Estimating the Perceived Returns to College*

Clint Harris[†]

March 8, 2021

Click Here for Latest Version

Abstract

The primary determinant of an individual's college attendance is their perceived lifetime return to college. I infer agents' perceived returns by assuming a dollar-for-dollar relationship between perceived returns and tuition costs in a binary choice model of college attendance. This method has the advantage of estimating perceived returns in terms of compensating variation without assuming rational expectations on actual returns or agent knowledge of tuition costs. Estimating the model using both maximum likelihood and moment inequalities, I find that the scale of the distribution of perceived returns is an order of magnitude lower than past work has found when assuming rational expectations on income returns. The low variance I find in perceived returns implies high responses to financial aid. I predict a 2.6 percentage point increase in college attendance from a \$1,000 universal annual tuition subsidy, which is consistent with quasi-experimental estimates of the effects of tuition assistance on college attendance. Because I estimate the complete distribution of perceived returns, my results can be used to predict heterogeneous effects of counterfactual financial aid policies.

JEL Codes: C31, D84, I21

Keywords: College Attendance, Tuition Subsidies, Returns to Education, Biased Beliefs

*I thank Tim Bond, Trevor Gallen, Soojin Kim, Kevin Mumford, Anita Alves Pena, Victoria Prowse, Miguel Sarzosa, Jeff Smith, Chris Taber, Matt Wiswall, and Owen Zidar as well as seminar participants at Case Western Reserve University, the European Association of Labor Economists Annual Meeting, Kansas State University, the Midwest Economics Association Annual Meeting, the National Tax Association Annual Meeting, Purdue University, the Southern Economic Association Annual Meeting, and the US Census Bureau for helpful comments. This research was conducted with restricted access to Bureau of Labor Statistics (BLS) data. The views expressed here do not necessarily reflect the views of the BLS.

†Wisconsin Institute for Discovery, University of Wisconsin-Madison, 330 N Orchard Street, Madison, WI 53715, USA; email: clint.harris@wisc.edu

1 Introduction

The existence of government financial aid for college suggests a concern that some individuals may make suboptimal choices about attending college. This concern is substantiated by studies such as Cunha, Heckman, and Navarro (2005) that find positive potential returns to college for many individuals who do not attend.[1] Because the actual returns to college are large for many people, errors in this decision will have economically significant effects on income and other outcomes. In order to inform policies that seek to affect individuals' college attendance decisions, it is not enough to estimate their actual returns to college; it is also necessary to estimate their perceived returns to college. Policies that cause perceived returns to look more like actual returns, either by providing information or introducing subsidies, will cause more efficient allocations of individuals into college.

In this paper I develop and implement a methodology for estimating the distribution of perceived returns to college. Using my method, I predict heterogeneous effects across the population on attendance for any given counterfactual change in well-publicized tuition subsidies regardless of whether the policy is applied uniformly across the population or is applied heterogeneously according to individuals' observed characteristics.[2] The primary contribution of this paper is that it is the first to estimate the distribution of perceived returns to college without depending on estimates of actual returns or assumptions regarding agents' knowledge of components of these returns other than pecuniary costs. I do this by estimating the causal effect of tuition on college attendance and comparing this to estimated relationships between individual characteristics and college attendance. I find that my estimates of perceived returns are consistent with the effects of tuition subsidies on attendance that previous studies of natural experiments have found, suggesting that this method can be used to successfully forecast the effects of counterfactual policies on college attendance.

The policy problem at hand is that while the socially optimal allocation of individuals into college requires assignment of individuals based on their actual social returns to college, individuals' actual attendance decisions are determined instead by their perceived private returns to college (and perceived ability to pay). If perceived and actual returns are different in sign or if individuals believe they are credit constrained, policy interventions that alter individuals'

[1]There is also evidence of negative returns for some individuals who do attend college.

[2]The caveat that any such policy must be well-publicized arises from the intuition that individuals will only respond to a policy if they are aware of its effects.

college attendance decisions can be welfare-improving. Information frictions interfere with optimal allocations of individuals into college most obviously by driving a wedge between perceived private returns and actual private returns, but also through interactions with other frictions. Specifically, information frictions interact with externalities if individuals are at all altruistic and have imperfect information about other individuals' preferences, and information frictions interact with credit constraints if perceived credit constraints are different from actual credit constraints.

It follows that in order to fully inform policy, we require estimates of both perceived private utility returns and actual social returns. The social return is comprised of actual private pecuniary returns, actual private nonpecuniary returns, and public returns associated with college attendance. Examples of work on these individual elements include Carneiro, Heckman, and Vytlacil (2011) who find that college attendance is strongly associated with private pecuniary returns to college, Oreopoulos and Salvanes (2011) who find that average nonpecuniary returns to college are potentially even larger than pecuniary returns, and Iranzo and Peri (2009) who find that pecuniary externalities from college are comparable in magnitude to typical estimates of private pecuniary returns. Estimates of perceived returns as obtained in this paper thus contribute a necessary piece of this policy puzzle.

A major advantage of the methodology employed in this paper is that because I do not rely on estimates of actual returns to infer perceived returns, I do not need to parse out the individual contributions of private pecuniary returns, private nonpecuniary returns, and externalities (insofar as they are internalized through altruism) to perceived returns. This allows me to avoid the difficulties involved in estimating these objects as well as the potentially greater difficulties involved in confidently establishing relationships between them and perceived returns.[3] Because the method I use relies on revealed preference arguments regarding observed college attendance, it naturally obtains estimates in terms of the underlying variable that drives attendance, namely, perceived utility returns. The conversion of these utility returns into a dollar scale is accomplished with a straightforward assumption on the perceived marginal cost to students of each dollar of tuition.

Existing research regarding agents' perceived returns to education relies on elicitation or

[3]For instance, because this method does not rely on earnings data, it is immune to selection bias from unobserved earnings for individuals who are not in the workforce. As a result, I have no need to take steps to correct for it such as excluding women from my sample (as is sometimes done in the literature on returns to education because of their low labor force participation relative to men).

estimation (or some combination thereof) of beliefs. Each of these present the researcher with substantial challenges. Elicitation can suffer from a lack of availability, as common data sources infrequently contain responses regarding beliefs about all of the objects of interest to researchers, and can suffer from a lack of reliability, as individuals' survey responses to questions about their beliefs may not correspond to the notion of beliefs used by the researcher.[4] These concerns are reduced for common experimental applications in which availability can be addressed by the experimental design, and reliability is improved both by increased researcher control over question framing and weaker required assumptions about the relationship between respondents' responses to questions and their actual beliefs.[5] In contrast, estimation of beliefs has the benefit that it is based on agents' observed choices rather than potentially unreliable survey responses, but has the disadvantage that beliefs and preferences cannot be jointly estimated, so assumptions must be made about agent preferences to estimate beliefs.[6] These approaches can be blended together by using elicited information on the subset of agent beliefs for which such information is available and reliable and using revealed preference to estimate other beliefs. A more comprehensive discussion of elicitation and estimation of beliefs can be found in Manski (2004).

Because of the lack of availability of reliable elicited information on perceived returns to college in known data sources, I will rely on estimation of beliefs by revealed preference.[7] Cunha and Heckman (2007) provide a valuable overview of related work which estimates heterogeneous ex ante and ex post returns to various education levels in a variety of environments.[8] The method used in these papers (referred to as the CHN method, after Cunha, Heckman, and Navarro) relies on estimates of the distribution of ex post returns to estimate ex ante returns. The main assumption here is that if agents act in accordance with a given component of their real returns

[4]Individuals' responses regarding beliefs may differ from the beliefs sought by the researcher if they are confused about the question, if demand effects are present, or if interpretation is required to translate responses from the form in which they are provided by respondents to the form in which they are relevant to the economic model. The existence of the experimental literature on how best to elicit beliefs such as Trautmann and Van De Kuilen (2015), further suggests the salience of these concerns.

[5]Jensen (2010), Zafar (2011), and Wiswall and Zafar (2015) are good examples of experimental research in which beliefs are elicited and these concerns are minimal. Because these papers use beliefs as predictors of heterogeneous treatment effects, it is not required that elicited beliefs correspond directly to actual beliefs, but only that they are a valid proxy for actual beliefs, a much weaker assumption.

[6]The problems with jointly estimating beliefs and preferences are described in more detail in Manski (1993).

[7]I am aware of no data source which elicits beliefs about individuals' net present value lifetime returns to college, the object of interest regarding college attendance. Even if such a data source existed, the reliability of responses would be suspect if respondents could conceivably vary in their interpretation of the question. For instance, if respondents differ in whether they incorporate beliefs about nonpecuniary costs into their responses about lifetime returns, the resulting distribution of elicited returns would lack a consistent interpretation.

[8]This includes Carneiro, Hansen, and Heckman (2001, 2003); Cunha and Heckman (2006); Cunha, Heckman, and Navarro (2005, 2006); Navarro (2005); and Heckman and Navarro (2007).

(such as the component associated with cognitive ability), they have full information on that component of returns.[9]

The estimation of perceived returns to college in this paper relies on the same revealed preference intuition, but uses estimates of the effect of tuition on attendance from both maximum likelihood and moment inequalities developed by Dickstein and Morales (2018) (henceforth, DM) to identify the scale of perceived returns rather than using estimates of real returns. These methods require the specification of a known relationship only between tuition and perceived returns to college which results in an estimated distribution of perceived returns with minimal dependence by construction on real returns.[10] This improvement occurs because the methods in this paper provide estimates of perceived returns conditional on agent characteristics without requiring that the researcher take a stance on whether these characteristics or their effects on returns are strictly known or unknown to agents, allowing for the possibility that agents have partial knowledge or even biased beliefs about the associated components of returns to college.[11] Allowing for partial knowledge of each component of returns allows for the estimated distributions of perceived returns and actual returns to differ in scale, while allowing for biased beliefs on each component of returns allows the distributions to differ in position.

The plan of the rest of this paper is as follows. Section 2 introduces the empirical model. Section 3 describes the econometric strategy and the assumptions required for identification. Section 4 discusses the data used in estimation of the model. Section 5 provides the results and discusses their implications. Section 6 considers some counterfactual policies. Section 7 concludes.

2 Model

The two-sector generalized Roy (1951) model provides a useful framework for considering selection based on potential outcomes. I denote schooling choices with S_i, where $S_i = 1$ denotes

[9]CHN assume rational expectations when identifying ex ante returns. Specifically, they assume that individuals' beliefs about known components of returns are equal to the components' actual individual-specific true values and that beliefs about unknown components of returns are equal to their average values. The first of these assumptions can mistake the scale of perceived returns if agents act on partial information about certain components of returns, while the second restricts unknown components of real returns from having an effect on perceived returns, effectively ruling out systemic bias in perceived returns.

