our respondents, it is no surprise that educational derailments were common as well: roughly one third of those who enrolled in postsecondary education left before finishing, many of whom were enrolled in trade schools. This is likely a significant underestimate of the final dropout rate in this sample due to right-censoring.

Care of family members is another factor affecting educational investments. The need to care for family prompted Chanel, 21, to change her postsecondary pathway. After high school, Chanel enrolled in a nearby university and started working towards a bachelor's degree in psychology. After Chanel had spent two and a half years at the university, her mother suffered an injury requiring knee replacement surgery, prompting her to leave work to recover. Chanel explained: "She hadn't been workin' for a while so that's why, you know, she needed like some extra income." As the only child still living at home, Chanel took on the responsibility of caring for her mother. She left her four-year university and instead enrolled in a for-profit institution to focus her training in the medical field. She explains the appeal of the occupational credential, given her family's situation: "I wanted to start immediately so I could have some money to help out." Having a credential in the medical field after a few months meant that Chanel was able to both help care for her mom and enter the workforce sooner than she would have, had she finished her bachelor's degree.

Thus, despite the time and effort that Chanel had already invested in her bachelor's degree, the sudden onset of caretaking responsibilities and financial need led to a choice to

[13]This figure does not include the deaths of grandparents (primary caregivers for some youths) due to natural causes, which many respondents had experienced as well.

pursue a quicker credential and, in turn, earlier access to jobs. This choice is not difficult to rationalize given the circumstances and constraints Chanel faced but does not reflect an information deficit or initial resource constraint.

Sierra, 20, echoed the sentiment that family instability might affect her postsecondary educational trajectory when she described her initial plans to attend college after graduating from high school: "I mean times get hard. One day times get, you know, they be good, everything paid on time then next time we might have a downfall." She also talked about how her family was currently experiencing a "hard time" that was disrupting her postsecondary plans. As Sierra was graduating from high school, her mother lost her job and her sister became pregnant, so, to support her family, Sierra chose to work in the service industry rather than attend school.

> We had like this little downfall in our family whereas I had to wait a while [to go to school], I had to wait . . . I remember my mother had got laid off and my sister was pregnant so, you know, my mother was like, you know, she gotta pay the bills. She was all looking forward, we were all looking forward so much.

Sierra and her family had been "looking forward" to a future in which she followed through on her original plans to go to college. Sierra still wanted to pursue some sort of college but, as of her interview, remained hesitant about the path forward: "So I mean I waited and I waited but now I feel as though I'm ready. So now I'm looking into it—like should I take a all year-round school or should I just take some classes? You know, I'm looking into it. I wanna make sure."

The onset of familial responsibilities for Sierra, then, was accompanied by a significant change in her college-going plans. Rather than pursuing the bachelor's degree that she and her family had been anticipating, she became hesitant about her path forward in a context in which a family "downfall" was always possible.

Issac's (23) story shows how different kinds of negative shocks can happen in quick succession and compound the difficulties students face in their postsecondary educational trajectories. As Isaac finished high school, he received scholarship opportunities from multiple four-year schools to play basketball, but he wanted to remain close to his family and ultimately chose to attend a community college in South Bend, Indiana (his family left Baltimore after our study began). He managed to stay on this track despite getting arrested after high school and spending six months in jail. However, just as Isaac was preparing to finish his two-year program, his sister was diagnosed with a rare disease and fell into a coma. Isaac explained how this unforeseen development in his family caused him to leave school, as he was just about to get his degree:

> She was in a coma for about two months, and she passed away, which, I would have almost—I would have almost finished my Associate's Degree, but I came back here. So she had my nephew, she passed away, and right now it's just me. I had a scholarship, and that's when my sister was still alive, you know, everything went downhill from there.

In this recollection, Isaac revisited his choice to stay close to family instead of taking advantage of one of the basketball scholarships he was offered with some regret. As of his last interview, he had not returned to school after leaving his two-year program and had instead cycled through jobs, working in warehouses, with temp agencies, and as a hotel security agent in order to provide for his nephew. Despite the planning and significant effort that Isaac put into his postsecondary pathway, the sudden shift in his caretaking responsibilities made it increasingly unlikely that he would return to earn any postsecondary credentials.

In addition to instability within one's immediate family, relationships with romantic partners also came with the possibility of unexpected caretaking responsibilities and shocks. For example, in the case of Vicky, 20, her boyfriend had been living with her and her family for a handful of years when he was involved in a violent altercation that put him in a coma and left him with lasting cognitive impairments. As a result, Vicky became his full-time caregiver, helping him bathe and feed himself. Vicky's mother remained unemployed and struggled with alcoholism, which Vicky said caused her to feel that she was "the responsible one" in the house. The sudden shift to this role of caregiver left Vicky feeling uncertain about the future. As she described, "Some nights I go to sleep wondering like what's goin on, what's gonna happen the next day." She had been attending a nearby community college but, similarly to Chanel, left to enroll in a medical certificate program at a for-profit school. She preferred this to the community college because it was more directly focused on the medical field, in which she hoped to get a job quickly so that she could bring more money into the household. In fact, Vicky considered getting multiple certificates in different occupational areas to hedge her bets.

In at least one case, experiences with death caused adolescents to abandon postsecondary education not because they lost interest but because the impact of loss derailed their daily lives. After graduating from high school, for instance, Tiara, 19, enrolled at a community college but dropped out due to several unanticipated deaths in her family. When we last spoke with her, she had personally arranged a funeral for one of her cousins. She explained, "Just, people kept dying. Every month I done had somebody in my family that died or somebody that was real close to my family. I actually dropped my classes because I kept having a death in family like, even like, if you miss too many days you automatically fail, so instead of failing, I just withdrew from the classes."

Given the pervasiveness of parental incarceration, absence, and death in their lives (see Table 1), a number of youths in our sample experienced a sudden onset of new responsibilities as they prepared to leave high school or shortly thereafter. Such experiences left respondents with the sense that they would continue to be called upon to provide caregiving or financial resources for their family members in the future, limiting their postsecondary options and affecting their decision-making process. We examine this anticipation next.

Expectations: Reasoned Unease: Some of the youth interviewed associated their experiences of adverse shocks with the expectation that similar events would happen again in the future. Taniya, 18, experienced a tumultuous childhood because of her father's alcohol abuse, and she described how these experiences had made her worry about his early death: "If he started drinking again, I see him drinking I'm just gonna have to prepare [. . .] like one day he's just not gonna be here. I guess I just got mentally prepared for that." Taniya was not alone in expressing these feelings: all but one of the youth in our sample spoke of their own accord about anticipating future adverse shocks.

Some respondents reported that experiencing and anticipating these events made them feel out of control and uncertain about their futures. Erica, 21, described experiencing the deaths of several aunts and uncles, her father, and a classmate in a brief span of time. She explained that these losses left her feeling anxious about death, including her own: "I really don't take death well. I have anxiety [. . .] my grieving process is not so much 'Oh, I miss the person,' it's 'Oh, will I die from that,' or death, facing death, yeah." Matthew, 21, felt he could not anticipate when his life might end. He had been arrested multiple times starting at age 14 and lived in a neighborhood where, as he describes it, violence is a part of everyday life. When asked where he thought he would be in five years, Matthew responded, "I ain't going to call it. I can't call it. Not when anything could happen." In response to questions about his goals for the future, he replied, "I'm trying to live right now [. . .] I ain't thinking about the future. I might not even make it to the future, so."[14]

Death and violence, in particular, appeared frequently in respondents' discussions about their neighborhoods and families. They described violence and death as being regular occurrences and yet still unpredictable. Some of their friends or acquaintances died over minor disagreements, or even for seemingly no reason at all. As Sierra noted, "Nowadays you fight, you either getting stabbed, shot up, you losin' your life over a little argument, over a little argument." Bridget, a junior in high school, commented: "You just wake up and hear someone just got killed. Like, DANG can't people go one day without killing?" Christopher,

[14]As we show in the next section using NLSY97 data, this concern about dying in the near future is much more common among young people, especially disadvantaged youths, than might be expected.

21, described the violence in his former public housing neighborhood and the potential for responding to such violence if it affected his family, "Every day you would hear a gunshot, or every day you would see somebody fighting, and you never know, like, when is that going to happen to you, or when, you know, it was gonna be your family member, or when you was gonna have to bear it all just go out there and just fight a war or whatever." Many respondents made comments such as "life is short" and "you can die today or tomorrow." Certainly, such concerns would loom large in youths' decision-making about the future.

These experiences and the choices that rationally follow them make the educational trajectories of so many of our respondents fraught and utterly uncertain. At 21 years old, Tony had experienced a somewhat stable childhood in comparison to other respondents. Of course, this is true only relatively—he had witnessed several incidents of violence throughout his childhood and suffered periods of anxiety and depression. He had two brothers who sold drugs and lived with him sporadically between stints in jail. However, Tony also benefited from a solid support system and an older sister whom he considered a role model. Still, he worried about making the same mistakes as his brothers. Although Tony hoped to pursue a bachelor's degree and become a pharmacist, he was concerned about his ability to finish his current program at community college. As he put it, "I'm just tryin' to, I want everything to stay like it is now. I really don't want no changes until I get my degree."

When asked where he expected to be in five years, he responded, "I still don't think that's enough time, but I'll probably still be here, still be in school, chasing my education, yep." Thus, even Tony—enrolled in a degree program, planning to pursue a bachelor's degree, and feeling that "everything is good"—still struggled to predict how long his educational trajectory would last, or whether unanticipated circumstances might derail it.

Choices: Anticipation of Shocks and Decision-Making: Shock-induced derailment of the kind experienced by Tiara and other respondents illustrates how youths' educational plans were thrown off course by unexpected family needs and the loss of friends and relatives. However, some youths' experiences affected their education not just directly via interruption but also indirectly through shifts in how they thought about their future, including the plausibility of longer-term plans.

For example, Elijah's experiences with instability led him to continuously re-optimize his educational decision-making. Elijah took a year off after graduating from high school and worked with his cousin, an experience that inspired a professional interest in the auto industry. Yet, despite his aspiration to work with cars, his postsecondary strategy was to pursue multiple, distinct educational options so that if one career possibility did not work out, he would have an alternate plan: "Like I thought about givin' like, like getting in the automotive industry and get a couple big, get a couple years in here and then go back to school for the heating and ventilation, so I could have two certificates." Like Vicky, Elijah described his educational decisions as if they were backup plans. He explained: "Just do what you can and if [something bad] happen, it happen, just make sure you know what to do for it not to happen next time." Rather than pursuing a longer degree, Elijah sought multiple short-term programs to obtain credentials focused on specific skills, and thus serve as an insurance policy against any instability that he might encounter in the future. This perceived need to try to accumulate multiple credentials quickly or to develop educational or professional back-up plans may have diverted some youth, like Elijah, off the pathway to a four-year degree.

Few stories, though, could be as stark an example of uncertainty-induced postsecondary choice as that of Rhiannon, 22. When Rhiannon was in 11th grade, her older brother was murdered, the victim of a random shooting while out with friends. As a result, her mother became very protective of Rhiannon and her younger brother. Rhiannon told us, "I wasn't able to be in high school and do much of anything, my mom was always worried that something would happen to me." Nevertheless, as a high school senior at one of the best high schools in Baltimore, she applied to ten colleges and was accepted to all of them, some with substantial scholarship offers. Ultimately, though, she decided she would choose the school closest to home because, as she explained, "I had never really been outside of Baltimore and I was just *afraid that something would happen while I was away*, and [this school] was close enough to home but far enough away to get away from Baltimore" (emphasis added). In this instance, the negative shock of losing her brother did not create a concrete obstacle to college-going for Rhiannon; rather, it shaped her thinking about what kind of school was reasonable, given her competing desires to leave home and to be nearby in case some kind of tragedy befell her family again.

These experiences are emblematic of a larger narrative that emerges from the 150 interviews, which we examine further and model more explicitly in the rest of this paper: disadvantaged youth anticipate that the instability to which their lives are subject may make it impossible or unreasonable to complete a bachelor's degree, leading them to rationally opt for shorter and less lucrative degree programs. The impact of adverse shocks on choices and outcomes is clear in these instances. However, a narrative interview is not necessarily emblematic of a population-level pattern. In Section 3 we thus turn to an analysis of a large-N data set. Yet, we note that it is on the foundation of the qualitative interviews that we are able to propose a novel narrative of postsecondary decision-making among disadvantaged youth.

2.4 Issues and Alternative Explanations

One prevalent critique of the use of qualitative data in the social sciences is that, given diverging opinions about sampling, methods, and analysis, it can be difficult to assess the quality of work (Small and Calarco, 2022; DeLuca, 2023; Jerolmack, 2023). Some have concerns that snippets, quotations, or even respondents can be cherry-picked to support a specific model, mechanism, or explanation of a phenomenon (see, for instance, Morse (2010)). In our study, we have a systematically sampled group of disadvantaged Black young adults and interviewed 75% of them, which reduces concerns about biased response rates and the extent to which that might shape our claims. Moreover, respondents did not differ significantly from non-respondents in the MTO data on any observable characteristics. Thus, we know at least that the sample is representative of very low-income youth from families living in high-crime, high-poverty neighborhoods, and we are less concerned that our sample is selected in a way for which we cannot account. Furthermore, we examined each interview and coded it for relevant characteristics so that our hypotheses emerge from analysis of all the complete interviews rather than a handful of quotes.

As this section has shown, our collection, coding, and analysis of qualitative data is systematic and follows guidelines analogous to best practices when analyzing quantitative data; these include the provision of a sampling frame, avoidance of directly posing research questions to respondents and multiple hypothesis testing, and upfront prescription of coding procedures. But the methodology we develop here goes further in two ways. First, we suggest the *a priori* use of qualitative data to identify novel and plausible stories in a systematic manner in addition to its *a posteriori* use in corroborating findings from large-N data (like that in Section 3). Second, we outline and provide an example (in Section 4) of the formalization of the narrative that emerges from qualitative analysis as an econometric model that can be used in counterfactual policy analysis. With such a model, researchers can examine the relevance and magnitudes of the key mechanisms that qualitative data reveal—they can go beyond the extensive margin of *whether* such factors exist and instead interrogate the intensive margin of *how much* they matter.

Crucially, our approach also suggests that researchers use qualitative data not only to detect novel narratives, but also to examine (and potentially reject) existing alternatives to the degree possible, especially in cases in which data limitations make a large-N, quantitative approach impossible or inappropriate. To be sure, the goal is not to design a model that encompasses every single possible factor or mechanism that comes up in the narrative interviews, nor to reject wholesale all others. Instead, we aim to use qualitative data to assess whether leading explanations put forth in extant research are supported by what respondents say, as well as to uncover potential alternatives. In this sense, qualitative data allows us to interrogate and reject classes of models while supporting and shaping others.

Here, we consider several potential factors in educational decision-making and then discuss either how they are incorporated into the model we construct or how they are at odds with our qualitative data. To name a few, students may choose post-secondary pathways because of financial constraints, a personal dislike of school, information deficits, a contextual "culture of despair," or attitudinal impatience. The structural econometric model we develop and employ in Section 4 allows for a (dis)utility of school, which incorporates both monetary and non-monetary costs—that is, the mechanisms by which financial constraints and a dislike of school would affect educational choice. Indeed, one of this paper's main findings is that if researchers impose beliefs onto agents in a model of decision-making that these agents do not actually have, estimates of these school utility parameters will be biased.

Much extant literature emphasizing the importance of information constraints (Dynarski et al., 2021; Hoxby and Turner, 2015) has implicitly assumed that individuals making educational choices have the (likely biased) belief that if they enroll in postsecondary school, they will certainly finish; this research focuses instead on deficits of information regarding the pecuniary returns to graduation or the likelihood of acceptance or receipt of financial aid. We make the implicit assumption about completion expectations explicit and then relax it, allowing students' beliefs to align with demographic population completion rates. It is also worth noting that in the narrative interview data summarized above, there is little evidence that respondents fail to understand the link between education and higher earnings.