[10]Rational expectations is one example of the assumption on beliefs about tuition. Some assumed dependence between perceived returns and actual returns is retained by the assumption that agents' expectations of tuition can be defined in terms of actual tuition.

[11]In brief, the methods used in the current paper rely on an accurate assumption about the perceived cost to students of one dollar of tuition, while the CHN method relies on an accurate assumption about the mappings from real returns to perceived returns for components of returns depending on whether they are known or unknown.

the choice of the college sector for agent i and $S_i = 0$ denotes the choice of the high school sector. Let $Y_{1,i,t}$ denote agent i's ex post earnings in time t conditional choosing the college sector. Let $Y_{0,i,t}$ denote the analogous object conditional on choosing the high school sector. Let $P_{i,t}$ denote i's ex post net psychic costs associated with the college sector in year t, which expresses in monetary terms costs associated either with college itself or with outcomes of the college decision, such as job characteristics. Finally, let $T_{i,t}$ denote pecuniary (tuition) costs which captures tuition paid if the college sector is chosen. I allow the interest rate $r_{S,i,t}$ to vary by time period, individual, and educational choice.

I define ex post net present value income in the college sector as

$$Y_{1,i} = \sum_{t=0}^{\infty} \frac{Y_{1,i,t}}{1 + r_{1,i,t}}, \tag{1}$$

net present value income in the high school sector as

$$Y_{0,i} = \sum_{t=0}^{\infty} \frac{Y_{0,i,t}}{1 + r_{0,i,t}}, \tag{2}$$

net present value psychic costs, expressed in monetary units, as

$$P_i = \sum_{t=0}^{\infty} \frac{P_{i,t}}{1 + r_{1,i,t}}, \tag{3}$$

and net present value pecuniary (tuition) costs as

$$T_i = \sum_{t=0}^{\infty} \frac{T_{i,t}}{1 + r_{1,i,t}}. \tag{4}$$

From the perspective of the agent, $\{Y_{1,i,t}, Y_{0,i,t}, P_{i,t}, T_{i,t}\}$ are random variables at time $t = 0$. I take the time series indefinitely into the future noting that there will be no difference between potential outcomes in the high school and college sectors for sufficiently large t due to agent death.

Denote i's (mis)information set as $\mathcal{I}_i$, which contains their potentially inaccurate and/or uncertain perceptions about future earnings, future net costs, and their determinants, such as interest rates, where $G_i(\mathcal{I}_i)$ gives the joint cdf of i's beliefs about potential outcomes. Under perfect credit markets or linear utility, the net present value potential outcomes are sufficient to describe agents' decisions. I assume agents choose their college attendance decision to maximize perceived net present value utility such that

$$S_i = \begin{cases} 1 \text{ if } \int_{-\infty}^{\infty} u_i(Y_{1,i} - P_i - T_i\gamma)\, \mathrm{d}G_i(\mathcal{I}_i) \geq \int_{-\infty}^{\infty} u_i(Y_{0,i})\, \mathrm{d}G_i(\mathcal{I}_i) \\ 0 \text{ otherwise,} \end{cases} \tag{5}$$

wherein $u_i(\cdot)$ gives i's utility function and γ gives the mapping of tuition costs into net present value earnings, allowing for the student to not bear the full cost of attendance. Note that the above expression is quite flexible. It allows not only for differences in earnings between the college and non-college sectors, but for different interest rates and differences in longevity (via beliefs regarding potential outcomes as t increases) as well as differences between agents' utility functions.

If $u_i(\cdot)$ is monotonic, $\exists!$ scalar Y_i that satisfies

$$\int_{-\infty}^{\infty} u_i(Y_{1,i} - P_i - T_i\gamma)\, \mathrm{d}G_i(\mathcal{I}_i) = \int_{-\infty}^{\infty} u_i(Y_{0,i} + Y_i)\, \mathrm{d}G_i(\mathcal{I}_i) \tag{6}$$

where I define Y_i as i's perceived return to college attendance. Y_i is compensating variation; it has the policy-relevant interpretation as the dollar amount required to be given to i in the high school sector at time zero that would induce them to attend high school rather than college.[12] Given Y_i's interpretation as compensating variation, it is a sufficient statistic for the education choice,

$$S_i = \begin{cases} 1 \text{ if } Y_i \geq 0 \\ 0 \text{ otherwise.} \end{cases} \tag{7}$$

This frames perceived returns to college as the latent variable in the college attendance decision. The primary challenge in estimating this latent variable will be identifying the scale. To do this, I note that $\exists!$ scalar certainty-equivalent tuition $\widetilde{T}_i$ that satisfies

$$\int_{-\infty}^{\infty} u_i(Y_{1,i} - P_i - \widetilde{T}_i\gamma)\, \mathrm{d}G_i(\mathcal{I}_i) = \int_{-\infty}^{\infty} u_i(Y_{0,i} + Y_i)\, \mathrm{d}G_i(\mathcal{I}_i). \tag{8}$$

Assuming constant absolute risk aversion utility with risk parameter φ_i allows extraction of the

[12] Note that when agents are certain about potential outcomes and credit markets are perfect, $Y_i = Y_{1,i} - Y_{0,i} - P_i - T_i$. Under linear utility, the same equality holds if we define $(Y_{1,i}, Y_{0,i}, C_i)$ as agents' expected values of potential outcomes regardless of market structure.

constant terms, Y_i and $\widetilde{T}_i$, yielding[13]

$$\begin{aligned} Y_i &= W_i - \widetilde{T}_i\gamma \\ W_i &= ln\Big(\int_{-\infty}^{\infty} e^{-\varphi_i Y_{0,i}}\,\mathrm{d}G_i(\mathcal{I}_i)\Big) - ln\Big(\int_{-\infty}^{\infty} e^{-\varphi_i (Y_{1,i}-P_i)}\,\mathrm{d}G_i(\mathcal{I}_i)\Big). \end{aligned} \tag{9}$$

Given some observed characteristics X_i, I model the uncertain components of perceived returns as

$$W_i = X_i\beta + \epsilon_i, \tag{10}$$

such that perceived returns can be expressed as

$$Y_i = X_i\beta - \widetilde{T}_i\gamma + \epsilon_i, \tag{11}$$

wherein ϵ_i is a mean zero error term and both β and ϵ_i capture not only ex ante perceived earnings and costs, but also differences in risk-aversion and misperceptions regarding tuition costs.[14] The goal of this paper is to estimate the distribution of Y_i, relying on an assumption on γ (e.g. $\gamma = 1\ \forall i$) using this expression of the latent variable and the revealed preference argument articulated in (7).

3 Empirical Strategy

It follows from the model that, given X_i and an assumption on γ, Y_i can be estimated using the revealed preference argument in (7) with assumptions on ϵ_i and $\widetilde{T}_i$. The assumption on γ is based on causal intuition, so a causal identification strategy is required to obtain consistent estimates of Y. The first, and likely most familiar, challenge is the possibility of correlation between perceived tuition and the unobserved error. The second challenge involves potential bias from using the necessity of using observed tuition in place of the unobserved certainty-equivalent tuition. Here, I discuss three alternative methods for addressing these challenges and describe the conditions under which they consistently estimate perceived returns.

It is helpful to the exposition to begin by presenting the general model of agents' beliefs

[13]CARA implies that certainty equivalent tuition is increasing linearly in shifts of the uncertain distribution of tuition by a constant. The perceived return increases linearly with certainty equivalent tuition by the definition of compensating variation.

[14]Note that nonlinearities in net uncertain returns can be captured by the linear-in-parameters production function by including polynomials of observed characteristics.

about tuition that will be common to all three methods,

$$Tuition_i = \widetilde{T}_i + X_i\alpha + \nu_i, \tag{12}$$

where tuition observed by the researcher is composed of the component that is known to agents, $\widetilde{T}_i$, an unknown component that is explained by the controls, $X_i\alpha$, (where α gives the matrix of misperceptions attributable to each control), and an unknown mean-zero residual, ν_i.[15] Rational expectations amounts to the assumption $\alpha = 0$ while perfect information adds the assumption that $\nu_i = 0 \ \forall i$. In general, this will allow us to represent an empirically tractable version of perceived returns,

$$\begin{aligned} Y_i &= X_i\beta - \widetilde{T}_i\gamma + \epsilon_i \\ &= X_i\beta - Tuition_i\gamma + X_i\alpha + \nu_i\gamma + \epsilon_i \\ &= X_i\theta - Tuition_i\gamma + \nu_i\gamma + \epsilon_i, \end{aligned} \tag{13}$$

by substituting in tuition observed by the researcher for agents' unobserved perceived tuition and defining $\theta = \beta + \alpha$. This representation presents misperceptions as omitted variables in the case where $Var(\nu_i \neq 0)$ which will causes bias if the researcher does not include all relevant controls in the model. The distinction between the extent to which each control contributes to misperceptions in prices, α, and to perceived returns, β, is presented to emphasize that the methods in this paper are *robust* to systematic bias in perceptions, even though they are not separately identified.

First, I consider a model in which misperceptions in tuition are fully explained by X_i such that $\nu_i = 0 \ \forall i$, and perceived tuition are exogenous.[16] In this model, the tuition observed by the researcher can be used in place of perceived tuition in the latent variable equation. This combined with the assumed lack of conditional correlation between tuition and unobserved components of perceived returns leads to consistent estimates of perceived returns from a simple adjustment to standard binary choice methods where the coefficient on $Tuition_i$ is constrained to be equal to the assumed value γ.

[15] Constant absolute risk aversion utility produces the result that certainty-equivalent tuition moves additively with shifts in the entire distribution of beliefs about tuition. The same result holds regardless of the utility function when there is no uncertainty on tuition such that $\nu_i isdegenerate$.

[16] The simplest case of this assumption on perceived tuition is perfect information, in which perceived tuition are equal to actual prices. Assuming perfect information in this setting will have no effect on the estimation procedure, but will affect the interpretation of estimated parameters.

Second, I consider a model in which agents have imperfect information on tuition that is not captured by X_i, while maintaining that both $\widetilde{T}_i$ and ν_i are exogenous in the perceived returns equation. In words, the condition on ν_i implies that the magnitude and direction of unexplained misperceptions in tuition are independent of perceived returns. In this setting, the minor adjustments to standard methods described above would provide consistent estimates if we could condition on agents' unobserved certainty-equivalent tuition. Here, the moment inequalities developed by Dickstein and Morales (2018) provide bounds on perceived returns using instruments, Z_i, which are independent of agents' unexplained misperceptions about prices.

Third, and finally, I consider a model in which agents have imperfect information on tuition, and both certainty-equivalent tuition and unexplained misperceptions on tuition may be correlated with the unobserved component of perceived returns. In this setting, even if we observed certainty-equivalent tuition, their correlation with unobserved components of perceived returns would produce bias in estimates of the effect of tuition on selection in the standard binary choice framework, which will invalidate the (causal) assumed value for the coefficient on certainty-equivalent tuition, γ. This method requires instruments, Z_i, which are uncorrelated with both agents' unexplained misperceptions about tuition and unobserved components of perceived returns.

3.1 Degenerate Misperceptions on Exogenous Prices

First, I will describe a method for inferring perceived returns with a simple adjustment to common binary choice methods. This procedure will provide consistent estimates of the perceived returns distribution under two assumptions that are likely to be violated in many applications. This method assumes that observed tuition and unobserved idiosyncratic perceived returns, ϵ_i, are uncorrelated conditional on X_i. This amounts to assuming both that tuition, conditional on X_i, is not correlated with the perceived return to college or with misperceptions on tuition.

Here, controlling for the explanatory variables X_i is sufficient to achieve conditional unconfoundedness for observed tuition in this setting. Assuming that the distribution of the error term is given by

$$\epsilon_i | X_i, Tuition_i \sim \mathcal{N}(0, \sigma^2), \tag{14}$$

and applying the decision rule described in (7) is sufficient to consistently estimate perceived returns by maximum likelihood.[17] The probability of selection is given by

$$Pr(S_i = 1|X_i, Tuition_i) = \Phi\Big(\frac{X_i\beta - Tuition_i\gamma}{\sigma}\Big), \tag{15}$$

where $\Phi(\cdot)$ denotes the standard normal cdf. Defining $\{\beta^*, \gamma^*\} = \{\frac{\beta}{\sigma}, \frac{\gamma}{\sigma}\}$ for notational convenience and taking γ as given, the parameters $\{\beta^*, \gamma^*\}$ are the values that maximize the log-likelihood:

$$\begin{gathered}\mathcal{L}(\beta^*, \gamma^*|X_i, Tuition_i) = \\ \sum_i S_i \log\Big[\Phi\Big(X_i\beta^* - Tuition_i\gamma^*\Big)\Big] + (1 - S_i)\log\Big[1 - \Phi\Big(X_i\beta^* - Tuition_i\gamma^*\Big)\Big].\end{gathered} \tag{16}$$

The estimates of perceived returns are then given by

$$Y_i|X_i, Tuition_i \sim \mathcal{N}(X_i\hat{\beta} - Tuition_i\gamma, \hat{\sigma}^2), \tag{17}$$

where imposing the assumed value on γ (rather than the standard assumption $\sigma = 1$) is the only difference from a standard Probit.[18]

3.2 Non-Degenerate Misperceptions on Exogenous Prices

It is unlikely that a researcher would have access to all variables associated with misperceptions on tuition. For one, tuition prices are often not contracted on prior to attendance and are unlikely to be captured by other observed covariates. Individuals are also likely uninformed about tuition after it have been set. For instance, Grodsky and Jones (2007) find that individuals' elicited beliefs about tuition for postsecondary education were consistently inaccurate, with substantial variance in perceptions unexplained by a variety of controls.

This section adapts to the current setting a method developed by Dickstein and Morales (2018) that addresses the bias from imperfect information using moment inequalities. The moment inequalities allow for an independent error ν_i, but maintain that certainty-equivalent

[17] The normality assumption is unnecessary for any of the methods in this paper. For the current method, any error distribution is acceptable. I will make the restrictions on error distributions clear for other methods as they arise.

[18] Many statistical software packages, such as Stata, impose $\sigma = 1$ in their binary choice estimation commands. It is simple to convert these estimates into scaled estimates using

$$\hat{\sigma} = \frac{\gamma}{\hat{\gamma}^*} \quad ; \quad \hat{\beta} = \hat{\beta}^*\hat{\sigma},$$

taking to care to apply the delta method to obtain correct standard errors for the scaled estimates.

tuition and misperceptions on tuition are uncorrelated with ϵ_i. Recalling again the empirically tractable equation for perceived returns in (13), this method works by recognizing that $Cov(Price, \nu_i) = 1$ by construction under this assumption, as seen in (12), and leverages this insight to bound the parameters of the model. This produces problems analogous to those associated with measurement error, and Probit estimates of the effect of $Tuition_i$ on selection will suffer from attenuation bias.[19] Because the scale of the distribution of perceived returns is inferred from the estimated effect of tuition on selection, this causes upward bias in Probit estimates of σ.

The method uses two sets of moment inequalities to obtain bounds on the parameters of perceived returns, $\{\beta, \sigma\}$. I will present the inequalities and provide a brief discussion here. For the derivation and further discussion of the moment inequalities, see Dickstein and Morales (2018). In a rough sense, the inequalities work because the true parameters $\{\beta^*, \gamma^*\}$ will predict more extreme responses to observed tuition than are found in the data because some of the variation in tuition observed by researchers is not actually driving agents' decisions. The first set of inequalities are based on revealed preference, and the second are based on the score equation from a binary choice maximum likelihood model.

3.2.1 Revealed Preference Moment Inequalities

The conditional revealed preference moment inequalities are

$$\begin{aligned} &\mathbb{E}\left[-(1-S_i)(X_i\beta^* - Tuition_i\gamma^*) + S_i \frac{\phi(X_i\beta^* - Tuition_i\gamma^*)}{\Phi(X_i\beta^* - Tuition_i\gamma^*)} \middle| Z_i \right] \geq 0, \\ &\mathbb{E}\left[S_i(X_i\beta^* - Tuition_i\gamma^*) + (1-S_i)\frac{\phi(X_i\beta^* - Tuition_i\gamma^*)}{1-\Phi(X_i\beta^* - Tuition_i\gamma^*)} \middle| Z_i \right] \geq 0. \end{aligned} \tag{18}$$

These inequalities are consistent with the revealed preference argument that perceived returns are positive for those who select the investment and negative for those who do not. A formal proof and discussion of the revealed preference inequalities are provided by DM. Here, I provide a brief overview of the intuition.

Regarding the second inequality, consider an agent that selects the investment such that $S_i = 1$. Following the revealed preference argument articulated in (7) and the representation of

[19] Indeed, the structure of information frictions in the current setting are empirically indistinguishable from the case where the tuition observed by the researcher differ from agents' certainty-equivalent tuition only due to being affected by classical measurement error.

perceived returns in (11), it follows that this individual's perceived return is positive such that

$$S_i(X_i\beta - \widetilde{T}_i\gamma + \epsilon_i) \geq 0. \tag{19}$$

This object cannot be computed directly because researchers do not observe $\widetilde{T}_i$ or ϵ_i. However, the inequality will hold in expectation if it holds for individuals. Ideally, we would compute

$$\mathbb{E}\left[S_i(X_i\beta - \widetilde{T}_i\gamma + \epsilon_i)\middle| X_i, \widetilde{T}_i, S_i\right] \geq 0, \tag{20}$$

but $\widetilde{T}_i$ is not observed by the researcher. Taking the expectation conditional on $(X_i, Tuition_i, S_i)$ cannot be done because we do not know $\mathbb{E}[\widetilde{T}_i|X_i, Tuition_i, S_i]$ or $\mathbb{E}[\epsilon_i|X_i, Tuition_i, S_i]$.

As an alternative, we can condition instead on Z_i to maintain the inequality if Z_i is independent of ν_i. First, note that

$$\mathbb{E}\left[S_i(X_i\beta - Tuition_i\gamma)\middle| Z_i\right] = \mathbb{E}\left[S_i(X_i\beta - \widetilde{T}_i\gamma)\middle| X_i, \widetilde{T}_i\right], \tag{21}$$

by the law of iterated expectations. Because Z_i is not associated with the misperception, ν_i, $\mathbb{E}[Tuition_i|Z_i] = \mathbb{E}[\widetilde{T}_i|Z_i]$.[20] Second, note that

$$\mathbb{E}\left[S_i\epsilon_i\middle| X_i, \widetilde{T}_i, S_i\right] = (1 - S_i)\sigma\frac{\phi(X_i\beta^* - \widetilde{T}_i\gamma^*)}{1 - \Phi(X_i\beta^* - \widetilde{T}_i\gamma^*)} \tag{22}$$

and that

$$\begin{aligned}
&\mathbb{E}\left[(1 - S_i)\sigma\frac{\phi(X_i\beta^* - Tuition_i\gamma^*)}{1 - \Phi(X_i\beta^* - Tuition_i\gamma^*)}\middle| Z_i\right] \\
&\geq \mathbb{E}\left[(1 - S_i)\sigma\frac{\phi(X_i\beta^* - \widetilde{T}_i\gamma^*)}{1 - \Phi(X_i\beta^* - \widetilde{T}_i\gamma^*)}\middle| X_i, \widetilde{T}_i\right],
\end{aligned} \tag{23}$$

by law of iterated expectations and Jensen's inequality.

Intuitively, if observed tuition is a mean-preserving spread of certainty-equivalent tuition, using it in place of certainty-equivalent tuition will mistakenly increase expected perceived returns unconditional on selection for half of the sample and decrease them for the other half by a common average amount λ. For the half for whom this expectation increases, the expectation of the error conditional on selection approaches zero as lambda increases. For the half for whom

[20] If misperceptions, ν_i, are correlated with ϵ_i (and thereby S_i), this latter equality will not hold, and the inequalities will not necessarily hold at the true parameter values.

it decreases, the expectation of the error conditional on selection approaches positive infinity as lambda increases. In most cases, this second effect will dominate the overall expectation of the error conditional on selection.[21] The first inequality follows from similar intuition applied to individuals who do not select the investment.

3.2.2 Odds-Based Moment Inequalities

The conditional odds-based moment inequalities are

$$\begin{aligned} &\mathbb{E}\left[\left(S_i \frac{1-\Phi(X_i\beta^* - Tuition_i\gamma^*)}{\Phi(X_i\beta^* - Tuition_i\gamma^*)} - (1-S_i)\right)\middle| Z_i\right] \geq 0, \\ &\mathbb{E}\left[\left((1-S_i) \frac{\Phi(X_i\beta^* - Tuition_i\gamma^*)}{1-\Phi(X_i\beta^* - Tuition_i\gamma^*)} - S_i\right)\middle| Z_i\right] \geq 0. \end{aligned} \tag{24}$$

They are derived from the conditional score equation,

$$\mathbb{E}\left[S_i \frac{\phi(X_i\beta^* - \widetilde{T}_i\gamma^*)}{\Phi(X_i\beta^* - \widetilde{T}_i\gamma^*)} - (1-S_i)\frac{\phi(X_i\beta^* - \widetilde{T}_i\gamma^*)}{1-\Phi(X_i\beta^* - \widetilde{T}_i\gamma^*)}\middle| X_i, \widetilde{T}_i\right] = 0. \tag{25}$$

Considering the bottom inequality, the score function can be rearranged to be a function of the model-predicted odds of selecting state one,

$$\mathbb{E}\left[\left((1-S_i)\frac{\Phi(X_i\beta^* - \widetilde{T}_i\gamma^*)}{1-\Phi(X_i\beta^* - \widetilde{T}_i\gamma^*)} - S_i\right)\middle| X_i, \widetilde{T}_i\right] = 0. \tag{26}$$

The advantage of this transformation is that the odds-ratio is globally convex in its arguments. Replacing the unobserved $\widetilde{T}_i$ with $Tuition_i$ changes the equation into an inequality by application of Jensen's inequality due to the global convexity of the odds ratio. As the index of the odds ratio increases, the model-predicted odds of a given outcome approach positive infinity, while the odds approach zero as the index decreases. When the index is replaced with a mean-preserving spread of itself (via replacing perceived tuition with prices), this first effect will usually dominate the second regardless of the distributional assumption.[22] This inequality holds when taking its expectation conditional on Z_i by law of iterated expectations. The first inequality follows from similar intuition for those who do not select the investment.