Two remaining alternative hypotheses regarding the causes of suboptimal postsecondary enrollment, feelings of "despair" and individual impatience, are conceptually related. Despair, or fatalism, might manifest as a lack of educational investment because the individuals experiencing it believe the probability that *the education itself* will actually improve their long-term outcome is low (Browman et al., 2019). The mechanisms via which such despair might have an effect on educational choices are somewhat muddy (Moffitt, 2016), and as Kearney and Levine (2016) note the next step in examining them is to find out how they show up "on the ground." We thus explored the possibility that this specific kind of economic despair plays a significant role in our qualitative interview data. Specifically, we examined a random sample of half of our 150 MTO interview transcripts and coded them for discussions of any thoughts or feelings resembling this kind of despair, and found just one case that could feasibly be construed as such. Instead, we see much stronger evidence of resilience, hopefulness, and awareness regarding the value of education (including pecuniary benefits), alongside individuals' uncertainty about their personal futures as captured in our econometric model.[15] We also find little qualitative evidence of impatience (which one might

[15] This is also consistent with long-standing literature in the sociology of education arguing that low-income,

attempt to model using low discount factors or other aberrant discounting rules). Respondents consistently engage in long-run planning and thought about how their lives might pan out, suggesting that they care and think a good deal about the future.

While the narrative interviews did not include specific questions about, for instance, despair with respect to education or economic outcomes as a decision-making factor, they were conducted in the open, wide-ranging way described in Section 2.1 to allow respondents to talk organically about their motivations for and against any particular choice they made. It would be hard to argue that, were it present, we would not find evidence of such despair in conversations held in these Baltimore neighborhoods—among the poorest and most violent in America. Those youths interviewed were unlikely to see optimistic evidence about the return to education when they looked at family members—only 13% had a caregiver who had ever attended any postsecondary program, with even fewer completing bachelor's degrees and finding jobs in their fields that helped them stay out of poverty. The lack of discussion of despair broadly construed, along with the prevalence of discussion of concerns about future adverse events, strongly suggests that the factors we include in our model are among the most important to examine. In sum, our qualitative data tend to point us away from some other explanations for suboptimal educational choices and toward a novel one, which we explore quantitatively in the next section.

3 Nationally Representative Data Analysis

3.1 Introduction to the NLSY97

If we believe that adverse shocks (and histories thereof) influence young people's educational decision-making in the ways suggested in Section 2, we would next want to demonstrate that this relationship is policy-relevant and economically significant. In Section 4, we accomplish this by formulating a structural model of that decision-making process. However, rather than assume that shocks and derailment are important factors in educational decisions beyond the small and relatively homogeneous sample examined in the previous section, in this section, we explore the idea using a nationally representative sample of a non-MTO population. The goal is to provide generalizable evidence for the concern that motivates the structural model—namely, that anticipation of adverse-shock-related derailment influences individuals' decisions about their educational investments.

We examine how adverse shocks relate to expectations and educational decisions using data from the National Longitudinal Survey of Youth 1997 (NLSY97). The NLSY97 sampled 8,984 individuals born between 1980 and 1984 who lived in the U.S. at the initial survey

black youth can hold "abstract" beliefs about the value of education in general but question whether it will payoff for them in a "concrete" way (e.g., Mickelson (1990)).

date in 1997, surveying them at first annually and then (starting in 2011) biennially. In addition to the traditional background variables we use as controls, including race, ethnicity, family income, and mother's education, the NLSY97 survey contains a wealth of questions regarding respondents' childhood experiences of what we have termed adverse shocks, as well as their expectations about their own futures, including educational and adverse-shock-related outcomes. All data regarding childhood shock histories and expectations for the future were gathered at the initial survey in 1997, when respondents were all approximately high-school-aged. We utilize this nationally-representative data through the 2015 survey, and thus observe individuals' responses regarding their educational and employment outcomes through at least age 30.

The results reported here qualitatively replicate in the Education Longitudinal Study [16] (ELS) and High School Longitudinal Study[17] (HSLS) given variable availability. However, the NLSY97 has the advantages of tracking longer records of adult outcomes and, importantly for our purposes, includes questions about students' expectations regarding future adverse shocks, as well as a more detailed question about expectations regarding future bachelor's degree attainment. Moreover, as it is based on a household (versus individual) sampling frame, the data continue to track individuals who have left school. Thus, the NLSY97 provides more precision in corroborating the qualitative story that emerged in Section 2, as well as more flexibility in specifying a model that conforms to that story in Section 4. Accordingly, we focus our analysis on NLSY97 data. To be clear, the results of this analysis are descriptive; while causal stories can be told, our goal in this section is merely to show that the model we posit in Section 4 makes sense in the world described by a nationally representative survey *and* the qualitative data. The results can be found in Appendix A.

Table A1 summarizes key variables for the full NLSY97 analytic sample, as well as by race and family income. There are notable racial and socioeconomic status (SES) disparities in educational attainment. Adverse shocks are fairly common in the NLSY97 analytic sample, which can broadly be classified as either "family shocks" (e.g., changes in household structure

[16] We present some analogous ELS results in Appendix B. The ELS is a nationally representative longitudinal survey of individuals who were in 10th grade and thus likely aged 15 or 16 in 2002, with the first survey wave occurring in that year and subsequent surveys administered in 2004, 2006, and 2012 (the most recent). Respondents were asked in each wave, among other things, about adverse events in their lives. They were asked about their expectations regarding their own educational attainment in the first two waves. However, the educational expectation question was cruder than that in the NLSY97, expectations about other kinds of events are missing, and individuals are only observed through about age 25.

[17] We present additional results using the HSLS results in Appendix C. The HSLS is a nationally representative longitudinal survey of individuals who were in 9th grade in 2009, with the first survey wave occurring in that year and subsequent surveys administered in 2012 and 2016. Student expectations regarding financial aid and academic ability in a postsecondary context are the relevant additional variables we explore. However, like with the ELS, HSLS respondents are quite young, which does not allow us to analyze noncompletion.

or adverse shocks to a parent) or "victimization shocks" (e.g., seeing a shooting or being the victim of a crime). The most common shocks include experiencing a non-parental family death (50%) and no father in the household (24%). In accordance with the group differences observed in educational attainment, there are notable racial and SES disparities in exposure to adverse shocks, though not all gaps occur in the same direction. For example, some shocks, like exposure to divorce and bullying, are more common among white than Black students. And some shocks, like witnessing a shooting and having no father in the household, are more than twice as likely among Black relative to white respondents. All the adverse shocks were more (or just as) common among low-income households, suggesting that an examination of the role of shocks in individuals' lives is indeed relevant beyond the sample of highly disadvantaged youth that was the focus of Section 2.

Panel D of Table A1 summarizes NLSY97 respondents' beliefs about the future—their subjective probabilities of earning a bachelor's degree by the age of 30 and their subjective probabilities of experiencing a few shock events in the year following the first interview (when all respondents would be teenagers of high school age). Many of the patterns here echo those observed elsewhere in Table A1. For example, with the exception of getting drunk, Black students place significantly higher probabilities on a variety of adverse events, including being the victim of a crime, being arrested, dying, and having an early pregnancy. There are also strong income gradients in the likelihood of anticipating these adverse shocks, again with the exception of the "getting drunk" question. In sum, the descriptives presented in Table A1 are consistent with the narrative that evolved in the qualitative data: Black and economically disadvantaged youth have lower educational attainment, experience more adverse shocks in childhood, and *anticipate* more frequent adverse shocks in the future.

3.2 Verifying Narratives in the Large-N NLSY97 Dataset

In this section, we estimate descriptive multivariate regressions using the NLSY97 data that correspond to the themes described in Section 2. There, our analysis focused on an economically disadvantaged sample of Black youth in Baltimore. To demonstrate the external validity of these narratives, we use the entire NLSY97 sample, which is intended to be nationally representative. However, we corroborate that the same patterns exist, and in some cases are more pronounced, in subsamples of Black and low-income individuals in Appendix D.

We begin by showing that a history of adverse shocks is indeed related to lower educational attainment, as the stories of Sierra and Vicky in Section 2 suggested. Next, we demonstrate that students' beliefs about the future (expectations) are correlated with personal histories of adverse shocks just as they were for Taniya, Erica, and Matthew. Finally, we show that relatively low expectations about future educational attainment are associated with lower attainment, and we explore an explanation for the connections between shocks, expectations, and outcomes: students' decisions about where and how to attend postsecondary school, like the choice Rhiannon made to stay near her family. Individuals who are more pessimistic about their probability of completing a bachelor's degree, or about the likelihood of experiencing adverse events in their own futures, are less likely to enroll in four-year bachelor's degree programs.

Adverse Shocks Predict Educational Attainment: Columns 1 and 2 of Table A2 suggest this is so: we regress an indicator for dropping out of college, conditional on enrollment, on demographic and background variables as well as indicators for having experienced different types of "family shocks" (such as parental incarceration, unemployment, or death) or "victimization shocks" (like seeing a shooting or being the victim of a crime). Experiencing shocks generally, and three specific types of shocks (i.e., changing schools, seeing a shooting, and feeling unsafe), significantly predict dropping out of college.

Adverse Shocks Shape Beliefs and Expectations: Columns 3 through 6 of Table A2 similarly examine the ability of students to predict self-reported subjective probabilities of future events such as earning a college degree or experiencing an adverse shock. Experiencing any type of childhood shock significantly reduces respondents' educational expectations and increases their expectation of experiencing a future adverse shock. And the same incidents that derail schooling are also the strongest predictors of low educational outcomes and of experiencing additional adverse shocks in the future.

Expectations and Beliefs Predict Educational Attainment: In Table A3 we regress indicators for bachelor's degree attainment (conditional on enrolling in some kind of postsecondary school) on sets of background, experienced shock, college expectation, and anticipated shock variables. Columns 3 and 4 show that, even net of demographic background, expectations regarding college completion and future adverse shocks, respectively, predict educational attainment. Column 5 includes demographic background, shock history, college expectations, and anticipated shock variables in the same model and finds qualitatively similar results. Together with the qualitative evidence described in Section 2, the descriptive patterns documented in Table A3 suggest a relationship between educational expectations and college completion. Moreover, the subjective probability of death remains a strong predictor of attainment after conditioning on educational expectations and prior shocks. This suggests two things. First, expectations about future shocks are associated with expectations about educational success. Second, anticipating extremely adverse shocks (e.g., death) affects educational success.

A specific mechanism through which expectations may translate into educational outcomes is the choice of postsecondary institution type, as with Elijah and Rhiannon. We probe this hypothesis in the NLSY97 data by estimating models similar to those in Table A3, but replacing the graduation outcome with an indicator for initially enrolling in a two-year program. The results are presented in Table A4. These estimates are consistent with the qualitative and quantitative results described thus far. Specifically, prior shocks such as feeling unsafe, changing schools, and being bullied predict choosing a two-year postsecondary program. So too does expecting future adverse shocks (e.g., death). Finally, there is a clear negative gradient in the relation between expecting to complete a four-year degree and enrolling in a two-year institution. Institution type selection is, therefore, a candidate mechanism through which attainment and adverse shock expectations affect educational outcomes. This response to anticipated shocks may explain why so many of our interview respondents ended up at short-term credentialing institutions—their choice of postsecondary institution was in part an expression of their beliefs about, or fears for, the future.

4 Structural Model: Post-Secondary Education with Noncompletion Risk

4.1 Model Specification

To this point, we have used qualitative interviews and a nationally representative data set to provide evidence that adverse shock events are common in the lives of disadvantaged youth. Moreover, we have demonstrated that these shocks are associated with more frequent derailment of educational trajectories, students dropping out or transferring into shorter and less lucrative programs, and more generally, the ways in which students think about the future, financial crisis, family tragedy, and death. If we take our evidence of such a story seriously, we might also take seriously the possibility that reverberations of these shocks influence outcomes such as educational trajectories, which in turn may have implications for policies designed to increase educational retention or attainment.

In this section, we assess this possibility with a structural model of dynamic discrete choices. Each year, agents in the model can choose to invest in their post-secondary education by attending two-year- or four-year-degree granting programs or to not attend school. A crucial feature of the model is that the agent may choose to enroll in a year of education but not complete it. Resulting noncompletion is operationalized in the model as a probability faced by each agent that, even if they pay the upfront cost of enrolling in school, their accrued educational attainment does not increase. The increment of educational attainment is thus stochastic, and enrolling is a risky investment. The probability of noncompletion is computed using the NLSY97 and varies by demographic characteristics that capture different sources of disadvantage. We show that model estimates—and thus policy conclusions—hinge on the noncompletion rates we incorporate into the model. While the probabilities we estimate using the NLSY97 are likely an improvement upon simply assuming away noncompletion risk, direct measures of noncompletion probabilities that agents hold when making education decisions would be better still.

Model Specification: In each period t, $t = 1, ..., T-1$ (corresponding to ages 18 to 29), agents indexed by i maximize their expected lifetime utility by choosing to work and earn income or to enroll in a four-year (bachelor's) degree program or a two-year (associate's) degree program. Period T (corresponding to age 30) is a terminal period in which agents receive as a lump sum the present-discounted value of the utility that would be generated by an infinitely-lived agent consuming out of income corresponding to their highest degree attained.

State Variables: The state variables that play a role in the decision-making process include demographic and background variables as well as educational histories. The four key background variables we use to capture potential sources of disadvantage or difficulty that could affect noncompletion and thus educational trajectories: indicators for underrepresented minority ("URM") status, a low-income family of origin ("Low-income"), a history of adverse shocks ("Shocks"; at least two of the shocks listed in Table A1), and poor high school performance ("Low GPA"; a high school grade point average below 3.0 on a 4.0 scale). These four binary variables are summarized below as X_i, and we use possible combinations of their values to generate sixteen distinct groups of agents. Age (denoted age_{it}) is also a relevant state variable, since it dictates the time remaining in the model for each agent to complete their education. The variable y^s_{it} represents the total years *completed* by individual i at time t and in each enrollment type $s \in \{1, 2\}$, where superscript $s = 1$ represents two-year programs and $s = 2$ four-year programs. $r^s_{i,t-1}$ indicates enrollment and completion of a year from school type $s \in \{1, 2\}$ in the prior period. D^s_{it} indicates possessing a degree from a school of type $s \in \{1, 2\}$. Finally, we assume that $D_{it} \in \{0, 1, 2\}$ (absent the superscript) indicates the highest degree attained prior to period t. We collect all these state variables, including X_i, in the vector Z_{it}.

Choices and Flow Utility: Agents derive utility from earned income and, if enrolled in school, the utility cost of one year in education, with the choice for agent i at time t denoted d_{it}. Options are work ($d_{it} = 0$), which generates income, or educational enrollment ($d_{it} = 1$

for a two-year program and $d_{it} = 2$ for a four-year program).[18] The income generated by work depends on degree attainment status not only after the model ends at age 30 but in any year before that when individuals decide to work—if an agent has already earned a degree, they earn a larger income. The flow utility function of choice $d \in \{0, 1, 2\}$[19] is specified as:

$$\mu_d(Z_{it}) + \varepsilon_{itd} = \frac{(e_{D_{it}})^{1-\gamma}}{1-\gamma} + u_d(X_i) + \varepsilon_{itd}, \tag{1}$$

where $e_{D_{it}}$ is income (measured in ten-thousands of 2013 dollars) generated by highest degree D_{it}, and $u_d(X_i)$ is the utility of the schooling choice (which could be positive or negative), where we normalize the flow utility of the non-school option to zero ($u_0(X_i) = 0$). The ε_{itd} is an unobserved mean-zero idiosyncratic utility factor distributed Type 1 Extreme Value. Individuals derive utility from income e through a constant relative risk aversion (CRRA) utility function with parameter γ. The estimated parameters of interest are the $u_d(X_i)$, which capture enrollment utility relative to the non-school option.

Noncompletion Probabilities: A key feature of our model is that while educational *enrollment* is a choice, educational *completion* is the result of a random draw; an individual's probability of noncompletion in each period of enrollment is denoted $\alpha^s(X_i)$, again for each enrollment type $s \in \{1, 2\}$, where superscript $s = 1$ represents two-year programs and $s = 2$ four-year programs. We describe how we calculate $\alpha^s(X_i)$ directly from the data in Section 4.2. Agents rationally forecast their probability of noncompletion based on the rate of its occurrence in their demographic group, i.e., they know the probability $\alpha^s(X_i)$ of noncompletion but must make enrollment decisions before a completion outcome is realized. If an individual enrolls but does not complete the educational year, they incur the flow utility cost of a year in school $u_d(X_i)$, but their total years of education do not increase in the following period. When deciding whether and in what type of program to enroll, agents in the model consider flow utilities and the enhancement to future income that education can yield but also the risk of derailment and the opportunity cost of lost earned income in an educational pursuit that does not result in such a future income enhancement.[20]

[18]Our focus on two-year- versus four-year degrees is due to data limitations. In the NLSY97, students are asked about enrollment in these two types of institutions.

[19]For clarification, $s \in \{1, 2\}$ can take two values, one for each of the two educational enrollment options, while $d \in \{0, 1, 2\}$ can take three values, corresponding to the two educational options along with the third option $d_{it} = 0$ of not choosing school.

[20]We do not model a direct cost to re-enrolling after noncompletion, but as time T approaches, the fact that individuals are "running out of time" to complete a degree will mechanically reduce the option value of enrollment.

State Transition Probabilities: At $t = 1$, all values in Z_{it} are equal to zero except X_i and age_{it}. All variables contained in X_i are constants, and age evolves deterministically ($age_{it} = age_{i,t-1} + 1$). Let $r^s_{i,t-1} \in \{0, 1\}$ indicate enrollment and completion for educational choices $s \in \{1, 2\}$, i.e., $r^1_{i,t-1} = 1$ if the student enrolled in and completed a year at a two-year institution and is equal to zero if the student did not enroll in a two-year institution or did enroll but did not complete the year. Completed years evolve according to the following formula for all $s \in \{1, 2\}$:

$$y^s_{it} = y^s_{i,t-1} + r^s_{i,t-1}. \tag{2}$$

For example, if an agent had completed two years in a bachelor's program as of time $t - 1$, chose $d_{it} = 2$ at that time, and received a favorable completion draw, their number of completed bachelor's program years would increment up to three at the start of period t. With an unfavorable draw, $r^2_{i,t-1} = 0$, the recorded number of completed educational years for that agent would remain two.

Students who enrolled in a school of type $s \in \{1, 2\}$ at time $t - 1$ will have a degree of that type as of time t if they have completed the requisite number of years at that type of school; individuals who were not enrolled at time $t - 1$ retain their degree attainment status from that period to time t. This progression is captured by state variables D^s_{it}, so that:

$$D^1_{it} = D^1_{i,t-1} + \left(1 - D^1_{i,t-1}\right) \left(\mathbb{1}\left[y^1_{i,t-1} = 1\right]\right) \left(\mathbb{1}\left[r^1_{i,t-1} = 1\right]\right) \tag{3}$$

$$D^2_{it} = D^2_{i,t-1} + \left(1 - D^2_{i,t-1}\right) \left(\mathbb{1}\left[y^2_{i,t-1} = 3\right]\right) \left(\mathbb{1}\left[r^2_{i,t-1} = 1\right]\right). \tag{4}$$

That is, if an agent lacks a degree of the relevant type, has one completed year in an associate's program or three in a bachelor's program, and successfully completes another in the relevant program type, then they are recorded as having earned the degree for that program type. Each degree attainment status is an absorbing state (i.e., once a student possesses a degree, they cannot lose it). Annual income is affected by an agent's highest attained degree status, so earning a bachelor's degree while already possessing an associate's affects earnings, but not the reverse.[21]

The Dynamic Programming Problem: Agents make enrollment decisions to maximize their expected lifetime utility, which means they take expectations over future realizations of ε_{itd} as well as the evolution of their experience and degree-earning state variables, in part

[21] While earning an associate's degree and then transferring to a bachelor's program with a year or two worth of existing credit is possible in principle, actually earning a bachelor's degree on this path is rather rare (see discussion in Odle and Russell (2023)), so we do not incorporate this in the model. When calculating α for each group, however, we do try to account for the cases in which such a transfer does occur; see Section 4.2.

dictated by draws against $\alpha^s(X_i)$ if they ever decide to enroll in school. A solution to an agent's problem is a set of policy functions that map each period's potential sets of state variables to choices. Under standard assumptions, including stationarity and the distribution of the flow utility error term ε_{itd}, we can rewrite this problem using a Bellman equation. Using derivations discussed in Rust (1987), we can express the net-of-error value of expected lifetime utility of choice d given state variables Z_{it}, denoted $\bar{v}_d$ as follows:

$$\begin{aligned}\bar{v}_d\left(Z_{it}\right) = \mu_d(Z_{it}) &+ \mathbb{1}\left[d=0\right]\beta E\left[V(Z_{i,t+1}|Z_{it})\right] \\ &+ \mathbb{1}\left[d=s\right]\left\{\alpha^s(X_i)\beta E\left[V(Z_{i,t+1}|Z_{it}, r^s_{it}=0)\right]\right. \\ &\left.+(1-\alpha^s(X_i))\beta E\left[V(Z_{i,t+1}|Z_{it}, r^s_{it}=1)\right]\right\}\end{aligned} \tag{5}$$

for $s \in \{1,2\}$, β is the annual discount rate, $\mathbb{1}[\cdot]$ is an indicator function, and V is the value function that captures continuation payoffs, including future realizations of $\alpha^s(X_i)$ and ε_{itd}. We solve the dynamic programming problem via backward induction, starting with optimal enrollment choices in the last period in which decisions are made ($T-1=12$, or age 29) based on the potential to earn a new degree and enhance the future earnings stream, and working backwards to time $t=1$.[22]

4.2 Empirical Implementation, Estimation, and Identification

Empirical Implementation: We continue to use the NLSY97 data for this structural estimation because they provide us with lengthy individual histories of educational completion and post-educational earnings. The sample we use is limited to individuals with information on the state variables described above and observed education choices between ages 18 to 29, leaving 6,004 individuals.[23] A summary of this sample is provided in Table 2 for the full sample and for each of the possible combinations defined by the four binary indicators: URM, Low-income, Shocks, and Low GPA. Panel A defines each group (e.g., Group 1 is not URM, not Low-income, did not experience shocks, and had a high school GPA of 3 or higher). In Panel B, we show graduation rates conditional on enrollment for each group (which are used to construct noncompletion rates). For example, Group 8 individuals (URM, low-income, multiple shocks, high GPA) complete an associate's degree within two years of starting it with 20% probability and a bachelor's degree within four years with a 29% probability. Group 1 individuals, who are not URM or low-income and did not experience shocks,

[22] At T, agents receive $\frac{1}{1-\beta}$ times the utility of the average annual income earned by individuals with their particular degree attainment status.

[23] This is smaller than the full sample used in Section 3, but larger than estimation samples that require other covariates and outcome variables (Tables A2-A4), particularly because the state variables used to construct the structural estimation sample are also present in the models estimated in Section 3.

finish an associate's degree within two years with 43% probability and a bachelor's degree within 4 years with 41% probability, conditional on enrollment.

We assume that all individuals who work in a given decision period, as well as all individuals throughout their infinite lives after time T, earn an approximation of the typical income earned in the NLSY97 data by workers in their demographic and degree-attainment categories in 2013,[24] rounded to the nearest hundred. Demographic categories include the 16 groups indicated above; the five degree-attainment categories we use are bachelor's and associate's degree holders, bachelor's and associate's program dropouts, and those who never attended any postsecondary school.[25]

Some demographic-education combinations are rather rare, and averages can be significantly affected by outliers, so "typical incomes" for each group were estimated as follows. Income in 2013 was regressed on our four demographic indicators as well as indicators for the five educational outcome categories and gender.[26] Regression estimates capturing the effect of each demographic and educational outcome indicator *across* demographic groups were then used to predict incomes *within* those much smaller groups. Figures for each demographic group are reported in Table E1. For each agent i and in each period t, consumption equals income $e_{d,D_{it}}$, determined by d_{it} and D^d_{it}. This simplification eases estimation and elides concerns about income growth or interruptions in work after age 30 without obscuring the main point of our results.[27]

We assign rates of noncompletion ($\alpha^1(X_i)$ and $\alpha^2(X_i)$) in each type of postsecondary enrollment (associate's or bachelor's programs, respectively) for each of the sixteen demographic groups on which we estimate the model, i.e., there are 32 different non-completion rates used in the model. Obtaining these values is indirect because there are no explicit variables recording noncompletion. Instead, we assume fixed annual rates of noncompletion $\alpha^s(X_i)$, which we must infer from data on the number of times students enrolled and whether

[24] We measure in 2013 because NLSY97 respondents are in their late 20's and early 30's at this time and have mostly stopped attending school, and thus reported incomes have begun to stabilize for even the most educated.

[25] Allowing income to vary for each combination of years of schooling is a plausible extension to the model. We omit this variation in part because it is unclear if students with a single year of school and no two- or four-year degree, for example, have a certificate or some kind of coursework. We have experimented with various potential specifications that incorporate returns to a single year of post-secondary education that does not result in a degree and generally find that doing so has a relatively minor impact on our main results.

[26] A gender dummy was included because it is strongly correlated with GPA (positively) and whether an individual is employed in 2013 (negatively), so the relationship between income and having a low GPA can be reasonably estimated.

[27] The median earned income amount reported by individuals enrolled in postsecondary school is about \$4,300. Certainly, students are consuming in some form while in school, so we use this dollar amount as their base consumption with all additional utility coming from the flow utility of enrollment itself and the idiosyncratic error term.

or not they completed a degree. For example, some students who enroll in an associate's degree program may report an associate's degree after exactly three years rather than two. This would suggest that at least one year did not count fully (i.e., was disrupted by a shock of some sort), which we interpret as a single instance of noncompletion. The $\alpha^s(X_i)$ must account for these instances in addition to dropouts, since our shocks do not necessarily lead to dropping out, so we need to develop a method for calculating their annual probability.

Our preferred approach uses on-time graduation rate for each group in each program type. On-time graduation reflects an educational history devoid of *either* a continuously enrolled noncompletion or a dropout. Thus, if p is the probability that an individual experiences either of these two types of educational derailment in a given year, m is the number of years an on-time student would need to graduate, and π is the on-time graduation rate in the relevant group and program type,

$$\pi = (1 - p)^m. \tag{6}$$

For instance, the on-time bachelor's degree graduation rate in Group 1 is 41 percent. Solving equation 6 with $\pi = .41$ and $m = 4$, we find that $p = 0.200$, which is the noncompletion rate

Table 2: Educational Outcomes, Noncompletion, and Model Results by Group in Structural Model Sample

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Panel A: Group definition																
URM	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Low-income	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Shocks	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes
Low GPA	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: On-time graduation rates																
Assoc. 2 Yr	0.43	0.23	0.40	0.24	0.35	0.22	0.28	0.20	0.23	0.16	0.19	0.16	0.21	0.14	0.19	0.14
Bach. 4 Yr	0.41	0.27	0.29	0.24	0.42	0.29	0.38	0.29	0.31	0.19	0.16	0.22	0.29	0.16	0.25	0.18
Panel C: Derived noncompletion probabilities																
α_1	.343	.524	.368	.514	.408	.530	.467	.547	.519	.604	.560	.606	.546	.626	.570	.632
α_2	.200	.280	.264	.297	.197	.264	.214	.264	.257	.335	.372	.312	.265	.367	.293	.353
Panel D: Utility parameter estimates with rational consideration of failure probability																
$u_1(X_i)$	-1.25	-1.04	-2.31	-1.98	-1.40	-1.30	-2.13	-2.28	-1.24	-1.16	-1.95	-1.85	-1.44	-1.48	-2.50	-2.33
	(.031)	(.031)	(.032)	(.033)	(.031)	(.032)	(.034)	(.034)	(.032)	(.032)	(.033)	(.034)	(.033)	(.032)	(.035)	(.035)
$u_2(X_i)$	-0.82	-1.11	-1.61	-2.08	-1.38	-1.56	-2.37	-2.70	-1.57	-1.62	-1.98	-2.68	-1.95	-1.79	-3.01	-2.90
	(.022)	(.022)	(.022)	(.026)	(.023)	(.024)	(.028)	(.030)	(.025)	(.026)	(.025)	(.033)	(.028)	(.025)	(.033)	(.032)
Panel E: Utility parameter estimates with no consideration of failure probability																
$u_1(X_i)$	-1.75	-2.70	-3.68	-4.17	-2.63	-3.22	-4.06	-4.80	-3.10	-3.49	-4.37	-4.80	-3.54	-4.16	-5.23	-5.75
	(.031)	(.035)	(.044)	(.053)	(.034)	(.039)	(.051)	(.057)	(.038)	(.042)	(.055)	(.057)	(.043)	(.053)	(.053)	(.046)
$u_2(X_i)$	-1.41	-2.34	-2.87	-3.64	-2.21	-2.82	-3.48	-4.18	-2.77	-3.27	-3.90	-4.51	-3.25	-4.73	-4.62	-5.08
	(.022)	(.025)	(.027)	(.034)	(.024)	(.027)	(.032)	(.037)	(.027)	(.031)	(.036)	(.038)	(.030)	(.035)	(.036)	(.032)
Panel F: Two-year (Associate's) enrollment probability																
Baseline	0.06	0.09	0.05	0.07	0.07	0.09	0.07	0.07	0.10	0.11	0.08	0.10	0.10	0.09	0.06	0.08
3/4 Rate	0.05	0.07	0.04	0.06	0.06	0.07	0.05	0.08	0.08	0.06	0.08	0.08	0.08	0.07	0.05	0.06
1/2 Rate	0.04	0.05	0.03	0.04	0.05	0.05	0.04	0.06	0.06	0.04	0.05	0.06	0.06	0.04	0.04	0.03
Panel G: Four-year (Bachelor's) enrollment probability																
Baseline	0.26	0.19	0.16	0.09	0.19	0.17	0.10	0.08	0.12	0.11	0.06	0.06	0.10	0.11	0.04	0.07
3/4 Rate	0.28	0.26	0.23	0.18	0.23	0.23	0.15	0.16	0.18	0.20	0.16	0.15	0.17	0.24	0.09	0.20
1/2 Rate	0.29	0.29	0.27	0.25	0.25	0.27	0.20	0.23	0.23	0.27	0.26	0.24	0.23	0.31	0.18	0.29
Observations	940	264	171	189	382	181	152	206	726	507	239	430	399	360	285	573

Columns represent subsamples of respondents who had certain characteristics or not as indicated by Panel A: individuals from URM (underrepresented minority) groups; those from low-income households (total less than $35,000); those who experienced at least two of our recorded adverse shocks in childhood; and those whose high school GPA was below 3.0. Panel B displays on-time graduation rates for those in each group who enroll in the relevant program type. The α in Panel C are annual probabilities of academic noncompletion determined using values in Panel B, which differ between two-year programs ($\alpha^1(X_i)$) and four-year programs ($\alpha^2(X_i)$). Utility parameter estimates are in Panels D and E. In Panel D, individuals rationally consider probabilities of noncompletion; in Panel E, we assume they ignore these probabilities. Parameters u_1 and u_2 are the flow utility values of attending an associate's and a bachelor's program, respectively. Predicted choice probabilities for enrollment at age 18 and different noncompletion probability levels are presented in Panels F (two-year programs) and G (four-year). "3/4 Rate" rows contain enrollment probabilities when both noncompletion probabilities (for associate's and bachelor's programs) are reduced to three-fourths the rate reported in Panel C; "1/2 Rate" rows reduce both noncompletion probabilities to half that base rate.

we use for Group 1 students in four-year programs in the structural model. We have experimented with alternative methods to calculate the $\alpha^s(X_i)$ that use additional information on trajectories, e.g., that take explicit account of all possible permutations of years completed versus enrolled that support the data.[28] We discuss these approaches in Appendix E, but find that they yield $\alpha^s(X_i)$ that are virtually identical to values we calculate with our preferred method.

To avoid counting part-time enrollees as not graduating on time when they accrue too many "years in school," we count part-time enrollment as a half-year. We also account for the possibility that individuals may report attending two-year schools with the goal of transferring to a four-year program and attaining a bachelor's degree rather than an associate's. If those individuals do eventually earn a bachelor's degree and never accrue more than two years enrolled at a two-year school, they are treated as having "graduated on time" for the purposes of calculating the associate's program noncompletion rate.[29]

We report the $\alpha^s(X_i)$ we construct in Table 2, Panel A. For instance, our calculations suggest that Group 1 individuals experience one of these kinds of educational setbacks in around 34 percent of years enrolled in associate's programs and in 20 percent of years enrolled in bachelor's programs.[30] Rates are higher for all other groups, reflecting the disadvantaged status of agents who are URM, come from a low-income household, or a history of prior adverse shocks.[31]

Identification and Estimation of Structural Utility Parameters: Identification ar-

[28]For example, if we observe that an individual enrolled six different years in a four-year granting institution and the obtained a four-year degree, there are many possible sequences of noncompletion, which a more elaborate procedure could explicitly account for.