[21]Global convexity of $\mathbb{E}[\epsilon_i|\epsilon_i < \kappa]$ in κ is necessary for the inequalities to hold regardless of the value of κ and the variance of the misperception term. This condition is satisfied by both the normal and logistic distributions.

[22]Global convexity of the odds ratio is necessary for this condition to hold for all values of the index and for all magnitudes of mean-preserving spreads. This condition is satisfied by log-concave distributions, such as the normal and logistic.

It may seem like the two moment inequalities in (24) would be redundant as they are both derived from transformations of the same score function. The convexity of the odds ratios essentially causes the model to overpredict both the odds of selection and the odds of nonselection when using tuition instead of certainty-equivalent tuition. The underlying, erroneous, assumption is that agents are responding to all of the variation in tuition rather than only to perceived variation in tuition. If this were the case, observed selection behavior would imply that agents are less price-sensitive than they actually are, which would drive attenuation bias in the effect of tuition on selection if the potential for misperceptions is ignored. A further discussion of the intuition behind these inequalities is available in Dickstein and Morales (2018).

3.2.3 Estimation Using Moment Inequalities

Under the information assumptions provided, the true parameter $\psi^* = \{\beta^*, \gamma^*\}$ will be contained within the set of parameters that satisfy the inequalities, which I define as Ψ_0^*. First, because it is computationally expensive to compute the inequalities conditional on Z, I will instead use unconditional inequalities that are consistent with the conditional inequalities described above. Additionally, in small samples it is possible that the true parameters will not strictly satisfy these inequalities, so it is necessary to construct a test of the hypothesis that a given value $\psi_p^* = \{\beta_p^*, \gamma_p^*\}$ is consistent with the inequalities. To do this I employ the modified method of moments procedure described by Andrews and Soares (2010). A description of the estimation procedure is available in Appendix B.

3.3 Non-Degenerate Misperceptions on Endogenous Prices

In this section I describe a control function approach that has several advantages over the moment inequalities for the current application. First, it is robust to endogeneity in both perceived and unknown prices.[23] Second, it allows for the misperception term ν_i to be correlated freely with ${}_i$ (it does not assume that it is an i.i.d. white-noise error). Third, it is substantially less computationally costly, allowing for the inclusion of a richer set of explanatory variables. This provides for a broader set of heterogeneous policy predictions conditional on observed covariates. Fourth, it provides point estimates of model parameters. Fifth, it places no restrictions

[23]DM do not use the moment inequalities to address endogeneity because their structural model addresses it by providing a forecasting equation for profit as a function of revenue and costs. In their framework, revenue forecasts profit at a constant marginal rate defined by the demand elasticity. Naturally, if the relationship were causal, this constant marginal rate would be 1.

on the distribution of the error term.[24]

The control function approach uses instruments Z_i that are uncorrelated with both ν_i and ϵ_i in the following system of equations:

$$\begin{aligned} Y_i &= X_i\beta - \widetilde{Tuition}_i\gamma + \epsilon_i \\ \widetilde{T}_i &= Z_i\pi + u_i \\ Tuition_i &= \widetilde{T}_i + X_i\alpha + \nu_i = Z_i\delta + u_i + \nu_i, \end{aligned} \tag{27}$$

wherein π provides the mapping of the instruments to certainty-equivalent tuition, while δ provides the mapping to actual tuition. For the J instruments $[z_{i,1}, ..., z_{i,J}]$ that are not contained in X_i, the corresponding $[\delta_1, ..., \delta_J] = [\pi_1, ..., \pi_J]$, while $[\delta_1, ..., \delta_J] = [\pi_1 + \alpha_1, ..., \pi_J + \alpha_J]$ for the instruments that are in both X_i and Z_i. The second line represents what would be the first stage in a two-step instrumental variables procedure if certainty-equivalent tuition were observable. The third line combines the assumption on beliefs given in (12) with the first stage on certainty-equivalent tuition to obtain an estimable first stage equation in terms of observed prices.[25]

Given the above, I derive a control function that can be used to estimate perceived returns as follows,[26]

$$\begin{aligned} Y_i =& X_i\beta - \widetilde{Tuition}_i\gamma + \epsilon_i \\ =& X_i\beta - Tuition_i\gamma + \nu_i\gamma + \epsilon_i \\ =& X_i\beta - Tuition_i\gamma + (u_i + \nu_i)\rho_{u\nu} + \eta_i \\ =& X_i\beta - Tuition_i\gamma + (\widehat{u_i + \nu_i})\rho_{u\nu} + \zeta_i. \end{aligned} \tag{28}$$

The second line follows directly from the representation of beliefs about tuition given in (12).

[24] This last advantage is quite small in practice. Formally, the assumed distribution for the error in perceived returns must be both log-concave and have globally convex expectations given selection and non-selection. As the normal and logistic distributions satisfy both of these requirements, this produces no substantive restriction for applications that would make otherwise make one of these distributional assumptions. Furthermore, as discussed above, the inequalities will often hold locally even if they do not hold globally for other distributions, making it probable that the inequalities will happen to hold at the true parameter vector in applied work under other distributional assumptions.

[25] The presentation here assumes a parametric first stage. It is also possible to construct a nonparametric first stage using the expectation of $Tuition_i$ conditional on Z_i.

[26] As an alternative, we could perform a two stage procedure with $Y_i = X_i\beta - Z\hat{\delta}\gamma - u_i\gamma + \epsilon_i$, where $\hat{\delta}$ is obtained from the first stage regression of $Price$ on Z. This formulation will obtain valid estimates of β and $\omega^2 = Var(-u_i\gamma + \epsilon_i)$, leading to $\hat{Y}_i \sim \mathcal{N}(X_i\hat{\beta} - \widehat{Price}_i\gamma, \hat{\omega}^2)$. This approach has the disadvantage of moving the individual-specific information on perceived returns contained in $(\widehat{u_i + \nu_i})$ into the error term, while the control function approach conditions on this observed variation.

The third line substitutes in the linear projection of the composite error $\nu_i\gamma + \epsilon_i$ on the first stage error $u_i+\nu_i$, wherein $\rho_{u\nu} = \mathbb{E}[(u_i+\nu_i)(\nu_i\gamma+\epsilon_i)]/\mathbb{E}[(u_i+\nu_i)^2]$.[27] The fourth line substitutes the estimated residuals from the first stage regression of $Tuition_i$ on Z_i in for their unobserved true values, generating a new error, ζ_i. This new error will converge asymptotically to η_i, but will differ in small samples due to sampling error in the estimation of the residual from the first stage, $(\widehat{u_i + \nu_i})$. Note that it is unnecessary (and impossible) in this setting to distinguish between u_i and ν_i, and that (u_i, ν_i, ϵ_i) can be freely correlated if Z_i is a valid instrument for certainty-equivalent tuition.

To estimate perceived returns, I assume that the error in the perceived returns control function is normally distributed,

$$\eta_i \sim \mathcal{N}(0, \sigma^2), \tag{29}$$

where I represent the standard deviation of the error using the same parameter, σ, that represented the standard deviation of ϵ_i in the preceding sections, recognizing that $\eta_i = \epsilon_i$ by assumption in those settings. The log-likelihood for the control function approach is then given by

$$\begin{gathered}
\mathcal{L}\Big(\beta^*, \gamma^*, \rho^*_{u\nu}|X_i, (\widehat{u_i + \nu_i})\Big) = \\
\sum_i S_i \log\left[\Phi\Big(X_i\beta^* - Tuition_i\gamma^* + (\widehat{u_i + \nu_i})\rho^*_{u\nu}\Big)\right] \\
+(1 - S_i)\log\left[1 - \Phi\Big(X_i\beta^* - Tuition_i\gamma^* + (\widehat{u_i + \nu_i})\rho^*_{u\nu}\Big)\right].
\end{gathered} \tag{30}$$

Estimates of perceived returns are obtained by plugging the estimated parameters and the assumed coefficient on perceived tuition into the latent variable equation:[28]

$$Y_i|X_i, Tuition_i, (\widehat{u_i + \nu_i}) \sim \mathcal{N}\Big(X_i\hat{\beta} - Tuition_i\gamma + (\widehat{u_i + \nu_i})\hat{\rho}_{u\nu}, \hat{\sigma}^2\Big). \tag{31}$$

In addition to being robust to a wider variety of information settings than the preceding alternatives, the control function approach has the added benefit of providing estimates of $\rho_{u\nu}$. These estimates are useful primarily because they allow for more precise estimation of idiosyncratic variation in perceived returns. Secondly, they provide insight into the type and

[27] Note that under the assumptions of the preceding section, $Cov(u_i, \nu_i\gamma + \epsilon_i) = 0$, and $Cov(\nu_i, \nu_i\gamma + \epsilon_i) = \gamma$ such that $\rho_{uv} = \gamma$.

[28] I present the estimation of the first stage generally here for simplicity, and because the insights of this paper are invariant to the estimator used. In the simulations, I assume normality on the first stage error and use two-stage conditional maximum likelihood to estimate perceived returns, following Rivers and Vuong (1988).

magnitude of bias expected from the preceding methods.

4 Data

The primary dataset used is the Geocode file of the 1979 National Longitudinal Survey of Youth (NLSY79). The NLSY79 is a longitudinal, nationally representative survey of 12,686 youths who were 14-22 years old when they were first surveyed in 1979. Respondents were first interviewed in 1979, were interviewed annually through 1994 and have been interviewed biannually since then. This data source provides a wide variety of information on individuals who were between the ages of 14 and 22 in 1979. Vitally, it provides information on the college(s) that individuals attended if they attended college as well as loans and financial aid received during college. Because the Geocode file provides detailed geographic information, it can also be combined with other datasets to obtain average tuition for both local colleges in individuals' counties of residence at age 17 and actual tuition for the college that they attended. The geographic information is also useful for obtaining information on local labor market characteristics. This dataset also includes a rich set of information on individuals' academic abilities and family characteristics that are predictive of college attendance, including information about the percentage of college costs that individuals pay themselves. As a final advantage, this dataset has been used extensively in the related literatures on ex ante returns such as Cunha, Heckman, and Navarro (2005) and Cunha and Heckman (2016) and the effects of policy interventions on college attendance such as Dynarski (2003) so that the relationships between this paper's results and those of existing work can be readily attributed to differences in methodology rather than differences across datasets. I merge this dataset with data on colleges from the Integrated Postsecondary Education Data System (IPEDS), and data on local and state labor markets from the Bureau of Economic Analysis (BEA) and the Bureau of Labor Statistics (BLS).

Other than dropping 41 individuals who reported graduating from college without ever attending college, I do not impose any limitations on the sample. Notably, because I do not use actual income to infer perceived returns, there is no reason to exclude women due to fertility and labor force participation concerns as is common in the literature. The initial sample of 12,686 is reduced to 5,492 due to missing observations for variables of interest. A description of the data is provided in Table 1.[29]

[29] The transformation of ASVAB scores to have positive support is required because a cancellation that takes place in the derivation of the moment inequalities requires that each variable's support in the data have a common

Table 1: Description of the Primary Variables

	Overall	High School Graduates	College Attendees
Attend College	0.501	0.000	1.000
Female	0.501	0.482	0.520
Black	0.104	0.117	0.092
Hispanic	0.047	0.057	0.038
High School GPA	2.480	2.124	2.835
Mother Education	11.833	11.005	12.658
Father Education	11.997	10.821	13.169
Number of Siblings	3.128	3.485	2.772
ASVAB Score Subtest 3	1.436	0.620	2.250
ASVAB Score Subtest 4	1.903	0.423	3.378
ASVAB Score Subtest 5	2.367	1.119	3.610
Family Income	69675	58858	80456
Broken Home	0.286	0.337	0.234
Age in 1979	16.231	16.081	16.381
Urban Residence at Age 14	0.757	0.723	0.791
Average County Wage at Age 17	11.318	11.418	11.218
State Unem. Rate at Age 17	7.242	7.395	7.090
Net Tuition	11,142	7,151	15,120
Local Tuition at Age 17	12,575	12,559	12,590
Observations	3324	1843	1481

Notes: Means are of all NPSY79 samples. Parents' education is in years. High school GPA is out of a maximum value of 4. All dollar values are adjusted to 2018 values using a 3% interest rate. Each ASVAB test score is transformed to have unit variance and zero mean. Broken home indicates the absence of either biological parent in the home for any year from birth to age 18.

I choose college attendance as the decision of interest.[30] This assumes that individuals who attend college do so because they perceive the return to completing a 4-year degree to be positive. If any individuals begin college intending to drop out because their perceived returns to fewer than 4 years of college are positive but their perceived returns to 4 years of college are negative, I will overestimate their perceived returns to college by using attendance as the relevant decision. I expect that such individuals are rare. I find that approximately 57% of my sample attended college after high school. This rate is somewhat lower in my data than the current average because college enrollment was lower in the early 1980's (when the individuals in my sample were attending college) and because the NLSY79 contains oversamples of poor whites and minorities who are less likely to attend college than average. I code an individual

sign. This transformation has no substantive effect on the estimation as the constant in the model will adjust to offset it. Average county wages come from the Bureau of Economic Analysis, and state unemployment rates come from the Bureau of Labor Statistics. These are matched to the primary dataset using the NLSY79's geographic information.

[30]The same estimation procedure could be performed on graduation, but would somewhat complicate the interpretation as dropping out suggests dynamic changes in information. Extensions of the estimation strategy that allow for ordered decisions and information dynamics would be well suited to investigating differences between attendance and completion and are left to future research.

as having attended college if they explicitly report having received a college degree by age 23 or if they report a highest grade attended above 12 by age 23.

Because I focus on a single decision at a single point in time, I do not convert the NLSY79 dataset into panel data. I instead use the longitudinal data to obtain information about the timing of college attendance, college tuition, and scholarships in years prior to receipt of a bachelor's degree and to obtain retrospective information that influences the college attendance decision. I use four times the average present value (in 2018 terms, using a 3% interest rate) of tuition in all years prior to receipt of a bachelors degree to construct a total present value tuition measure for individuals who complete a 4-year degree. I use the information on tuition for individuals who complete college and attend 4-year colleges to impute counterfactual 4-year tuition for individuals who did not attend college.

Noting that I only observe tuition for individuals that attend college and that individuals only attend college if their perceived return to college is positive, I impute tuition in a manner that is consistent with the model of college attendance described above. I control for variables X_T while accounting for selection with the system:

$$Tuition_i = \begin{cases} X_{iT}\alpha_T + Local\_Tuition\_17_i\alpha_{LT} + \xi_{iT} & \text{if } S_i = 1 \\ . & \text{otherwise,} \end{cases} \tag{32}$$

$$S_i = \begin{cases} 1 \text{ if } Z_{iT}\alpha_S - Local\_Tuition\_17_i\alpha_{LS} + \xi_{iS} > 0 \\ 0 \text{ otherwise,} \end{cases} \tag{33}$$

in which a variable within Z_T satisfies the exclusion restriction that it is not included in X_T. Because I do not observe tuition for people who do not attend college (the very problem I seek to address), I use local tuition at age 17 (the instrument for tuition) in these equations instead of actual tuition. Secondly, I argue that distance from college at age 14 provides variation in selection that does not otherwise affect tuition, such that I can exclude it from X_T while including it in Z_T.

Using distance to college as an instrument for educational attainment was introduced by Card (1993) to estimate the effect of education on earnings. Its use is similar here in predicting college attendance, but the identification here relies on it having no effect on tuition conditional on other controls, while making no assumption on its effect on earnings. An additional concern specific to tuition is that distance to college may be associated with college prices, for instance if more rural areas are more likely to have small community colleges than urban areas. I address this concern by controlling for local tuition in county of residence at age 17 as well as including an indicator variable for living in an urban county. Conditional on AFQT, local tuition at age 17, urbanicity of residence, and the other controls in X_T, I argue that distance to college only contributes a measure of the potential costs associated with housing and transportation associated with college attendance, which should predict attendance without otherwise affecting tuition.

Relatedly, use of average local tuition at age 17 as an instrument for the effect of college attendance on earnings was introduced by Kane and Rouse (1995). I include this as a control for the imputation of tuition while using it as an instrument in estimation of perceived returns. For individuals who live in a county with a college, I use the enrollment-weighted average tuition of public 4-year colleges in their county. For individuals who do not live in a county with a college, I use the state-level enrollment-weighted average. Instead of relying on this instrument affecting selection without otherwise affecting earnings as it has primarily been used in the past, I rely on it affecting tuition without otherwise affecting selection.

One concern with the use of distance to college to instrument for selection into college is that it has been shown to be correlated with AFQT, a measure of ability. To address this concern, I include each ASVAB subtest, from which the AFQT score is computed, as controls in all estimated equations. Hansen, Heckman, and Mullen (2004) show that years of schooling at the time of testing affects AFQT scores, so rather than using raw ASVAB scores, I use the residual of each test score after controlling for years of education. I make no other adjustments to any of the variables in the data.

At first glance, the imputation of tuition and the estimation of the model may seem circular because I estimate a selection equation to impute tuition and then use imputed tuition to estimate a very similar selection equation for the main results. A succinct chronological ordering of each step in the estimation procedure is helpful for dispelling this potential confusion. First, I estimate the selection equation (33) using variables that are observed for everyone in my sample. Because I only use these estimates to control for selection, I am uninterested in the scale of the latent variable of this equation as well as causal effects of any variables on the latent variable. Second, I use these estimates to impute tuition with (32) while controlling for selection from (33) with the exclusion restriction that distance to college affects attendance but not potential tuition. Third, I instrument for this imputed tuition with local tuition at age 17 in the first stage. Fourth, I use the instrumented value of tuition to estimate the causal effect of tuition on selection.[31] Finally, I apply the normalization assumption that tuition affects perceived returns at known marginal rate γ. Table 2 shows the variables that are and are not included in each estimated equation both for the tuition imputation and for the main results.

Table 2: List of Variables Included and Excluded in Each System

Variable Name	Tuition (Observation) (Z_T)	Tuition (Imputation) (X_T)	Return IVs (Z)	Return (X)
Imputed Tuition	·	·	·	✓
Local Tuition, Age 17	✓	✓	✓	·
ASVAB (All Tests)	✓	✓	✓	✓
Mother's Education	✓	✓	✓	✓
Mother's Education Squared	✓	✓	✓	✓
Father's Education	✓	✓	✓	✓
Father's Education Squared	✓	✓	✓	✓
Number of Siblings	✓	✓	✓	✓
Number of Siblings Squared	✓	✓	✓	✓
Urban at Age 14	✓	✓	✓	✓
High School GPA	✓	✓	✓	✓
High School GPA Squared	✓	✓	✓	✓
Broken Home	✓	✓	✓	✓
Average County Wage, Age 17	✓	✓	✓	✓
State Unemployment, Age 17	✓	✓	✓	✓
Distance to College	✓	·	·	·

Notes: I rely on distance to college affecting attendance without directly tuition. I further rely on local tuition at age 17 affecting tuition without otherwise affecting perceived returns, conditional on the other controls. I do not include distance to college in the main equation because tuition is imputed from this variable and all other objects in X_T, such that including it in the main equation would produce perfect collinearity on imputed tuition.

I estimate a value for γ using data from the NLSY79 on the proportion of college tuition paid for by the student. This data is only available in 1979, so I impute a value for the proportion of costs paid using ordinary least squares. In practice, I will use this $\hat{\gamma}(X_i)$ when estimating perceived returns, such that each individual is allowed to differ in the amounts of pecuniary costs they bear. The estimation of $\gamma(X)$ is described in detail in Appendix D.

Finally, Because the moment inequalities are estimated using a grid search, they are highly computationally expensive. For this reason, I use principal components to reduce the parameter space to a constant, a coefficient on tuition, a coefficient on the first principal component of variables associated with individuals' general ability, and a coefficient on the first principal component of variables associated with individuals' local geographic characteristics.[32] The

[31]For the moment inequalities, steps 3 and 4 are integrated into one step.

[32]It takes approximately 16 hours to estimate this 4-parameter model. Because the primary time cost is in the grid search, adding any additional variables can be expected to increase the computation time required

details of the principal component analysis are presented in Appendix C.