[29]These outcomes are relatively rare, especially for disadvantaged groups. Accounting for this complex educational path has a noticeable reductive effect on non-URM student noncompletion rates, since white students earn a bachelor's degree after such transfers at a meaningful (though not large) rate. But URM noncompletion rates are almost completely unaffected.

[30]It is difficult to calculate similar values in the ELS or HSLS for comparison since those surveys were not performed every year like the NLSY97. However, it is worth noting that 36 percent of ELS respondents who enroll in some postsecondary school never earn a degree of any kind and a further 11 percent earn only a certificate of some sort, to say nothing of those who may take extra years to finish. The National Center for Education Statistics (2022) find that, in keeping with data from the recent past, around two-thirds of four-year-program enrollees in the U.S. earn a degree within six years. DeLuca et al. (2016) compiled data for colleges in the Baltimore area that show overall graduation rates as low as 14 percent, with many around one-third, in bachelor's degree-granting institutions. All these numbers are generally in keeping with the numerical values we employ here.

[31]It is worth noting that these data-based α values are subject to selection, since individuals with very high expected noncompletion risk are more likely to never enroll in the first place, and thus not show up in our on-time graduation calculation. It's not immediately clear what cross-group differentials we should expect in this selection effect, but it seems reasonable to guess that it might be stronger in groups with less enrollment—i.e., our disadvantaged demographic groups.

guments for the type of dynamic discrete choice model we have specified here are standard and discussed in Magnac and Thesmar (2002). If we specify how agents form expectations over transitions to future states and fix the discount factor β, we can identify the utility parameters $u_d(X_i)$, subject to a normalization. We first impose the assumption that agents in the model hold rational expectations calculated from the data as discussed in the previous section; in fact, this is one of our main points: young people making educational decisions act on rational beliefs about their probability of completing an educational program, which is a function of their demographic characteristics and earlier life circumstances. As we show, moreover, different $\alpha^s(X_i)$ would lead to different estimates of the utility parameters.

We thus set β to 0.95 and normalize the utility cost of choosing to work to zero. This means that the utility costs of education are properly interpreted as relative to the cost of not going to school, i.e., working. As noted, we also assume that the unobserved idiosyncratic utility factors ε_{itd} are drawn from a Type 1 Extreme Value distribution. Finally, we fix the CRRA parameter $\gamma = 0.75$, which lies in a range established by prior estimates in the literature (see, for instance, Hurd (1989) or Blau and Gilleskie (2008)).[32] Given these assumptions, agents' choices in the data identify the $u_d(X_i)$.[33]

With the model and noncompletion rates specified, we estimate the $u_d(X_i)$ separately for each of the sixteen demographic groups via maximum likelihood in a nested fixed-point algorithm (Rust, 1987), using the NLSY97 data. Given candidate parameters, values of the net-of-error choice specific value functions $\bar{v}_d(Z_{it})$ are calculated by backward induction from time T (age 30) based on the relevant group's value of the noncompletion probabilities $\alpha^s(X_i)$. Given distributional assumptions on ε_{itd}, the $\bar{v}_d(Z_{it})$ are used to calculate the implied probability that each individual would make the work or enrollment choice they did in each year they appear in the NLSY97 data, and this probability becomes their contribution to

[32]While other values of γ affect the absolute magnitudes of our estimates of the $u_d(X_i)$, they do not affect the qualitative patterns we show. In principle, our setup would allow us to estimate γ, but doing so would require more variation in income and thus consumption arising from agents' choices, which would in turn require a more elaborate model.

[33]It is important to note that the identification arguments we rely on focus on the existence of a unique set of parameters that best fit the data—in this case, that maximize the log likelihood function. There remains a possibility of bias in estimates. In our case, a concern might be, for example, that individuals who faced adverse shocks systematically over-estimate the probability of completion relative to what we observe in the NLSY97 data (and use to model their expectations). If so, students go to school not because it is a low-utility-cost option, but because they believe they are likely to finish and thus over-estimate the payoff. This would lead us to over-estimate the utility of school (or under-estimate the utility cost of school). Our data work, however, has shown that stated expectations are predictive of choices and outcomes. Moreover, we argue that using estimates of completion rates from the NLSY97 data, despite the types of concern we raise here, leads to better estimates than assuming noncompletion risk is zero, which is not empirically supported. Below, we illustrate the impact on parameter estimates and policy conclusions if we ignore noncompletion risk. Broadly, this type of concern underscores why direct measurement of beliefs about noncompletion could improve estimates.

the overall likelihood. The natural logs of these probabilities are summed, and the algorithm finds the parameters that maximize this sum.[34]

4.3 Utility Parameter Estimates and Interpretation

Utility parameter values for each group on whom the model is estimated $u_d(X_i)$ are reported in Panels D and E of Table 2.[35] The key result to note comes from the comparison of the estimated valuation students apply to schooling in two divergent cases, one in which we model them as rationally worrying about the possibility of noncompletion and another in which we assume they do not account for it. Individuals are estimated to value school much less if we do not account for their rational expectations about noncompletion and derailment.

A translation of the utility value of school into dollars is reported for both models in Table E2.[36] The results make clear that when we do not account for individuals' concerns about noncompletion, we overestimate their distaste for school: with $\alpha^s(X_i)$ in the model, two-year programs generate negligible utility penalties, and the disutility of four-year enrollment is valued in the dozens or hundreds of dollars (e.g., \$729 per year for Group 4, minority low-income individuals with little or no shock history and a high GPA); when we assume $\alpha^s(X_i)$ away, we estimate that individuals' disutility from such programs is thousands of dollars per year (e.g., \$6,848 for Group 4).[37] Ignoring the anticipated probability of noncompletion means we estimate disutility from enrollment in a bachelor's program by about an order of magnitude.

We can demonstrate the basic role of $\alpha^s(X_i)$ in the model by asking a straightforward question: given students' estimated *actual* valuation of school under rational expectations, what would be the effect of reducing their anticipated probability of derailment? Using the parameters estimated in our main specification, we calculate enrollment probabilities at age 18 under reduced noncompletion rates for all groups. Results are reported in Panels F and G of Table 2. If noncompletion rates are reduced by one quarter of their existing level (and individuals are allowed to see and anticipate this change), enrollment in bachelor's degree programs increases by 2 percentage points (about 8 percent of the baseline rate) for Group

[34]See Appendix E for an expression of the likelihood function, as well as details regarding the computation of standard errors.

[35]The model fits group-specific choices virtually perfectly largely because there are separate utility parameters estimated for each group and educational option.

[36]This appendix also reports the numerical values that generate Figure 1.

[37]In the no-α model, estimates of schooling disutility are hundreds or thousands of dollars larger for all other groups. Keane and Wolpin (1997) estimate a "cost" of college attendance, including consumption value and tuition, of around \$3,000 for young white men in the 1979 cohort of the NLSY while not modeling rational noncompletion risk assessment, so our no-α estimates are roughly similar to those from other models that ignore such risk.

1,[38] while enrollment in associate's programs actually declines by about 1 percentage point for the same group. Indeed, two-year enrollment falls slightly for all groups under this α-reduction. Four-year program enrollment increases more for most groups (e.g., for Group 4 the increase is from 16 percent to 23 percent), given the greater importance of α to their decision-making process. All these effects are unsurprisingly stronger when noncompletion is further reduced to half its true frequency.

These numbers suggest the potential importance of the risk of derailment in individuals' decisions about their education. Programs aimed specifically at student support and retention (Evans et al., 2020) may affect student decision-making on this margin, increasing educational attainment and reducing inequality. However, it may be unrealistic to assert that noncompletion risk could be reduced as much as is supposed in this analysis. What does our model have to say about the potential effect of a more straightforward policy approach: subsidies paid to enrolled students? Does the presence of $\alpha^s(X_i)$ in the model affect our predictions?

4.4 Impacts of a Counterfactual Subsidy

Here, we calculate enrollment probabilities at age 18 for individuals from each group, first under parameters estimated using our main specification, then under parameters from our no-α specification. We compare the effects across the two cases of adding increments of 0.1 to the consumption value e if an individual is enrolled in a four-year educational program in a given period, effectively a subsidy to bachelor's students in increments of \$1,000 per year enrolled. Table E3 presents the effect of a \$10,000 subsidy. Enrollment in two-year programs by all groups is reduced to meager rates in response to the four-year financial incentive in both models. Increases in four-year enrollment in both models are dramatic. The subsidy is predicted to increase bachelor's degree program enrollment rates at age 18 by 17 percentage points for Group 1 when we acknowledge that students rationally anticipate the probability of noncompletion and dropping out. This prediction changes to a 14 percentage point increase in enrollment if we use utility estimates generated by a model that excludes expectations about noncompletion. For smaller \$1,000 subsidies, four-year enrollment increases by group fall in the 3 to 5 percentage point range.[39] The predicted subsidy effect is, similarly, meaningfully larger in the α-inclusive model for all groups.

Our estimated parameters do not produce precisely identical predictions across the two α-cases, even for the zero-subsidy scenario—age-18 enrollment rates with no subsidies are

[38] Group 1 are non-URM, high-income household individuals with little history of shocks and a high GPA.

[39] Surveying prior work, Deming and Dynarski (2009) find that the best estimates predict that a \$1,000 decrease in tuition costs increase college enrollment by about 4 percentage points, so again our exercises yield reasonable effect sizes.

higher across the board for the rational-α case—but our focus is on the *growth* of this difference as the subsidy value increases. We can see this clearly in Figure 2, Panel (a), in which we depict the change in enrollment rates within demographic groups that results from increasingly large subsidies to students. When individuals consider noncompletion risk (and their utility parameters are estimated under this consideration), a subsidy alleviates the financial concern associated with that risk; associate's degree program enrollment falls, and bachelor's degree program enrollments skyrocket for each group. Enrollment increases at similar rates at the lowest subsidy values among disadvantaged individuals as among those in Group 1; however, Group 1 enrollment levels off somewhat while larger subsidies continue to grow enrollment in disadvantaged groups. This can be seen, for instance, in non-minority, low-income individuals (Group 3) as well as minority students who have a history of adverse shocks (Group 6), whose respective enrollment rates are both up about 15 percentage points with a \$5,000 subsidy and 20 percentage points with a \$10,000 subsidy. The counterfactual policy closes enrollment gaps between advantaged and disadvantaged groups, more significantly so as the size of the subsidy grows.

Moreover, there is an important and large gap in the predicted effect of the subsidy between the two versions of our model. We can see this in Figure 2, Panel (b), which shows that the subsidy's effect is predicted to be larger in the rational model for all groups, and its effect is in particular larger for individuals who are more disadvantaged and more likely to be derailed. Looking at students with high GPA, the enrollment rate gap between the most well-off individuals (Group 1) and the least well-off (Group 8) in bachelor's enrollment rates closes as subsidies grow if we assume rational risk assessment (from 16 percentage points to 7 percentage points); however, if we ignore $\alpha^s(X_i)$ in the model, we predict that this gap is more stagnant (reducing from 13 percentage points to 10 percentage points). If we ignore $\alpha^s(X_i)$ we also meaningfully underestimate the policy effect on every group of students.

When we assume individuals ignore noncompletion risk and estimate their utility parameters under that assumption, subsidization does less to mitigate the disutility that enrollment is estimated to generate, particularly for groups with large true probabilities of noncompletion. There are smaller increases in bachelor's enrollment, and associate's enrollment falls in some better-off groups, indicating that some of the change for them may be coming from those who change between program types.

It should also be noted that the implications of this counterfactual exercise are all derived while holding derailment risk constant—increased subsidies are not permitted to affect $\alpha^s(X_i)$, only the utility value derived from enrollment. However, we might suspect that subsidy dollars could be used to avert the kinds of adverse shock events we have argued lead to such derailment, lowering that risk. For instance, a subsidy of sufficient size could substitute for lost family income or rent payments in the case of a "downfall" like those Sierra described in her interview. This might mean that a student whose family endures such an event can afford to keep taking classes, rather than transferring to a less lucrative or more local program (like Chanel and Vicky) or dropping out of school entirely (like Isaac). As we demonstrated above, if decision-makers are accurately assessing derailment risk, this could compound the positive enrollment effect that the model predicts subsidies would generate.

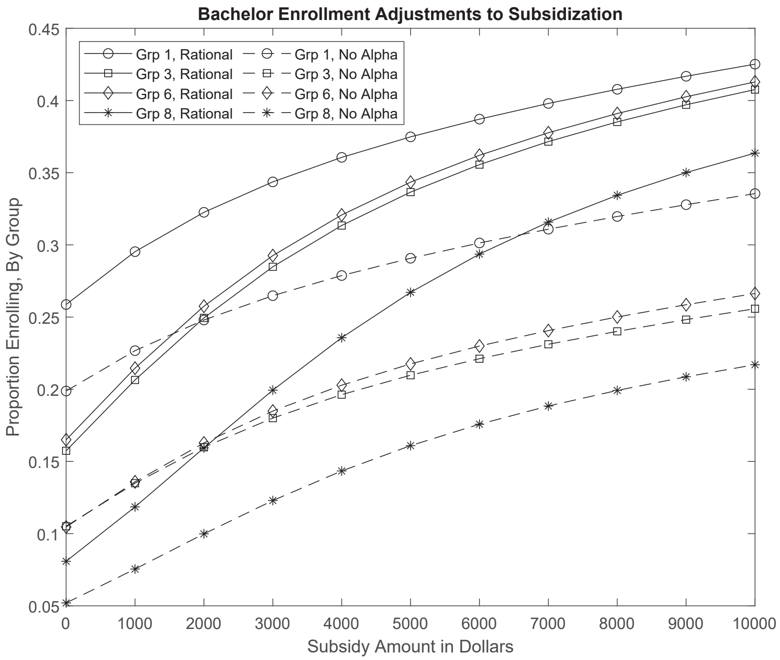
In summary, our results suggest that, by ignoring individuals' rational consideration of noncompletion and derailment risk in modeling their educational decision-making process, we lay the foundation for erroneous evaluation of the impacts of counterfactual policies, and therefore may discount the value of certain approaches to the related problems of enrollment and completion.

4.5 Welfare Analysis: Completion Probabilities versus Subsidies

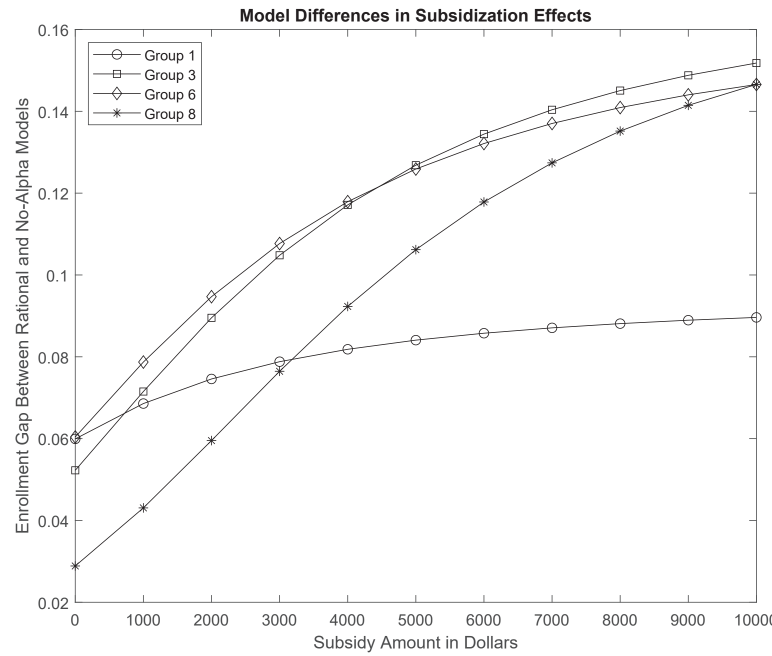
Using the structural model, we conduct a simple welfare analysis to further understand the consequences of noncompletion probabilities. In particular, we lower noncompletion probabilities and then compute a compensating variation by shifting subsidies. Essentially, we ask how large of a subsidy would be required to make individuals indifferent between it and a reduction in noncompletion probability. Using our α-inclusive utility parameter estimates, we simulate ten thousand individual education choice paths for each of the sixteen groups. These simulated paths each generate a series of utility values. We calculate the present discounted value as of age 18 and take the group average. We predict this average value for the base model, for a no-subsidy case in which each group's noncompletion probabilities $\alpha^s(X_i)$ are reduced by 10 percentage points—approximately the difference between non-URM, high-income students' noncompletion probabilities in bachelor's programs and those of URM, low-income students—and for various values of the student subsidy. This allows us to determine the yearly subsidy amount that would yield the same value as the ten percent decrease in noncompletion probability. Results are depicted in Figure A1.[40] The proportional average value across all groups is $2,412; that is, a 10-percentage-point reduction in noncompletion probabilities yields the same predicted lifetime utility increase as a subsidy to all four-year students of $2,412. The subsidy-equivalent value of α reductions tends to be larger for minority and otherwise disadvantaged groups; for instance, the projected lifetime benefit of a reduction in noncompletion for URM, low-income students with multiple adverse shocks in their backgrounds (Group 8) is equivalent to a student subsidy of $3,580. This is because noncompletion risk is a more prevalent consideration for these students than for those in advantaged groups, compared to the disutility of schooling. These results demonstrate

[40] Results are also found in Table E4.

Figure 1: Counterfactual Subsidy Evaluation with vs. without Noncompletion Risk



(a)



(b)

(a) Curves depict changes in the proportion of each demographic group enrolling in bachelor's programs at age 18 given increasing levels of subsidization for students. The solid lines depict decision-makers that rationally account for educational non-completion risk; the dashed lines, decision-makers that ignore this possibility (and have correspondingly estimated utility parameters). Groups are numbered as defined in Table 2. (b) Curves depict differences in the proportion of each demographic group enrolling in bachelor's degree programs at age 18 between the baseline ("rational") and no-α models, given increasing levels of subsidization for students. Groups are numbered as defined in Table 2.

that ignoring non-zero noncompletion rates can not only lead to biases in estimated utility parameters and thus erroneous policy conclusions (as described above) but also hide a source of welfare losses. Relatedly, doing so ignores the potential value of targeted support and retention programs, which could provide significant value to students and, indeed, may provide that value at lower cost than direct subsidization. Each policy increases enrollment and completion rates, but a model that more completely captures the mechanisms by which they do so allows for better-founded, more useful counterfactual policy evaluation and comparison.

5 Discussion

This paper introduces a novel, interdisciplinary mixed-methods approach that uses systematic analyses of rich qualitative data to guide the specification of a structural economic model of dynamic choice. Our approach is rooted in a crucial link between qualitative data gathered from open-ended interviews and structural modeling: at their core, both suggest narratives of behavior and decision-making. This fundamental link—which we have used in this paper to further our understanding of non-traditional post-secondary educational pathways—can be especially useful in contexts where data are lacking, the context is not well-understood, or the researcher is faced with myriad plausible modeling choices. In such cases, existing narratives, sometimes formalized as models based on extant data, may not fully reflect the barriers, circumstances, and constraints that individuals face or the rational thought processes in which they engage. This can lead to model misspecification; for example, researchers may inappropriately view decision-makers as irrational or deficient since they appear to make puzzling decisions, or as having nonstandard or self-defeating preferences that they do not in fact harbor. Seeking insight from decision-makers themselves and applying qualitative analytical methods to their responses could improve model specification and thus help to generate more useful policy proposals in these areas, and we hope this paper contributes to forging a path in that direction.

References

G. A. Akerlof. Sins of Omission and the Practice of Economics. *Journal of Economic Literature*, 58(2):405–418, 2020.

K. L. Alexander, D. R. Entwisle, and S. D. Bedinger. When Expectations Work: Race and Socioeconomic Differences in School Performance. *Social Psychology Quarterly*, 57 (4):283–299, 1994.

E. Anderson. *Streetwise: Race, Class, and Change in an Urban Community.* University of Chicago Press, Illinois, 1990.

P. Attewell, S. Heil, and L. Reisel. Competing Explanations of Undergraduate Noncompletion. *American Educational Research Journal*, 48(3):536–559, 2011.

H. S. Becker. *Tricks of the Trade: How to Think About Your Research while You're Doing It.* University of Chicago Press, Illinois, 1990.

H. S. Becker, B. Geer, E. C. Hughes, and A. L. Strauss. *Boys in White: Student Culture in Medical School.* University of Chicago Press, Illinois, 1961.

C. Belzil and M. Leonardi. Risk Aversion and Schooling Decisions. *Annals of Economics and Statistics*, 7-12(111):35–70, 2013.

P. Bergman, R. Chetty, S. DeLuca, N. Hendren, L. F. Katz, and C. Palmer. Creating Moves to Opportunity: Experimental Evidence on Barriers to Neighborhood Choice. *American Economic Review*, 114(5):1281–1337, 2024.

E. P. Bettinger and B. J. Evans. College Guidance for All: A Randomized Experiment in Pre-College Advising. *Journal of Policy Analysis and Management*, 38(3):579–599, 2019.

T. F. Bewley. A depressed labor market as explained by participants. *The American Economic Review*, 85(2):250–254, 1995.

D. Blau and D. Gilleskie. The Role of Retiree Health Insurance in the Employment Behavior of Older Men. *International Economic Review*, 49(2):475–514, 2008.

M. L. Boyd and S. DeLuca. Fieldwork with In-Depth Interviews: How to Get Strangers in the City to Tell You Their Stories. In J. M. Oakes and J. S. Kaufman, editors, *Methods in Social Epidemiology*, pages 239–253. Wiley: Jossey-Bass, New Jersey, 2017.

R. Bozick, K. Alexander, D. Entwisle, S. Dauber, and K. Kerr. Framing the Future: Revisiting the Place of Educational Expectations in Status Attainment. *Social Forces*, 88(5): 2027–2052, 2010.

A. S. Browman, M. Destin, M. S. Kearney, and P. B. Levine. How economic inequality shapes mobility expectations and behaviour in disadvantaged youth. *Nature Human Behaviour*, 3(3):214–220, 2019.

M. Brown, R. Chakrabarti, W. van der Klaauw, and B. Zafar. Understanding the Evolution of Student Loan Balances and Repayment Behavior: Do Institution Type and Degree Matter?. *Economic Policy Review*, 25(1):35–57, 2019.

E. Burland, S. Dynarski, K. Michelmore, S. Owen, and S. Raghuraman. The Power of Certainty: Experimental Evidence on the Effective Design of Free Tuition Programs. *American Economic Review: Insights*, 5(3):293–310, 2023.

B. Castleman and J. Goodman. Intensive College Counseling and the Enrollment Persistence of Low-Income Students. *Education Finance and Policy*, 13(1):19–41, 2018.

R. Chetty, M. O. Jackson, T. Kuchler, and J. Stroebel. Social Capital I: Measurement and Associations with Economic Mobility. *Nature*, 608(7921):108–121, 2022.

E. Chyn. Moved to Opportunity: The Long-Run Effects of Public Housing Demolition on Children. *American Economic Review*, 108(10):3028–3056, 2018.

R. Cox. Complicating Conditions: Obstacles and Interruptions to Low-Income Students' College 'Choices'. *Journal of Higher Education*, 87(1):1–26, 2016.

R. Darolia, C. Guo, and Y. Kim. The Labor Market Returns to Very Short Postsecondary Certificates., 2023.

S. DeLuca. Sample Selection Matters: Moving Toward Empirically Sound Qualitative Research. *Sociological Methods & Research*, 52:1073–1085, 2023.

S. DeLuca and E. Burland. Postsecondary Choices and Perceived Risk Among Low-Income High-Achieving Students: Leveraging Treatment Heterogeneity in College Access Interventions. Presentation at Annual Meeting of the American Sociological Association, 2023.

S. DeLuca, S. Clampet-Lundquist, and K. Edin. *Coming of Age in the Other America.* Russell Sage Foundation, New York, 2016.

S. DeLuca, L. F. Katz, and S. C. Oppenheimer. "When Someone Cares About You, It's Priceless": Reducing Administrative Burdens and Boosting Housing Search Confidence to Increase Opportunity Moves for Voucher Holders. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 9:179–211, 2023.

S. DeLuca, N. Papageorge, and J. Boselovic. Exploring the Tradeoff between Surviving and Thriving: Heterogeneous Responses to Adversity and Disruptive Events Among Disadvantaged Black Youth. *Russell Sage Foundation Journal of the Social Sciences*, 10(1): 103–131, 2024.

S. DeLuca, J. Groccia, N. Papageorge, and S. Sullivan. Sub-Baccalaureate Swirling: Understanding Timing, Trajectories, and Degree Completion for Community College Students. Mimeo, 2025.

D. Deming and S. Dynarski. Into College, Out of Poverty? Policies to Increase the Postsecondary Attainment of the Poor. NBER Working Paper 15387, 2009.

S. Dynarski, C. Libassi, K. Michelmore, and S. Owen. Closing the Gap: The Effect of a Targeted, Tuition-Free Promise on College Choices of High-Achieving, Low-Income Students. *American Economic Review*, 111(6):1721–1756, 2021.

K. Edin and L. Lein. *Making Ends Meet: How Single Mothers Survive Welfare and Low-Wage Work.* Russell Sage Foundation, New York, 1997.

W. N. Evans, M. S. Kearney, B. Perry, and J. X. Sullivan. Increasing Community College Completion Rates Among Low-Income Students: Evidence from a Randomized Controlled Trial Evaluation of a Case-Management Intervention. *Journal of Policy Analysis and Management*, 39(4):930–965, 2020.

F. Geiecke and X. Jaravel. Conversations at Scale: Robust AI-led Interviews with a Simple

Open-Source Platform. Available at SSRN: http://dx.doi.org/10.2129/ssrn.4974382, 2025.

S. Goldrick-Rab. *Paying the Price: College Costs, Financial Aid, and the Betrayal of the American Dream.* University of Chicago Press, Illinois, 2016.

N. Grigoropoulou and M. L. Small. The Data Revolution in Social Science Needs Qualitative Research. *Nature Human Behaviour*, 6:904–906, 2022.

B. Hart. Hanging In, Stopping Out, Dropping Out: Community College Students in an Era of Precarity. *Teachers College Record*, 121(1):1–30, 2018.

S. Hesse-Biber and R. B. Johnson, editors. *The Oxford Handbook of Multimethod and Mixed Methods Research Inquiry.* Oxford University Press, Oxford, 2015.

N. W. Hillman. College on Credit: A Multilevel Analysis of Student Loan Default. *Review of Higher Education*, 37(2):169–195, 2014.

M. M. Holland and S. DeLuca. "Why Wait Years to Become Something?" Low-income African American Youth and the Costly Career Search in For-profit Trade Schools. *Sociology of Education*, 89(4):261–278, 2016.

C. M. Hoxby and S. Turner. What High-Achieving Low-Income Students Know About College. *American Economic Review*, 105(5):514–517, 2015.

M. D. Hurd. Mortality Risk and Bequests. *Econometrica*, 57(4):779–813, 1989.

C. Iloh. Toward a New Model of College "Choice" for a Twenty-First-Century Context. *Harvard Educational Review*, 88(2):227–244, 2018.

C. Iloh and W. G. Tierney. Understanding For-Profit College and Community College Choice Through Rational Choice. *Teachers College Record*, 116(3):1–34, 2014.

A. A. Jack, editor. *The Privileged Poor: How Elite Colleges Are Failing Disadvantaged Students.* Harvard University Press, Massachusetts, 2019.

B. A. Jacob and T. Wilder. Educational Expectations and Attainment. NBER Working Paper 15683, 2010.

C. Jepsen, K. Troske, and P. Coomes. The Labor-Market Returns to Community College Degrees, Diplomas, and Certificates. *Journal of Labor Economics*, 32(1):95–121, 2014.

C. Jerolmack. What Good is Qualitative Literacy Without Data Transparency?. *Sociological Methods & Research*, 52:1059–1072, 2023.

N. Jones. *Between Good and Ghetto: African-American Girls and Inner-City Violence.* Rutgers University Press, New Jersey, 2010.

M. P. Keane and K. I. Wolpin. The Career Decisions of Young Men. *Journal of Political Economy*, 105(3):473–522, 1997.

M. P. Keane and K. I. Wolpin. The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment. *International Economic Review*, 42(4):1051–1103, 2001.

M. S. Kearney and P. B. Levine. Income Inequality, Social Mobility, and the Decision to

Drop Out of High School. *Brookings Papers on Economic Activity*, Spring 2016:333–380, 2016.

J. R. Kling, J. B. Liebman, and L. F. Katz. Bullets Don't Got No Name: Consequences of Fear in the Ghetto. In T. S. Weisner, editor, *Discovering Successful Pathways in Children's Development: Mixed Methods in the Study of Childhood and Family Life*, pages 243–281. The University of Chicago Press, Chicago, 2005.

V. Y. Liu, C. R. Belfield, and M. J. Trimble. The Medium-Term Labor Market Returns to Community College Awards: Evidence from North Carolina. *Economics of Education Review*, 44:42–55, 2015.

M. Lovenheim and J. Smith. Returns to different postsecondary investments: Institution type, academic programs, and credentials. In *Handbook of the Economics of Education*, volume 6, pages 187–318. Elsevier, 2023.

J. Ludwig and S. Mullainathan. Machine Learning as a Tool for Hypothesis Generation. *Quarterly Journal of Economics*, 139:751–827, 2024.

T. Magnac and D. Thesmar. Identifying Dynamic Discrete Decision Processes. *Econometrica*, 70(2):801–816, 2002.

R. K. Merton. Sociological Theory. *American Journal of Sociology*, 50(6):462–473, 1945.

R. K. Merton. *Social Theory and Social Structure.* Free Press, New York, 1968.

R. A. Mickelson. The Attitude-Achievement Paradox among Black Adolescents. *Sociology of Education*, 63(1):44–61, 1990.

R. Moffitt. Perspectives on the Qualitative-Quantitative Divide. *Poverty Research News*, 4 (1):5–8, 2000.

R. A. Moffitt. Comment on Income Inequality, Social Mobility, and the Decision to Drop Out of High School. *Brookings Papers on Economic Activity*, Spring 2016:385–391, 2016.

J. M. Morse. "Cherry Picking": Writing from Thin Data. *Qualitative Health Research*, 20 (1):3, 2010.

K. Myers and J. Oetzel. Exploring the Dimensions of Organizational Assimilation: Creating and Validating a Measure. *Communication Quarterly*, 1(2):164–182, 2003.

National Center for Education Statistics. Condition of education. U.S. Department of Education, Institute of Education Sciences, 2022.

T. K. Odle and L. C. Russell. The Impact of Reverse Transfer Associate Degrees on Education and Labor Market Outcomes. *Journal of Policy Analysis and Management*, 42(3):648–676, 2023.

P. Oreopoulos and U. Petronijevic. Making College Worth It: A Review of Research on the Returns to Higher Education., 2013.

N. W. Papageorge, S. Gershenson, and K. M. Kang. Teacher Expectations Matter. *Review* *of Economics and Statistics*, 102(2):234–251, 2020.

L. D. Pearce and J. H. Hardie. *Approaches to Mixed Methods Research.* SAGE Publications, Inc., California, 2024.

R. K. Raley, Y. Kim, and K. Daniels. Young Adults' Fertility Expectations and Events: Associations with College Enrollment and Persistence. *Journal of Marriage and Family*, 74(4):866–879, 2012.

L. A. Rhue. Racial Influence on Automated Perceptions of Emotions. Social Science Research Network, 2018.

M. Roderick, V. Coca, and J. Nagaoka. Potholes on the Road to College: High School Effects in Shaping Urban Students' Participation in College Application, Four-Year College Enrollment, and College Match. *Sociology of Education*, 84(3):178–211, 2011.

J. Rust. Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher. *Econometrica*, 55(5):999–1033, 1987.

P. Sharkey. The Acute Effect of Local Homicides on Children's Cognitive Performance. *Proceedings of the National Academy of Sciences*, 107(26):11733–11738, 2010.