5 Results

Table 3 shows estimates of the model parameters ($\{\beta, \sigma\}$) of perceived returns to college for the Probit, control function method, and moment inequality method using principal components. The bias in the Probit is evident in the insensible negative estimate of the standard deviation. Because it is identified from the assumption on γ, the estimate of the standard deviation will be negative when expected tuition is positively associated with college attendance, i.e. individuals who are likely to attend college are also likely to attend expensive colleges. The negative standard deviation affects the signs of the other coefficients because the estimates from the discrete choice model are multiplied by $\hat{\sigma}$ to convert them into dollar terms. Graphs of the implied distribution of perceived returns for the control function method and moment inequalities are shown in figure 1.

Table 3: Perceived Returns Estimates, 2018 Dollars, Principal Components

	Probit	IV Probit	Moment Inequalities
Constant	-0.451	0.120	[-10.310, 3.192]
	(0.321)	(1.300)	N/A
PC1(Ability)	-4.121	26.982	[23.511, 30.326]
	(0.407)	(33.084)	N/A
PC1(Location)	1.631	-0.532	[-4.554, 2.655]
	(0.321)	(2.545)	N/A
σ	-13.154	43.239	[32.246, 60.768]
	(1.338)	(57.458)	N/A
Observations	3324	3324	3324

Notes: Standard errors in parentheses. Parameters are marginal effects of the variable on perceived returns to college in thousands of dollars. The coefficient on tuition is assumed to be equal to $\lambda\gamma(X)$. Estimates are from equation (11) with the details varying by estimation method. See text for details.

The results in Table 3 are scaled using the estimated $\gamma(X)$ described in D. The unscaled results are presented in Table 4. It is worth noting that the moment inequality bounds in Table 4 (and consequently Table 3) fail to completely characterize the confidence set of parameters that satisfy the moment inequalities. The hyper-rectangle implied by the upper and lower bound on each parameter is larger than the actual confidence set of parameters that satisfy the moment inequalities. The resulting distributions of beliefs about returns to college are obtained for each point in this confidence set. A three-dimensional cut of the confidence set of $\{\hat{\beta}^*_{MI}, \hat{\gamma}^*_{MI}\}$ is

exponentially.

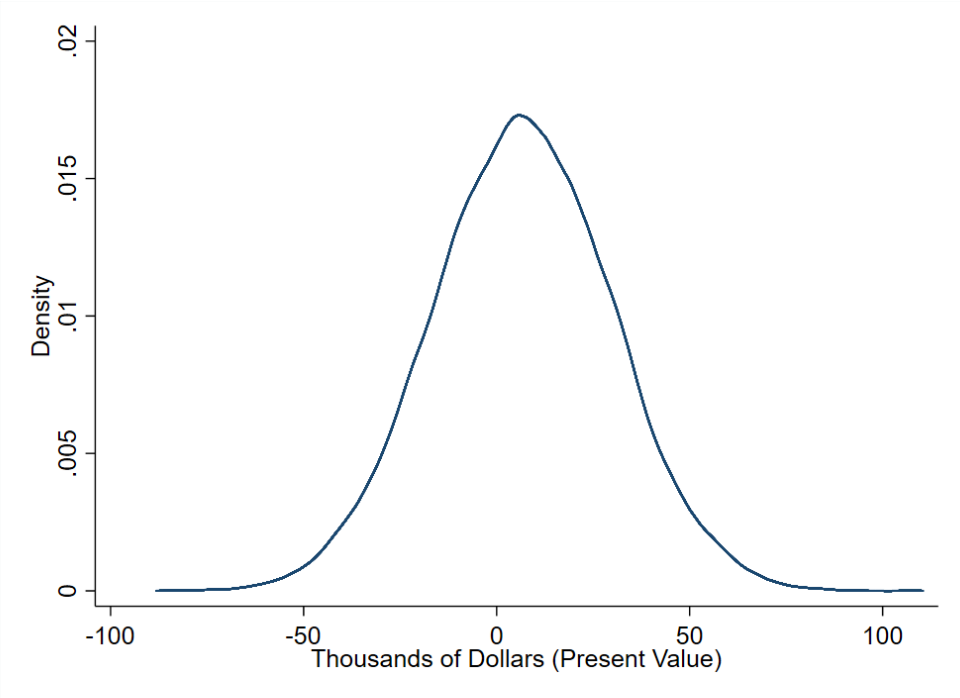

Figure 1: Perceived Returns to College, Control function method, Principal Components

Notes: Perceived returns across the population, weighted by 1988 sample weights, using principal components. The distribution is given by $Y = X\hat{\beta} - Z\hat{\delta}\lambda\hat{\gamma}(X) + \mathcal{N}(0, \hat{\sigma}^2)$.

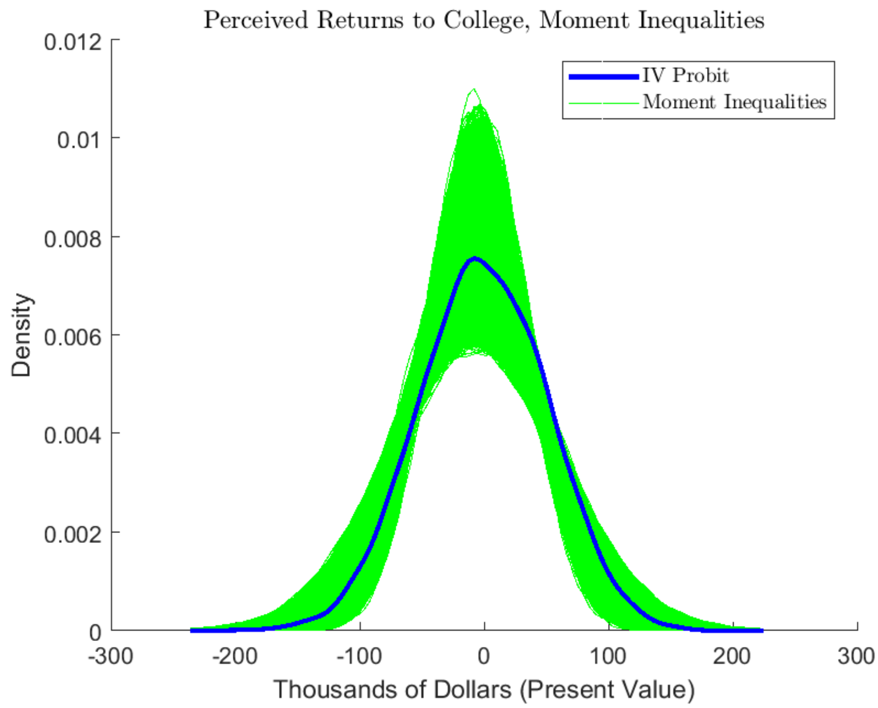

Figure 2: Perceived Returns to College, Moment Inequalities, Principal Components

Notes: Perceived returns across the population, weighted by 1988 sample weights, using principal components. Each point $\{\hat{\beta}_p, \hat{\sigma}_p\}$ in the confidence set (partially shown in figures 3 and 4) implies an entire distribution of beliefs about returns given by $X\hat{\beta}_p - Tuition\lambda\hat{\gamma}(X) + \mathcal{N}(0, \hat{\sigma}_p^2)$. I am unable to reject any of these implied distributions with 95% confidence. The distributions in blue are those with the lowest and highest values of $\hat{\sigma}$.

shown in Figure 3 for illustrative purposes. The first stage for the control function method is provided in F.[33]

Table 4: Perceived Returns Estimates, Unscaled, Principal Components

	(1) Probit	(2) IV Probit	(3) Moment Inequalities
Constant	-1.045	-1.407	[-11.331, 0.111]
	(0.080)	(0.082)	N/A
PC1(Ability)	0.354	0.573	[0.483, 4.443]
	(0.022)	(0.029)	N/A
PC2(Location)	-0.066	0.062	[-0.333, 1.661]
	(0.019)	(0.023)	N/A
Tuition	0.045	-0.047	[-0.476, -0.014]
	(0.005)	(0.010)	N/A
Observations	5492	5492	5492

Notes: Standard errors in parentheses. Parameters are marginal effects of the variable on perceived returns to college in standard deviations. For these results, I assume $\sigma = 1$. Estimates are from equation (11) with the details varying by estimation method. See text for details.

Turning to the comparison between the control function method and the moment inequalities, I note that the control function method point estimates fall within those of the moment inequality confidence sets. I further note that the bounds on the moment inequalities parameters are quite large, suggesting, for instance, that the standard deviation of the unobserved component of perceived returns to college is somewhere between $1,200 and $43,000. Because of the wide bounds on the moment inequalities and the suggestive evidence of the validity of the control function method for the purpose of this paper, I will focus on the control function method estimates for subsequent results and counterfactuals.

Because I find in this application that the control function method estimates are broadly consistent with the moment inequality estimates, I will use the full set of controls for the remaining analysis rather than the principal components. This is of interest for more precisely identifying the sources of variation in perceived returns to college. The control function method results when including all controls are presented in Table 5, with the corresponding visual representation of the distribution of perceived returns shown in Figure 4. The first stage and and unscaled estimates are provided in Appendix F.

I note first that the estimates of σ in the principal component specification and the full controls specification are statistically indistinguishable, suggesting that the two specifications give qualitatively similar results. The specification with full controls gives insight into which

[33] Recall that the moment inequalities make use of the instruments without estimating a first stage.

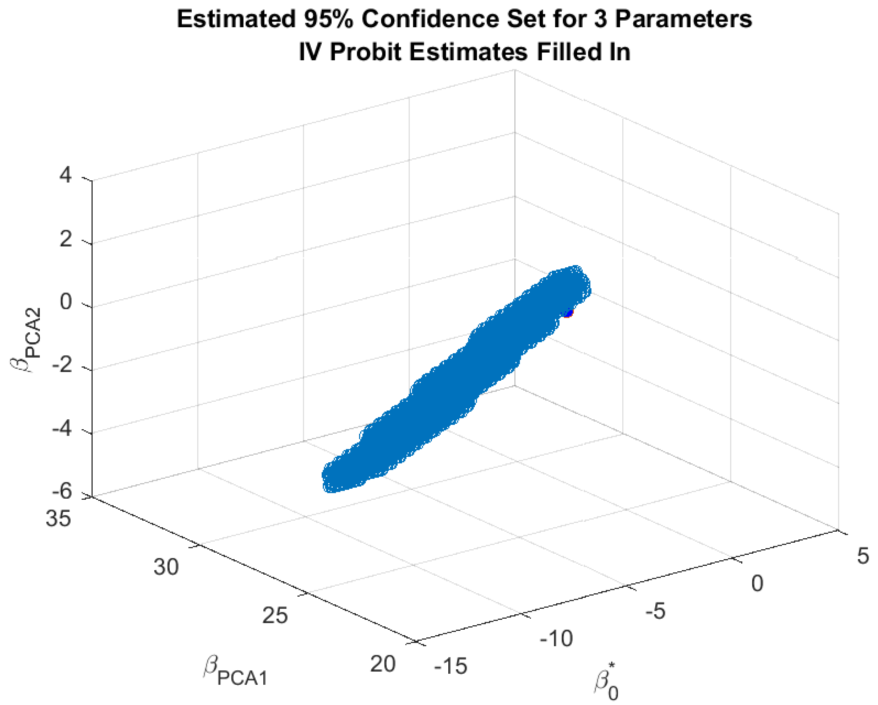

Figure 3: Unscaled Confidence Set for 3 Parameters, Moment Inequalities

Notes: The confidence set contains all combinations of parameter values that satisfy all of the moment inequalities in the unscaled discrete choice model. Note that the limits of the x, y, and z axes correspond to the results in Table 4, while the 95% confidence set is a subset of the 3-dimensional orthotope represented by these limits. The complete 95% confidence set is a 4-dimensional object. The scale transformation implied by the assumption on γ is performed on each point in the confidence set to obtain the results in Table 3.