M. L. Small and J. M. Calarco. *Qualitative Literacy: A Guide to Evaluating Ethnographic and Interview Research.* University of California Press, California, 2022.

M. Thelwall and T. Nevill. Is Research with Qualitative Data More Prevalent and Impactful Now? Interviews, Case Studies, Focus Groups and Ethnographies. *Library and Information Science Research*, 43(2):1–14, 2021.

S. Timmermans and I. Tavory. Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis. *Sociological Theory*, 30(3):167–186, 2012.

W. van der Klaauw. On the Use of Expectations Data in Estimating Structural Dynamic Choice Models. *Journal of Labor Economics*, 30(3):521–554, 2012.

D. A. Webber. Are College Costs Worth It? How Ability, Major, and Debt Affect the Returns to Schooling. *Economics of Education Review*, 53:296–310, 2016.

M. Wiswall and B. Zafar. Determinants of College Major Choice: Identification Using an Information Experiment. *The Review of Economic Studies*, 82(2):791–824, 2015.

G. T. Wodtke, D. J. Harding, and F. Elwert. Neighborhood Effects in Temporal Perspective: The Impact of Long-Term Exposure to Concentrated Disadvantage on High School Graduation. *American Sociological Review*, 76(5):713–736, 2011.

Appendix A Supplemental: Shocks and Their Effects in the NLSY

Table A1: NLSY97 Summary Stats by Demographics

	(1) All	(2) Black	(3) White	(4) Low Inc	(5) Mid Inc	(6) High Inc
Panel A: Demographics & Background						
White non-Hisp	0.67	0.00	1.00	0.53*	0.76	0.84
Black	0.15	1.00	0.00	0.24*	0.09	0.05
Hispanic	0.13	0.00	0.00	0.18*	0.10	0.05
Low income	0.34	0.54*	0.27	1.00	0.00	0.00
Mid income	0.44	0.27*	0.50	0.00	1.00	0.00
High income	0.09	0.03*	0.12	0.00	0.00	1.00
Mom no diploma	0.17	0.23*	0.11	0.31*	0.10	0.03
Mom HS diploma	0.36	0.42*	0.37	0.39*	0.37	0.22
Mom some college	0.25	0.24*	0.27	0.21†	0.30	0.22
Mom bachelor's	0.21	0.11*	0.25	0.08*	0.23	0.53
GPA 3 and up	0.44	0.28*	0.49	0.33*	0.48	0.61
GPA 2-3	0.48	0.58*	0.44	0.54*	0.46	0.36
GPA 1-2	0.07	0.12*	0.06	0.11*	0.06	0.03
Verbal score	0.03	-0.60*	0.17	-0.23*	0.08	0.30
Math score	0.01	-0.59*	0.16	-0.30*	0.08	0.39
Panel B: Final Attainment						
No degree	0.10	0.12*	0.08	0.16*	0.05	0.03
HS or GED	0.31	0.36*	0.30	0.41*	0.28	0.15
Some coll/assoc.	0.30	0.34*	0.28	0.29†	0.31	0.26
Bachelor or more	0.30	0.19*	0.34	0.15*	0.36	0.57
Panel C: Adverse Shocks						
No father in HH	0.24	0.46*	0.19	0.44*	0.13	0.05
Changed schools	0.11	0.11	0.11	0.16*	0.08	0.08
Break-in	0.10	0.12*	0.09	0.11	0.09	0.09
Bullied	0.11	0.09*	0.12	0.12	0.10	0.09
Seen shooting	0.10	0.21*	0.08	0.15*	0.08	0.06
Parent died	0.03	0.05*	0.03	0.04*	0.02	0.01
Other family died	0.50	0.57*	0.50	0.51	0.50	0.52
Parent hospitalized	0.09	0.10	0.09	0.09	0.09	0.10
Parent jailed	0.02	0.02	0.02	0.03†	0.02	0.01
Parents divorced	0.09	0.07*	0.09	0.10*	0.09	0.04
Parent unemp	0.08	0.09*	0.08	0.09*	0.07	0.05
Victim of crime	0.07	0.07	0.07	0.08*	0.07	0.04
Ever homeless	0.02	0.02	0.02	0.02*	0.01	0.00
Panel D: Expectations						
Coll exp 0-25	0.13	0.13	0.13	0.20*	0.11	0.03
Coll exp 25-50	0.17	0.18	0.16	0.22*	0.15	0.07
Coll exp 50-75	0.13	0.11	0.13	0.14†	0.13	0.10
Coll exp 75-100	0.57	0.58	0.58	0.45*	0.61	0.79
Exp: crime victim	0.14	0.17*	0.13	0.16*	0.14	0.12
Exp: arrest	0.10	0.13*	0.09	0.12*	0.09	0.07
Exp: death	0.18	0.23*	0.17	0.20*	0.17	0.13
Exp: pregnancy	0.07	0.11*	0.06	0.09*	0.06	0.05
Exp: get drunk	0.22	0.12*	0.24	0.22†	0.22	0.24
Observations	8984	2334	4413	3588	3478	671

NLSY97 sample-weighted means. Verbal and math scores have been standardized. "Low Inc" indicates childhood household income of less than \$35,000; "Mid Inc," in the range \$35,000-\$100,000; "High Inc," above \$100,000. Values marked with $*$ and † are statistically different from paired groups' means according to t-tests with p-values of 0.01 and 0.1, respectively; for these tests, Black means are paired with white, and low-income means are paired with high-income.

Table A2: Attainment and Expectations' Relationship to Past Shocks

	(1) Dropout	(2) Dropout	(3) Exp Coll	(4) Exp Coll	(5) Exp Shocks	(6) Exp Shocks
Black	0.11***	0.10***	0.08***	0.09***	0.03***	0.03***
Hispanic	0.10***	0.09***	0.02	0.03*	0.03***	0.03***
Asian or PI	-0.02	-0.02	0.10***	0.10***	-0.01	-0.01
Native American	0.03	0.02	0.03	0.04	-0.05	-0.05
Multiple races	0.02	0.00	-0.04	-0.00	0.09*	0.08*
Male	0.09***	0.09***	-0.09***	-0.09***	0.02***	0.02***
Sibling count	0.01	0.01	-0.00	-0.00	-0.00	-0.00*
Low income	0.06***	0.05**	-0.12***	-0.10***	0.02***	0.02**
Mid income	-0.00	-0.00	-0.05***	-0.04***	0.01	0.01
Mom no diploma	0.23***	0.22***	-0.23***	-0.22***	0.00	-0.00
Mom HS diploma	0.18***	0.18***	-0.16***	-0.15***	0.00	-0.00
Mom some college	0.09***	0.09***	-0.08***	-0.08***	-0.00	-0.01
Family shocks	0.02***		-0.01*		0.01***	
Victimization	0.07***		-0.02***		0.03***	
No father in HH		0.02		-0.03*		0.01
Changed schools		0.13***		-0.05**		0.04***
Parent died		0.02		-0.06*		0.03*
Other family died		-0.00		0.01		0.00
Parent hospitalized		0.00		0.04**		-0.02*
Parent jailed		0.06		-0.08*		0.04**
Parents divorced		0.04		-0.01		-0.00
Parent unemp		0.00		0.01		0.01
Break-in		0.04		0.02		0.02***
Bullied		0.03		-0.00		0.01
Seen shooting		0.10***		-0.02		0.05***
Feels unsafe		0.06***		-0.06***		0.02***
Victim of crime		0.01		-0.02		0.01
Ever homeless		0.13		-0.01		0.01
Constant	0.09***	0.02	0.99***	1.07***	0.07***	0.05***
Adjusted R^2	0.09	0.09	0.14	0.17	0.07	0.08
Outcome mean	0.23	0.23	0.74	0.74	0.13	0.13
Observations	4206	4194	2596	2586	2608	2598

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Regressions of an indicator for leaving college without a bachelor's degree after enrolled (Dropout), the NLSY97 subjective probability of earning a bachelor's degree (Exp Coll), and the average subjective probability of future shocks including pregnancy, arrest, and death (Exp Shocks) on various covariates. In columns 1, 3, and 5 indicators of past shocks are summed to arrive at two summary measures, "family shocks" (such as parental incarceration or unemployment or a death in the family) and "victimization" (such as seeing a shooting or being the victim of a crime). In columns 2, 4, and 6 individual shock indicators are included as covariates. "Outcome means" for each column's outcome variable calculated within that regression's estimation sample. All estimation is weighted using NLSY97 sampling weights.

Table A3: LPM for Bachelor's Degree, Given Enrollment

	(1)	(2)	(3)	(4)	(5)
Black	-0.04*	-0.00	-0.07***	-0.03	-0.03
Hispanic	-0.09***	-0.09***	-0.09***	-0.08***	-0.09***
Asian or PI	0.11	0.09	0.07	0.10	0.05
Native American	-0.12	-0.04	-0.11	-0.15	-0.07
Multiple races	-0.11	-0.06	-0.09	-0.08	-0.05
Male	-0.10***	-0.09***	-0.06***	-0.09***	-0.07***
Sibling count	-0.01	-0.01	-0.00	-0.01	-0.01
Low income	-0.23***	-0.16***	-0.18***	-0.22***	-0.13***
Mid income	-0.08***	-0.06**	-0.06**	-0.08***	-0.04
Mom no diploma	-0.40***	-0.36***	-0.31***	-0.39***	-0.29***
Mom HS diploma	-0.34***	-0.33***	-0.27***	-0.34***	-0.28***
Mom some college	-0.23***	-0.22***	-0.19***	-0.23***	-0.18***
No father in HH		-0.03			-0.02
Changed schools		-0.11***			-0.09***
Parent died		-0.08			-0.07
Other family died		-0.04*			-0.04**
Parent hospitalized		-0.03			-0.05
Parent jailed		-0.13**			-0.10*
Parents divorced		-0.05			-0.04
Parent unemp		0.01			0.01
Break-in		-0.07**			-0.07***
Bullied		-0.04			-0.04
Seen shooting		-0.10***			-0.07***
Feels unsafe		-0.09***			-0.07***
Victim of crime		-0.04			-0.03
Ever homeless		-0.14***			-0.13***
Coll exp 25-50			0.03		0.02
Coll exp 50-75			0.17***		0.15***
Coll exp 75-100			0.30***		0.26***
Exp: crime victim				-0.09*	0.00
Exp: arrest				-0.13**	-0.03
Exp: death				-0.12***	-0.11***
Exp: pregnancy				-0.06	0.04
Exp: get drunk				-0.06*	-0.06*
Constant	0.78***	0.98***	0.48***	0.83***	0.71***
Adjusted R^2	0.19	0.23	0.25	0.20	0.28
Observations	2503	2503	2503	2503	2503

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Linear probability model (LPM) with earning a bachelor's degree as the outcome, conditional on postsecondary enrollment. "Coll exp X-Y" are indicators for the individual's subjective probability of earning a bachelor's degree being between X and Y percent. The other "Exp" controls represent the subjective probabilities placed on the specified adverse future events. Enrolled sample rate of bachelor's degree attainment is 30 percent, or 0.30. All estimation is weighted using NLSY97 sampling weights.

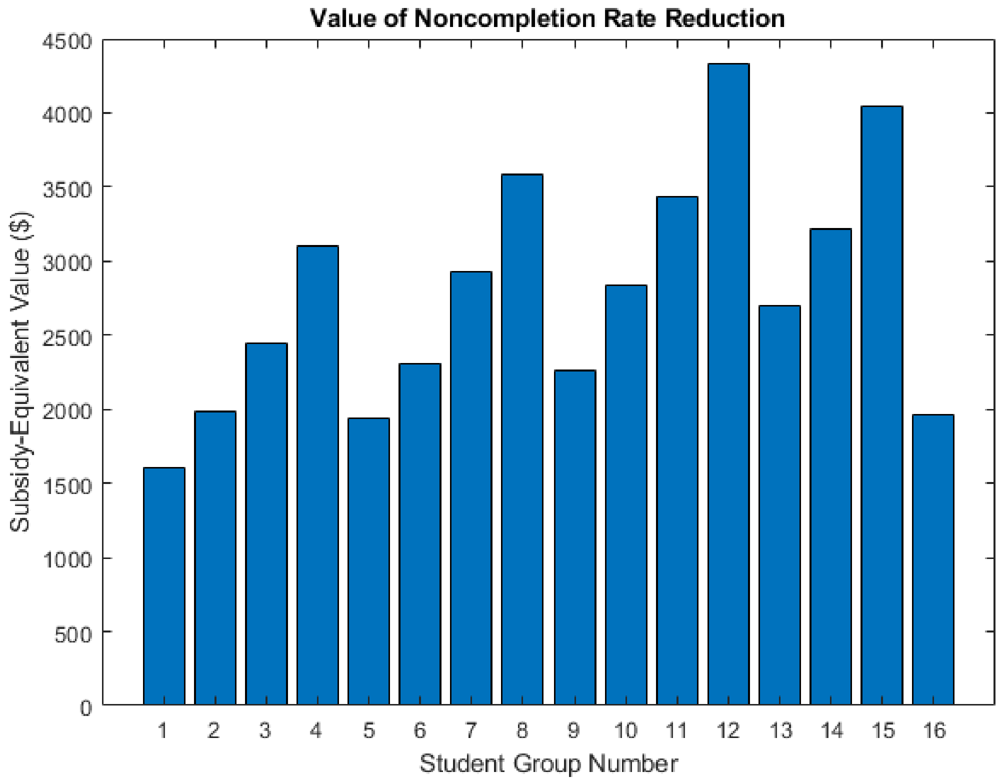
Table A4: LPM for Starting at a Two-Year Program, Given Enrollment

	(1)	(2)	(3)	(4)	(5)
Black	0.05	0.00	0.05	0.02	0.00
Hispanic	0.18***	0.18***	0.16***	0.16***	0.16***
Asian or PI	-0.06	-0.04	-0.03	-0.05	-0.02
Native American	0.26	0.13	0.23	0.27*	0.14
Multiple races	0.12	0.09	0.13	0.07	0.09
Male	0.05*	0.04*	0.02	0.04*	0.03
Sibling count	0.00	0.00	-0.00	0.01	0.00
Low income	0.21***	0.14***	0.17***	0.20***	0.11***
Mid income	0.09***	0.06*	0.07**	0.09***	0.05
Mom no diploma	0.24***	0.21***	0.19***	0.23***	0.16***
Mom HS diploma	0.22***	0.22***	0.18***	0.22***	0.18***
Mom some college	0.16***	0.15***	0.13***	0.16***	0.12***
No father in HH		0.06*			0.05
Changed schools		0.13***			0.10**
Parent died		0.13*			0.12
Other family died		-0.00			-0.00
Parent hospitalized		0.01			0.04
Parent jailed		0.14			0.07
Parents divorced		0.06			0.06
Parent unemp		0.04			0.03
Break-in		0.05			0.04
Bullied		0.12***			0.12***
Seen shooting		0.08**			0.06
Feels unsafe		0.11***			0.09***
Victim of crime		0.02			0.02
Ever homeless		0.06			0.05
Coll exp 25-50			-0.02		0.00
Coll exp 50-75			-0.19***		-0.15***
Coll exp 75-100			-0.33***		-0.25***
Exp: crime victim				0.06	-0.06
Exp: arrest				0.10	0.00
Exp: death				0.18***	0.20***
Exp: pregnancy				0.24***	0.12
Exp: get drunk				0.03	0.04
Constant	0.17***	-0.03	0.50***	0.12***	0.22***
Adjusted R^2	0.10	0.15	0.15	0.12	0.19
Observations	1784	1784	1784	1784	1784

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Linear probability model (LPM) with choice of a two-year postsecondary program as the outcome. "Coll exp X-Y" are indicators for the individual's subjective probability of earning a bachelor's degree being between X and Y percent. The other "Exp" controls represent the subjective probabilities placed on the specified adverse future events. Enrolled sample rate of two-year program enrollment is 51 percent, or 0.51. All estimation is weighted using NLSY97 sampling weights.

Figure A1



Value of 10-percentage-point reduction in noncompletion rates $\alpha^s(X_i)$ as expressed in the dollar value of annual student subsidies yielding the same lifetime utility increase. Group bars represent subsamples of respondents who had certain characteristics or not as indicated by Table E2.