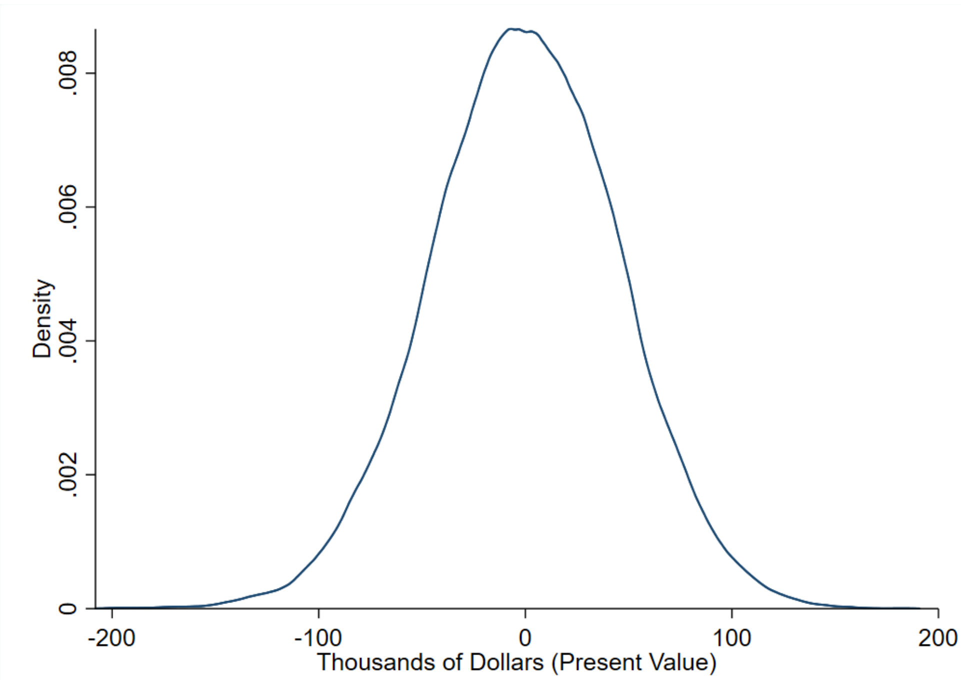

Figure 4: Perceived Returns to College, Control function method, All Controls

Notes: Perceived returns across the population, weighted by 1988 sample weights, for the full controls specification. The distribution is are given by $Y = X\hat{\beta} - Z\hat{\delta}\lambda\hat{\gamma}(X) + \mathcal{N}(0, \hat{\sigma}^2)$.

Table 5: Perceived Returns Estimates, 2018 Dollars, All Controls

	(1) Probit	(2) Std. Error	(3) IV Probit	(4) Std. Error
Constant	-2.107	(3.152)	-0.096	(1.042)
Female	-14.026	(7.384)	-6.681	(5.685)
Black	59.461	(8.540)	17.133	(17.066)
Hispanic	34.472	(9.554)	11.942	(12.620)
Deceased Father	-27.964	(44.749)	-19.777	(22.751)
Deceased Father x Before	82.788	(49.114)	40.634	(37.339)
Before	-5.321	(10.093)	-5.456	(7.166)
Senior Year	-28.145	(3.926)	-10.079	(9.288)
High School GPA	45.219	(18.734)	13.516	(11.192)
High School GPA Squared	2.383	(3.967)	0.794	(1.721)
Mother Education	-15.077	(4.322)	-2.819	(3.730)
Mother Education Squared	0.965	(0.208)	0.238	(0.262)
Father Education	-3.370	(3.677)	-1.208	(1.568)
Father Education Squared	0.413	(0.168)	0.139	(0.125)
Number of Siblings	-16.312	(3.282)	-4.308	(4.151)
Number of Siblings Squared	1.002	(0.296)	0.295	(0.283)
ASVAB Score Subtest 3	2.081	(1.102)	0.882	(0.861)
ASVAB Score Subtest 4	-0.405	(0.791)	0.054	(0.284)
ASVAB Score Subtest 5	-0.813	(0.767)	-0.099	(0.286)
ASVAB Score Subtest 6	1.036	(1.425)	0.401	(0.614)
ASVAB Score Subtest 7	-0.102	(0.404)	-0.192	(0.246)
ASVAB Score Subtest 8	-0.190	(0.271)	-0.032	(0.106)
ASVAB Score Subtest 9	-2.269	(1.020)	-0.817	(0.656)
ASVAB Score Subtest 10	2.751	(0.862)	0.532	(0.548)
ASVAB Score Subtest 11	-0.871	(0.977)	-0.177	(0.359)
ASVAB Score Subtest 12	0.933	(1.206)	0.411	(0.462)
ASVAB Score Subtest 3 Squared	-0.033	(0.141)	-0.034	(0.058)
ASVAB Score Subtest 4 Squared	0.156	(0.081)	0.049	(0.042)
ASVAB Score Subtest 5 Squared	-0.201	(0.063)	-0.054	(0.054)
ASVAB Score Subtest 6 Squared	-0.494	(0.277)	-0.170	(0.164)
ASVAB Score Subtest 7 Squared	-0.079	(0.029)	-0.015	(0.014)
ASVAB Score Subtest 8 Squared	0.002	(0.012)	-0.000	(0.004)
ASVAB Score Subtest 9 Squared	-0.390	(0.114)	-0.165	(0.148)
ASVAB Score Subtest 10 Squared	0.446	(0.101)	0.118	(0.111)
ASVAB Score Subtest 11 Squared	-0.008	(0.126)	-0.014	(0.045)
ASVAB Score Subtest 12 Squared	0.127	(0.184)	0.019	(0.069)
Family Income	0.000	(0.000)	0.000	(0.000)
Family Income Squared	-0.000	(0.000)	-0.000	(0.000)
Broken Home	-6.666	(6.499)	-1.553	(3.357)
Age 1979	-16.936	(4.242)	-5.902	(5.323)
Urban Residence at Age 14	19.175	(6.845)	5.621	(5.535)
Average County Wage at Age 17	-0.824	(1.224)	-0.028	(0.496)
State Unem Rate at Age 17	0.398	(1.695)	0.198	(0.564)
σ	106.208	(149.122)	31.328	(29.084)
Observations	3324	3324	3324	3324

Notes: All non-categorical variables are demeaned such that the constant gives the mean for white males. Parameters are marginal effects of the variable on perceived returns to college in thousands of dollars. The coefficient on tuition is assumed to be equal to $\lambda\hat{\gamma}(X)$. Estimates are from equation (11) with the details varying by estimation method. See text for details.

characteristics are associated with higher perceived returns as well as providing insight into the curvature of these characteristics. The results suggest that individuals with college-educated parents have about $30,000 higher perceived returns on average than those with parents who only completed high school(holding other variables at their means).[34]

The relationships between GPA and the various ASVAB scores are of further interest, as they are consistent with selection on gains in college attendance. For instance, GPA and the the first two ASVAB subtests on science and arithmetic predict high perceived returns. Meanwhile, ASVAB scores associated with nonacademic ability (such as subtests 7 and 9 on auto and shop information and mechanical comprehension) are associated with low perceived returns to college.[35] The negative relationship between subtests 5 and 6 (which measure word knowledge and paragraph comprehension) and perceived returns are also interesting in light of past findings of a negative relationship between verbal skills and wages, such as in Sanders (2015).

If my estimates of perceived returns are biased, it is likely that they overestimate the variance of the distribution. First, if local tuition at age 17 is associated with the unobserved component of perceived returns, it is likely to produce positive bias in estimates of the effect of tuition on attendance. This will happen if high local tuition is associated with higher perceived returns (i.e. people who live near elite universities expect their returns to college to be high, conditional on their other characteristics). I have attempted to account for this by including indicators of local labor market health. Second, If there is predictive power for actual tuition in local tuition at age 17 that is unknown to agents, estimates of the effect of tuition on attendance (the unscaled control function method or moment inequality estimates) will be biased toward zero if misperceptions are i.i.d. This is similar to the problems with assuming people know tuition perfectly, the predicted value for tuition will contain classical measurement error insofar as it is a measure of beliefs about tuition. Because the causal effect of perceived tuition on perceived returns should be negative, this bias moves the estimate of the effect of tuition on attendance in the positive direction.

Thus both likely sources of bias are positive, which would move the estimate of the effect of tuition on attendance closer to zero. Then applying the assumption on γ will produce upward bias in estimates of σ. In other words, this bias would cause me to conclude that tuition has a small effect relative to other factors, which would imply (because tuition is valued in dollars)

[34] Recall that the point at which a variable and its quadratic of opposite sign cross zero is given by $\beta_1 x + \beta_2 x^2 = 0 \implies x = \frac{-\beta_1}{\beta_2}$.

[35] Recall that the ASVAB tests have all been transformed to have unit variance and positive support.

that other factors have large effects in dollars. The effect of this is to overstate the variance of the distribution of perceived returns and to thus underestimate the effect of tuition subsidies/taxes on attendance.

6 External Validation and Policy Counterfactuals

Estimates of perceived returns are of interest to policymakers for identifying how many and what type of individuals value college at various levels. With this knowledge, it is possible to predict the number and type of individuals who will and will not attend college in the presence or absence of tuition subsidies or taxes. In this section, I will test the validity of the methodology of this paper by comparing the predicted effects of tuition subsidies on attendance from my estimates with those found in a natural experiment on Social Security Student Benefits studied by Dynarski (2003). Then, I will investigate the costs and effects of additional counterfactual policies.