Appendix B Supplemental: Evidence of Shocks' Effects in the ELS

Do students actually have to worry about adverse shock events derailing their educational paths? This is difficult to verify in the NLSY97 data, though, as we have seen above, there is certainly an association between adverse shocks in general and lower attainment. However, in this instance, the ELS, for which summary statistics are supplied in Table B1, can provide more conclusive results. Respondents to that survey supplied information regarding events that occurred in their lives in the period just *after* their expected date of graduation from high school, when those who attended postsecondary schools would have been enrolled. This information is not available in the NLSY97. Relevant shock events include whether their parents divorced, became unemployed, or died, whether another loved one died, whether the student or an immediate family member fell seriously ill, or whether the student was the victim of violence. We sum these indicators (labeled "shocks in college") and treat this sum as representative of the level of instability in a college student's life. This sum is visibly correlated with final attainment in Table B1. Now we can more formally analyze the effect of such instability among only those students who actually enroll in some kind of postsecondary program and look for evidence of adverse-shock-related derailment.

In Table B2, we report results from estimation in the ELS of an ordered probit model of final educational attainment, with the possible outcomes being (in the following order) no credential, a certificate, an associate's degree, and a bachelor's degree or more. Adverse shock events have a significant negative relationship with final attainment given enrollment, even in the presence of controls for demographic and socioeconomic background, as well as high school academic performance and standardized test scores. Similarly, significant relationships can be found in a linear probability model of bachelor's degree attainment. It would appear, then, that adverse shocks—typically subsumed into an error term and thus treated as unobserved heterogeneity—likely help to derail educational paths. Any young person who is concerned about life-altering negative events happening in their future might further reason that such events would make it difficult to complete a bachelor's or other degree.

Table B1: ELS Summary Stats by Attainment

	All	No Deg	HS	Some Coll	Bachelor
White	0.54	0.31	0.52	0.51	0.64
Black	0.12	0.19	0.12	0.15	0.07
Hispanic	0.14	0.28	0.17	0.16	0.08
Low income	0.28	0.49	0.42	0.31	0.15
Mid income	0.53	0.38	0.46	0.54	0.56
High income	0.15	0.02	0.04	0.10	0.27
Mother: No degree	0.13	0.38	0.23	0.13	0.05
Mother: HS diploma	0.27	0.33	0.40	0.29	0.18
Mother: Some college	0.33	0.20	0.28	0.37	0.31
Mother: Bachelor	0.27	0.09	0.09	0.21	0.45
GPA: 0-1	0.02	0.17	0.04	0.02	0.00
GPA: 1-2	0.17	0.59	0.36	0.20	0.02
GPA: 2-3	0.41	0.22	0.47	0.51	0.25
GPA: 3-4	0.40	0.02	0.13	0.28	0.73
Math score	50.71	40.73	44.36	49.04	56.87
Reading score	50.53	40.85	44.66	49.17	56.39
Composite score	50.66	40.16	44.13	49.04	57.08
No degree	0.03	1.00	0.00	0.00	0.00
HS diploma	0.10	0.00	1.00	0.00	0.00
Some college	0.48	0.00	0.00	1.00	0.00
Bachelor	0.38	0.00	0.00	0.00	1.00
Shocks in college	0.93	1.05	1.04	1.00	0.80
Observations	16197	356	1388	6406	5100

ELS means for group indicator variables, standardized test scores, and a count of adverse shock events during college. Test scores have themselves been statistically standardized. The "Some Coll" column includes all individuals who attended some kind of postsecondary institution but never earned an associate's or bachelor's degree.

Table B2: ELS: Ordered Probit for Attainment Given PS Enrollment

	(1)	(2)	(3)	(4)
Black	-0.53***	-0.51***	-0.41***	0.04
Hispanic	-0.49***	-0.48***	-0.32***	-0.07*
Asian or PI	0.08**	0.07*	0.15***	0.08*
Native American	-0.66***	-0.63***	-0.49***	-0.15
Multiple races	-0.26***	-0.24***	-0.22***	-0.12**
Male	-0.12***	-0.13***	-0.17***	-0.09***
Shocks in college		**-0.10***	**-0.09***	**-0.06***
Low income			-0.42***	-0.29***
Mid income			-0.22***	-0.17***
Mother: no degree			-0.62***	-0.34***
Mother: HS diploma			-0.49***	-0.29***
Mother: some college			-0.38***	-0.24***
GPA: 3-4				1.25***
GPA: 2-3				0.56***
GPA: 1-2				0.13
Reading score				0.01***
Math score				0.02***
HS Cutoff	-0.61***	-0.71***	-1.21***	1.19***
Associate Cutoff	-0.33***	-0.42***	-0.92***	1.52***
Bachelor Cutoff	-0.09***	-0.18***	-0.67***	1.81***
Observations	9239	9239	9239	9239

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Ordered probit for final attainment in ELS data. "Shocks in college" refers to the count of adverse events that happen to respondents while they are in postsecondary school, including parental divorce, unemployment, and death, the death of other loved ones, serious illness befalling the student or a family member, and violent victimization. A significant negative effect suggests that college-period shocks do reduce eventual attainment, even in the presence of increasing sets of covariates.

Appendix C Supplemental: Expectations and Institution Choice in HSLS

Summary statistics for the High School Longitudinal Study (HSLS) sample on the unique variables are summarized in Table C1, while essential demographic characteristics are shown in Table C2. The former include students' own beliefs regarding their ability to earn a bachelor's degree as well as whether they expect to do so, whether they anticipate qualifying for financial aid after high school, and even whether and why they would eventually take a break from postsecondary school. When we estimate the relationship between these variables and what kind of institution students eventually graduate from, these other new kinds of expectations do not eliminate a statistically significant estimated link between attainment expectations and outcomes. Results are provided in Table C3. Students who expect to earn a bachelor's are more likely to end up at four-year programs of the public and non-profit varieties, and less likely to get a credential from a two-year school. These results demonstrate that students' predictions regarding their own academic ability and financial aid eligibility do not drive the relationship between attainment expectations and institution choice we found in the NLSY97, enhancing the case for a direct causal link.

Table C1: HSLS Means for Expectations and Shocks Variables

Exp Bachelor's	0.62
Thinks capable of BA	0.49
Fin aid: will qual	0.43
Fin aid: won't qual	0.26
Fin aid: unsure	0.31
Plans 4yr enroll, 2009	0.53
Break: Academic	0.00
Break: Family	0.03
Break: Financial	0.02
Break: Work	0.02
Break: Unknown	0.01
Observations	23495

HSLS means for postsecondary school type indicator variables, adverse shocks, and expectations. "Shocks" variables are sums of various types of adverse shocks over the relevant time period (2009-2011 or 2012-2016). Expected earnings are in thousands of dollars per year.

Table C2: HSLS Means

	All	No Deg	HS	Some Coll	College
White	0.51	0.44	0.50	0.51	0.53
Black	0.10	0.12	0.13	0.12	0.03
Hispanic	0.16	0.23	0.19	0.17	0.14
Asian, HI, PI	0.09	0.04	0.05	0.06	0.14
Male	0.51	0.56	0.57	0.49	0.27
Low income	0.20	0.34	0.30	0.22	0.08
Mid income	0.30	0.18	0.29	0.36	0.32
High income	0.21	0.05	0.08	0.16	0.31
Mother no degree	0.09	0.25	0.16	0.09	0.00
Mother HS only	0.46	0.54	0.59	0.50	0.39
Mother Assoc.	0.18	0.13	0.15	0.22	0.27
Mother Bachelor	0.27	0.08	0.10	0.20	0.33
Math score	0.03	-0.76	-0.42	-0.04	0.48
Observations	23503	747	3670	3935	59

HSLS means for group indicator variables and standardized test scores. Test scores have been statistically standardized. The "Some Coll" column includes all individuals who attended some kind of postsecondary institution but never earned an associate's or bachelor's degree.

Table C3: MNL for First Postsecondary Program Type Given Plan to Enroll in Four-Year: HSLS

	Pub2	FP2	Pub4	NP4	FP4
Exp Bachelor's	-0.09***	-0.01**	0.06*	0.05*	0.01
Break: Academic	0.68	-0.19	2.07	-1.86	0.04
Break: Family	0.08**	0.02***	0.00	-0.12	-0.00
Break: Financial	0.10**	0.01	-0.13	-0.03	0.02***
Break: Work	0.03	0.01	0.04	-0.13	0.01
Break: Unknown	0.07	0.00	-0.12	0.07	-0.10
Thinks capable of BA	-0.02	-0.00	0.02	0.01	-0.00
Fin aid: will qual	0.00	0.00	0.00	0.00	0.00
Fin aid: won't qual	-0.01	-0.00	0.05**	-0.04**	0.00
Fin aid: unsure	-0.02	-0.00	0.01	-0.01	0.00
Black	-0.09***	-0.00	0.04	0.07***	0.01*
Hispanic	0.00	0.00	-0.03	0.02	0.00
Asian, HI, PI	-0.10***	0.01	0.11***	-0.03	0.00
Native American	-0.00	-0.14	0.15	0.07	-0.09
Multiple races	0.01	0.01	-0.02	-0.02	0.01**
Male	0.01	-0.02***	0.05***	-0.02*	-0.00
Low income	0.09***	0.01	-0.05*	-0.05**	0.00
Mid income	0.07***	0.02**	-0.07***	-0.03*	0.00
Mother no degree	0.12***	0.01	-0.00	-0.11**	-0.00
Mother HS only	0.05***	0.01*	-0.03	-0.03**	-0.00
Mother Associate's	0.07***	-0.00	-0.04*	-0.03*	-0.01
GPA: Academic	-0.09***	-0.01***	0.08***	0.06***	-0.01*
GPA: CTE	-0.02*	-0.00	0.04**	-0.01	-0.00
Math score	-0.05***	-0.01**	0.04***	0.03***	0.00
Observations	3768	3768	3768	3768	3768

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. "Break" variables indicate that the respondent took a break from school for the stated reason. "Fin aid" variables regard respondents' expectations for educational financial aid qualification. "CTE" indicates Career and Technical Education; "HI" and "PI" indicate Hawaiian and Pacific Islander, respectively. Column titles are "Public Two-Year Institution," "For-Profit Two-Year Institution," "Public Four-Year Institution," "Non-Profit Four-Year Institution" and "For-Profit Four-Year Institution," respectively.

Appendix D Supplemental: Analyses with Demographic Subsamples

In this appendix, we perform a few of the same analyses as in Tables A2 and A3, as well as a tweaked version of the analysis in Table A4, using two relevant demographic subsamples: Black individuals and those from low-income backgrounds. These subsamples more closely resemble the narrative interview sample explored in Section 2. The change in the program-type choice analysis is that the outcome is choosing to enroll in a four-year public or non-profit school, and the other available option is *any* other enrollment choice, including not enrolling in postsecondary school at all. Patterns are broadly similar to those in the main text: a history of adverse shocks predicts lower attainment and more dire expectations about the future, and those worse expectations are correlated with lower final attainment, potentially through the mechanism of program choice.

Table D1: Black Subsample: Attainment, Expectations, and Past Shocks

	(1) Dropout	(2) Exp Coll	(3) Exp Shocks
Male	0.14***	-0.01	0.06***
Sibling count	0.04***	-0.02*	-0.00
Low income	0.05	-0.11***	0.00
Mid income	-0.06	-0.07*	-0.01
Mom no diploma	0.14**	-0.21***	-0.03
Mom HS diploma	0.21***	-0.10***	-0.03
Mom some college	0.10*	-0.09**	-0.04*
Family shocks	0.02	0.01	0.01
Victimization	0.06***	-0.01	0.03***
Constant	0.17***	0.99***	0.13***
Adjusted R^2	0.08	0.07	0.06
Observations	945	658	660

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Regressions of an indicator for leaving college without a bachelor's degree (Dropout), subjective probability of earning a bachelor's degree (Exp Coll), and average subjective probability of future shocks including pregnancy, arrest, and death (Exp Shocks) on various covariates. Shock indicators are summed to arrive at two summary measures, "family shocks" (such as parental incarceration or unemployment or death in the family) and "victimization" (such as seeing a shooting or being the victim of a crime). All estimation is weighted using NLSY97 sampling weights.

Table D2: Low-Income Subsample: Attainment, Expectations, and Past Shocks

	(1) Dropout	(2) Exp Coll	(3) Exp Shocks
Black	0.12***	0.12***	0.02
Hispanic	0.08**	0.08***	0.03*
Asian or PI	-0.03	0.32***	-0.02
Native American	-0.02	-0.07*	0.01
Multiple races	0.05	-0.26***	0.26***
Male	0.09***	-0.10***	0.05***
Sibling count	0.02	-0.01	-0.00
Mom no diploma	0.24***	-0.33***	0.01
Mom HS diploma	0.21***	-0.24***	0.02
Mom some college	0.04	-0.12***	-0.00
Family shocks	0.03**	-0.01	0.02***
Victimization	0.05***	-0.02*	0.03***
Constant	0.13**	0.92***	0.06***
Adjusted R^2	0.08	0.13	0.10
Observations	1202	1036	1038

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Regressions of an indicator for leaving college without a bachelor's degree (Dropout), subjective probability of earning a bachelor's degree (Exp Coll), and average subjective probability of future shocks including pregnancy, arrest, and death (Exp Shocks) on various covariates. Shock indicators are summed to arrive at two summary measures, "family shocks" (such as parental incarceration or unemployment or death in the family) and "victimization" (such as seeing a shooting or being the victim of a crime). In columns 2, 4, and 6 individual shock indicators are included as covariates. All estimation is weighted using NLSY97 sampling weights.

Table D3: Black Subsample: LPM for Bachelor's Degree, Given Enrollment

	(1)	(2)	(3)	(4)	(5)
Male	-0.11***	-0.10***	-0.10***	-0.08*	-0.07*
Sibling count	-0.04***	-0.04***	-0.03***	-0.04***	-0.03***
Low income	-0.14**	-0.11*	-0.11*	-0.13**	-0.08
Mid income	-0.03	-0.01	-0.01	-0.04	0.00
Mom no diploma	-0.34***	-0.32***	-0.28***	-0.33***	-0.27***
Mom HS diploma	-0.28***	-0.27***	-0.25***	-0.29***	-0.24***
Mom some college	-0.17**	-0.17**	-0.14*	-0.19**	-0.15**
Family shocks		-0.02			-0.02
Victimization		-0.05***			-0.05***
Coll exp 25-50			0.04		0.07
Coll exp 50-75			0.04		0.04
Coll exp 75-100			0.21***		0.20***
Exp: crime victim				-0.00	0.05
Exp: arrest				-0.18*	-0.11
Exp: death				0.06	0.05
Exp: pregnancy				-0.16**	-0.11
Exp: get drunk				-0.11	-0.08
Constant	0.67***	0.71***	0.46***	0.70***	0.52***
Adjusted R^2	0.14	0.15	0.18	0.15	0.19
Observations	617	617	617	617	617

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Linear probability model with earning a bachelor's degree as the outcome. "Coll exp X-Y" are indicators for the individual's subjective probability of earning a bachelor's degree being between X and Y percent. The other "Exp" controls represent the subjective probabilities placed on the specified adverse future events. All estimation is weighted using NLSY97 sampling weights.

Table D4: Low-Income Subsample: LPM for Bachelor's, Given Enrollment

	(1)	(2)	(3)	(4)	(5)
Black	0.01	0.03	-0.02	0.02	-0.00
Hispanic	0.01	0.01	0.00	0.02	0.00
Asian or PI	0.25	0.22	0.17	0.24	0.14
Native American	0.06	0.03	0.12	0.06	0.09
Multiple races	-0.18***	-0.13**	-0.10***	-0.08	-0.01
Male	-0.07**	-0.06**	-0.04	-0.04	-0.02
Sibling count	-0.00	-0.01	-0.00	-0.00	-0.00
Mom no diploma	-0.26***	-0.25***	-0.18***	-0.25***	-0.17***
Mom HS diploma	-0.22***	-0.21***	-0.15**	-0.21***	-0.15**
Mom some college	-0.07	-0.05	-0.03	-0.06	-0.02
Family shocks		-0.03**			-0.03**
Victimization		-0.05***			-0.05***
Coll exp 25-50			0.02		0.01
Coll exp 50-75			0.09**		0.07*
Coll exp 75-100			0.19***		0.17***
Exp: crime victim				-0.08	-0.04
Exp: arrest				-0.15***	-0.07
Exp: death				-0.01	-0.02
Exp: pregnancy				-0.09**	-0.04
Exp: get drunk				-0.02	-0.03
Constant	0.37***	0.45***	0.20***	0.40***	0.30***
Adjusted R^2	0.07	0.09	0.11	0.08	0.14
Observations	982	982	982	982	982

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Linear probability model with earning a bachelor's degree as the outcome. "Coll exp X-Y" are indicators for the individual's subjective probability of earning a bachelor's degree being between X and Y percent. The other "Exp" controls represent the subjective probabilities placed on the specified adverse future events. All estimation is weighted using NLSY97 sampling weights.

Table D5: LPM for Starting at a Four-Year (Not For-Profit) Program, Given Enrollment

	(1)	(2)	(3)	(4)	(5)
Black	-0.03	0.01	-0.06**	-0.01	-0.01
Hispanic	-0.11***	-0.11***	-0.10***	-0.09***	-0.11***
Asian or PI	0.10	0.08	0.06	0.10	0.04
Native American	-0.26***	-0.17*	-0.26**	-0.28***	-0.19**
Multiple races	-0.15	-0.08	-0.12	-0.10	-0.07
Male	-0.08***	-0.08***	-0.04**	-0.07***	-0.04**
Sibling count	-0.01	-0.01	-0.00	-0.01	-0.01
Low income	-0.26***	-0.18***	-0.19***	-0.24***	-0.13***
Mid income	-0.11***	-0.08***	-0.08***	-0.10***	-0.06**
Mom no diploma	-0.36***	-0.33***	-0.26***	-0.35***	-0.24***
Mom HS diploma	-0.30***	-0.29***	-0.23***	-0.30***	-0.23***
Mom some college	-0.20***	-0.19***	-0.16***	-0.20***	-0.15***
No father in HH		-0.07***			-0.05**
Changed schools		-0.09***			-0.07***
Parent died		-0.08			-0.07
Other family died		0.00			-0.00
Parent hospitalized		-0.02			-0.05
Parent jailed		-0.11*			-0.06
Parents divorced		-0.05			-0.05
Parent unemp		0.00			0.00
Break-in		-0.04			-0.04
Bullied		-0.08***			-0.08***
Seen shooting		-0.07**			-0.04
Feels unsafe		-0.11***			-0.08***
Victim of crime		-0.02			-0.01
Ever homeless		-0.11**			-0.10*
Coll exp 25-50			0.02		0.01
Coll exp 50-75			0.18***		0.15***
Coll exp 75-100			0.34***		0.28***
Exp: crime victim				-0.00	0.09*
Exp: arrest				-0.14**	-0.04
Exp: death				-0.15***	-0.14***
Exp: pregnancy				-0.21***	-0.10*
Exp: get drunk				-0.03	-0.04
Constant	0.80***	1.00***	0.46***	0.84***	0.70***
Adjusted R^2	0.17	0.21	0.24	0.19	0.28
Observations	2511	2511	2511	2511	2511

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Linear probability model with choice of a four-year public or private non-profit school, given some postsecondary enrollment, as the outcome. "Coll exp X-Y" are indicators for the individual's subjective probability of earning a bachelor's degree being between X and Y percent. The other "Exp" controls represent the subjective probabilities placed on the specified adverse future events. All estimation is weighted using NLSY97 sampling weights.

Table D6: Black Subsample: LPM for Starting at a Four-Year (Not For-Profit) Program, Given Enrollment

	(1)	(2)	(3)	(4)	(5)
Male	-0.07*	-0.06*	-0.06*	-0.05	-0.05
Sibling count	-0.02	-0.02	-0.01	-0.02	-0.01
Low income	-0.10	-0.09	-0.07	-0.09	-0.07
Mid income	0.05	0.07	0.08	0.04	0.07
Mom no diploma	-0.38***	-0.39***	-0.32***	-0.38***	-0.34***
Mom HS diploma	-0.30***	-0.30***	-0.26***	-0.30***	-0.27***
Mom some college	-0.20**	-0.21**	-0.17**	-0.22***	-0.18**
Family shocks		0.01			0.01
Victimization		-0.03*			-0.02
Coll exp 25-50			0.00		0.01
Coll exp 50-75			0.03		0.03
Coll exp 75-100			0.21***		0.19***
Exp: crime victim				0.03	0.03
Exp: arrest				-0.13	-0.05
Exp: death				-0.07	-0.06
Exp: pregnancy				-0.12	-0.08
Exp: get drunk				-0.14*	-0.12*
Constant	0.64***	0.64***	0.43***	0.69***	0.49***
Adjusted R^2	0.13	0.13	0.17	0.14	0.17
Observations	620	620	620	620	620

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Linear probability model with choice of a four-year public or private non-profit school, given some postsecondary enrollment, as the outcome. "Coll exp X-Y" are indicators for the individual's subjective probability of earning a bachelor's degree being between X and Y percent. The other "Exp" controls represent the subjective probabilities placed on the specified adverse future events. All estimation is weighted using NLSY97 sampling weights.

Table D7: Low-Income Subsample: LPM for Starting at a Four-Year (Not For-Profit) Program, Given Enrollment

	(1)	(2)	(3)	(4)	(5)
Black	0.00	0.02	-0.03	0.01	-0.02
Hispanic	-0.01	-0.02	-0.02	-0.00	-0.02
Asian or PI	0.09	0.06	-0.02	0.08	-0.04
Native American	-0.25***	-0.28***	-0.18***	-0.25***	-0.21***
Multiple races	-0.20***	-0.15***	-0.10***	-0.08	0.00
Male	-0.06**	-0.06**	-0.03	-0.04	-0.01
Sibling count	0.00	-0.00	0.00	0.00	0.00
Mom no diploma	-0.33***	-0.32***	-0.23***	-0.32***	-0.22***
Mom HS diploma	-0.28***	-0.27***	-0.19***	-0.27***	-0.19***
Mom some college	-0.19***	-0.18**	-0.15**	-0.19***	-0.14**
Family shocks		-0.04***			-0.03**
Victimization		-0.05***			-0.04***
Coll exp 25-50			-0.00		-0.01
Coll exp 50-75			0.12***		0.11**
Coll exp 75-100			0.24***		0.22***
Exp: crime victim				0.01	0.06
Exp: arrest				-0.21***	-0.11*
Exp: death				-0.04	-0.05
Exp: pregnancy				-0.11*	-0.04
Exp: get drunk				-0.02	-0.04
Constant	0.47***	0.55***	0.26***	0.50***	0.37***
Adjusted R^2	0.06	0.08	0.12	0.07	0.15
Observations	986	986	986	986	986

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Linear probability model with choice of a four-year public or private non-profit school, given some postsecondary enrollment, as the outcome. "Coll exp X-Y" are indicators for the individual's subjective probability of earning a bachelor's degree being between X and Y percent. The other "Exp" controls represent the subjective probabilities placed on the specified adverse future events. All estimation is weighted using NLSY97 sampling weights.

Appendix E Supplemental: Structural Model Features and Results

In this appendix, we report on methods used to derive noncompletion probabilities $\alpha^s(X_i)$ from the NLSY97 data, though these are superseded by those used in the main specification based solely on on-time completion rates. We also write the likelihood function, describe how we compute standard errors for utility parameters, and provide various parameter and outcome values we use or estimate.

Appendix E.1 Construction of Candidate Noncompletion Probabilities

Let n be the number of years a student enrolls to earn a degree, m be the number of years an "on-time" student would need to complete to earn that degree, and $x = n - m$ (that is, the deterministic number of years students must not have "completed" while enrolled). If π is the proportion of enrollees who earned their degree after n years of enrollment, we can calculate an implied rate of continuous-enrollment noncompletion p among these students using the formula

$$\pi = \frac{(n-1)!}{(n-1-x)!x!} \times p^x(1-p)^m. \tag{7}$$

For example, among Group 1 students, the proportion of bachelor's program enrollees in the NLSY97 who take exactly 6 years to get their degree is 7.2 percent. In this case, $n = 6$, $m = 4$, $x = 2$, and $\pi = .072$. This generates the expression:

$$.072 = \frac{5!}{3!2!} \times p^2(1-p)^4. \tag{8}$$

Essentially, a noncompletion must have happened twice in the first five years of enrollment, since the sixth results in graduation and therefore must have been completed. The first term in the formula determines the number of permutations by which this could happen (here, ten). Then, the latter terms express the probability of two noncompletions and four successfully completed years occurring. Solving for p, this formula has multiple solutions, but the only feasible noncompletion probability[41] among them in this specific case is $p = 0.106$.

These implied noncompletion probabilities differ for different values of n, since in the data the probability is not actually constant. Moreover, to calculate $\alpha^s(X_i)$ values they must be combined with dropout rates, since both types of incidents represent educational disruptions or derailment. However, as noted in the main text, the on-time graduation rate suffices to capture all this information. We calculated values of p using (π, n, m, x) values for multiple years and groups, took multiple approaches to finding "central" or average values for each group, and combined these with group-specific dropout probabilities to get values

[41] Other solutions are infeasibly large values.

of $\alpha^s(X_i)$, but always found results quite similar to the derailment probabilities implied by on-time graduation rates.

Appendix E.2 Likelihood Function

Given our assumption that idiosyncratic ε_{itd} are drawn from a Type 1 Extreme Value distribution, we can derive a closed-form expression for agent i's probability of choosing d^* at age t for a candidate vector of utility parameters, denoted Θ:

$$P(d_{it} = d^*|Z_{it};\Theta) = \frac{exp\left[\bar{v_{d^*}}(Z_{it})\right]}{\sum_{d=0}^{2} exp\left[\bar{v_d}(Z_{it})\right]}.$$

This probability is calculated for each annual enrollment choice made by each agent in the data, and is the contribution of agent i at age t to the likelihood function, given the current parameter guesses. If there are N total agents, and each true decision outcome is expressed $\hat{d_{it}}$, the log-likelihood function can be expressed as

$$\mathcal{L} = \sum_{i=1}^{N}\sum_{t=1}^{T}\left[log\left(P(d_{it} = \hat{d_{it}}|Z_{it};\Theta)\right)\right].$$

Appendix E.3 Computation of Standard Errors

We compute standard errors in the estimation of structural utility parameters by constructing the Hessian of the likelihood function using the outer product measure. We calculate a numerical derivative of the likelihood for each estimated parameter (u_1 and u_2) by perturbing each parameter, solving for choice probabilities using the perturbed value, and computing a new likelihood.

Appendix E.4 Tables

The rest of this appendix contains tables and figures reporting the expected group income levels used in estimation, dollar-value translations of the utility parameters estimated by the model, the results of the subsidy-α equivalence exercise, and the calculated enrollment rates underlying Figure 1.

Table E1: Income by Demographic Group and Education in Structural Model Sample

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Panel A: Group Definition																
URM	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Low-income	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Shocks	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes
Low GPA	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Incomes in $1000's by Educational Outcome																
Assoc. Dropout	36.2	32.5	30.1	26.5	33.0	29.4	27.0	23.4	31.6	28.0	25.6	22.0	28.5	24.9	22.4	18.8
Bach. Dropout	40.6	37.0	34.6	31.0	37.5	33.9	31.5	27.8	36.1	32.5	30.1	26.5	32.9	29.3	26.9	23.3
Assoc. Degree	41.0	37.4	35.0	31.4	37.9	34.3	31.9	28.2	36.5	32.9	30.5	26.8	33.3	29.7	27.3	23.7
Bach. Degree	53.2	49.6	47.2	43.6	50.1	46.5	44.1	40.5	48.7	45.1	42.7	39.1	45.5	41.9	39.5	35.9
No Postsec	32.0	28.4	26.0	22.4	28.9	25.3	22.9	19.2	27.5	23.9	21.5	17.8	24.3	20.7	18.3	14.7
Observations	940	264	171	189	382	181	152	206	726	507	239	430	399	360	285	573

Expected incomes for subsamples used in structural estimation. Columns represent subsamples of respondents who had certain characteristics or not as indicated by Panel A: individuals from URM (underrepresented minority) groups; those from low-income households (total childhood household income less than $35,000); those who experienced at least two of our recorded adverse shocks in childhood; and those whose high school GPA was below 3.0. Panel B displays estimated typical incomes in thousands of dollars for those in each group who have the stated educational outcome. Income values were arrived at by the method described in Section 4.2. Observation counts are of individuals; the sample yields an average of 11.9 yearly observations per individual between 1997 and 2011 for the appropriate age range.

Table E2: Dollar Value of Education Utility

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
URM	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Low-income	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes	No	No	Yes	Yes
Shocks	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes
Low GPA	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel A: Rational consideration of noncompletion probability																
Two-year	-98	-45	-1108	-599	-150	-111	-798	-1061	-93	-71	-567	-461	-166	-189	-1537	-1158
Four-year	-17	-59	-261	-729	-143	-234	-1238	-2081	-236	-268	-603	-2035	-567	-400	-3201	-2772
Panel B: No consideration of noncompletion probability																
Two-year	-365	-2068	-7175	-11760	-1856	-4188	-10636	-20781	-3602	-5828	-14293	-20678	-6182	-11725	-29222	-42687
Four-year	-156	-1177	-2675	-6848	-934	-2477	-5715	-11897	-2311	-4472	-9001	-16112	-4356	-7526	-17822	-26004

Estimated annual dollar value of utility in each program type – two-year (associate's) or four-year (bachelor's) – for various demographic groups, rounded to the dollar. In Panel A, individuals rationally consider probabilities of noncompletion; in Panel B, we assume they ignore these probabilities. Columns represent subsamples of respondents who had certain characteristics or not as indicated by the top four rows: individuals from URM (underrepresented minority) groups; those from low-income households (total less than $35,000); those who experienced at least two of our recorded adverse shocks in childhood; and those whose high school GPA was below 3.0.

Table E3: Enrollment Probability at Age 18, Subsidized vs. Unsubsidized

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Panel A: Rational consideration of noncompletion probability																
Two-year	0.06	0.09	0.05	0.07	0.07	0.09	0.07	0.07	0.10	0.11	0.08	0.10	0.10	0.09	0.06	0.08
Two-year, Subsidy	0.02	0.03	0.01	0.02	0.02	0.03	0.02	0.02	0.03	0.04	0.03	0.04	0.03	0.03	0.03	0.03
Four-year	0.26	0.19	0.16	0.09	0.19	0.17	0.10	0.08	0.12	0.11	0.06	0.06	0.10	0.11	0.04	0.07
Four-year, Subsidy	0.43	0.43	0.41	0.39	0.39	0.41	0.35	0.36	0.38	0.41	0.38	0.37	0.38	0.44	0.31	0.40
Panel B: No consideration of noncompletion probability																
Two-year	0.05	0.06	0.04	0.05	0.06	0.06	0.05	0.05	0.07	0.08	0.06	0.07	0.07	0.07	0.04	0.06
Two-year, Subsidy	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Four-year	0.20	0.12	0.11	0.06	0.14	0.10	0.07	0.05	0.08	0.06	0.04	0.03	0.06	0.07	0.03	0.04
Four-year, Subsidy	0.34	0.28	0.26	0.22	0.29	0.27	0.23	0.22	0.25	0.24	0.21	0.20	0.24	0.23	0.18	0.20

Predicted choice probabilities for enrollment in two-year and four-year programs at age 18 by demographic group, based on model parameter estimates. In Panel A, individuals rationally consider probabilities of noncompletion; in Panel B, we assume they ignore these probabilities. "Subsidy" rows consider a $10,000 payment to students who enroll in a four-year program. Columns represent subsamples of respondents who had certain characteristics or not as indicated by the top three rows: individuals from URM (underrepresented minority) groups; those from low-income households (total less than $35,000); those who experienced at least two of our recorded adverse shocks in childhood; and those whose high school GPA was below 3.0.

Table E4: Present-Value Subsidy Equivalent of α-Reduction

Group	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Subsidy value	1,610	1,980	2,440	3,100	1,940	2,310	2,930	3,580	2,260	2,840	3,440	4,330	2,700	3,220	4,040	1,960

Annual student subsidy amount, in dollars, yielding equivalent present-value utility increases to a 10-percentage-point reduction in noncompletion rates $\alpha^s(X_i)$. Columns represent subsamples of respondents who had certain characteristics or not as indicated by Table E2.