6.1 Social Security Student Benefit

The Social Security Student Benefit was a policy from 1965 to 1982 that provided income assistance to children of deceased, disabled, or retired parents if they attended college. The financial reward was based on parental earnings, and was on average roughly $11,400 (2018 dollars) per year. This was sufficient to completely offset tuition costs for public institutions and to nearly do so even for many private institutions. Because this policy ended right as the individuals in the NLSY79 were deciding whether to attend college, this dataset was chosen by Dynarski (2003) to estimate the effects of the policy on educational outcomes including college attendance rates using differences in differences. I compare the implied effect of tuition aid on college attendance from the perceived returns I estimate to the results from her paper.

The primary result I attempt to match from Dynarski (2003) is the effect of the policy on attendance probabilities by age 23. Dynarski finds that the termination of this policy caused a 24.3% decrease in college attendance for the affected group, though these estimates were not significantly different from zero.[36] Assuming a linear effect of tuition on enrollment, she finds that a $1000 yearly subsidy caused a 3.6% increase in college attendance in year 2000 dollars.[37]

[36]Dynarski focused on the difference in attendance between children of deceased fathers before and after termination of the program. Fewer than 200 individuals in her data had deceased fathers, which likely contributed to the lack of significance despite the substantial point estimates.

[37]This amounts to a 2.1% increase in 2018 dollars using a 3% discount rate.

The validation exercise is thus to determine whether my estimates of perceived returns predict a 24.3% increase in enrollment from an \$45,600 (\$11,400/year x 4 years) subsidy. Because I use the same dataset to estimate perceived returns as Dynarski used to estimate the effects of the SSA Student Benefit, the results should be comparable. Because there was variation in benefits received, applying the \$45,600 uniformly across the population may somewhat misstate the policy effect according to the association between paternal death and perceived returns to college.[38] Finally, because Dynarski identifies the effect of student aid off of individuals with deceased fathers, my estimates of the effect of aid on the entire population will exceed hers if her treated group has lower responses to aid than average. The distribution of perceived returns implied by the estimates in Table 4 are shown in Figure 5 along with a counterfactual distribution showing the effect of a uniform \$45,600 subsidy to all potential college students. The predicted effect of the policy on attendance is given by the difference in the mass to the right of zero between the distributions. This effect is 26.0%, which is very close to the effect of 24.3% found by Dynarski (2003).

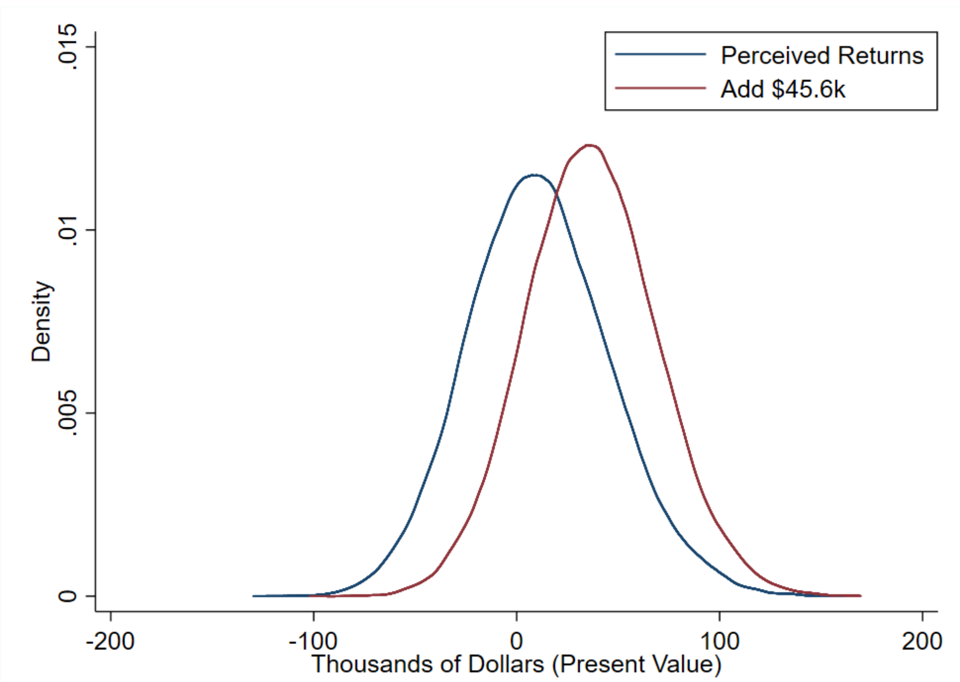

Figure 5: Effect of Universally Applied Aid Equivalent to Social Security Student Benefit

Notes: The shift in the presence of the policy comes from adding \$45,600 dollars to everyone, which is assumed to be well-publicized such that we have new perceived effective tuition for individual i given by $\widetilde{T}_i\hat{\gamma}(X_i) = \widetilde{T}_i\hat{\gamma}(X_i) + \$45,600\hat{\gamma}(X_i)$. The increase in mass just to the right of zero is the result of individuals with lower perceived returns paying a higher percentage (given by $\hat{\gamma}(X_i)$) of tuition than those with higher perceived returns, such that the tuition aid shifts them relatively further to the right. The shift visually looks smaller than \$45,600 because the average $\hat{\gamma}(X) = 0.61$.

[38]The NLSY79 does not have benefit amounts received, only parental mortality status. Dynarski used average benefits and data on parent mortality to infer the effect of benefit amounts.

An advantage of the methodology employed in this paper is that by obtaining the complete distribution of perceived returns, I do not rely on an assumption of a linear (or other) effect of tuition on attendance when computing effects of other counterfactual policies. For instance, the difference in differences methodology employed by Dynarski clearly identifies the effect of the \$11,400 annual tuition subsidy, but relies on a linear assumption on the effect of tuition is made to infer the effect of a \$1000 annual subsidy. Thus, while Dynarski infers a 2.1% effect of \$1,000 dollars (3.6% in year 2000 dollars), the predicted effect of a \$1,000 annual subsidy using the methodology employed in this paper is 2.6%. This larger effect is found because the average mass of the distribution of perceived returns is higher between \$0 and -\$4,000 than it is between \$0 and -\$45,600 dollars, such that the marginal effect of aid falls as aid rises.[39] I argue that the ability of the methodology described in this paper to closely match the results from a cleanly identified study of a natural experiment bodes extremely well for its validity and predictions for a wide variety of potential policies.

6.2 Attendance Target with Cost-Minimization

Given the external validity of the results as demonstrated above, it is possible to use my estimates of perceived returns to predict the effects of other potential policies. Here I describe the cost-minimizing policy that reaches a given attendance target, given the results above.[40] I choose $A = 75\%$ as the target level of college attendance, but any target could be chosen.

The cost-minimizing schedule of student aid conditional only on observables is shown in Figure 6. To derive it, I begin by noting the attendance probability for individual i, conditional on observables and financial aid offer a_i, is given by

$$Pr\left(S_i = 1|\hat{Y}_i, a_i\right) = \Phi\left(\frac{\hat{Y}_i + a_i\hat{\gamma}_i}{\hat{\sigma}}\right), \tag{34}$$

where $\hat{Y}_i = \mathbb{E}[Y_i|X_i, Tuition_i]$.[41] The expected cost to the government for this financial aid offer is then given by

$$\mathbb{E}[C_i|\hat{Y}_i, a_i] = a_i\Phi\left(\frac{\hat{Y}_i + a_i\hat{\gamma}_i}{\hat{\sigma}}\right), \tag{35}$$

where a_i is spent by the government on individual i only if they choose to attend college. Note

[39] It is worth noting that Dynarski (2003) suggested this possibility.

[40] Defining such a concrete target may be appealing to policymakers. For instance, President Obama specifically stated a goal of the U.S. having the highest proportion of college graduates in the world.

[41] Note that while the government spends a_i on individual i, the individual's return to college increases by $a_i\gamma$ because they pay γ proportion of their schooling costs.

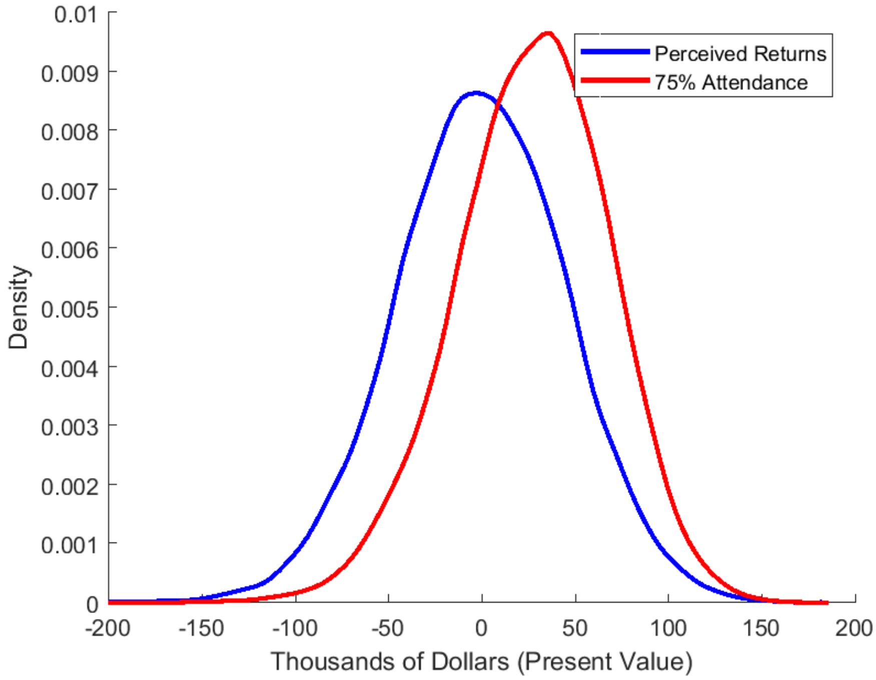

Figure 6: Cost-Minimizing Aid for Attendance Target

Notes: The red line shows perceived returns to college in the presence of the cost-minimizing policy that achieves the attendance target of 75%. This is the same proportion of the population that I predict will attend in the preceding section with the universally-applied subsidy of the same magnitude as the Social Security Student Benefit. The average cost per individual for that policy is $39,800, while the average cost in the cost-minimizing policy shown here is $29,400. Visual comparison of Figure 6 and Figure 7 show that the cost-minimizing policy shifts perceived returns to college less for individuals with high perceived returns than for those with low perceived returns.

that the government must pay a_i to person i even if they would have gone to college in the absence of the policy. Avoiding aid for individuals who are likely to go to college in the absence of aid will play an important role in the cost-minimization.

The attendance target implies that the government receives a constant marginal benefit, b, from any individual attending college.[42] Choosing a_i to set expected marginal benefit equal to expected marginal costs gives

$$b = \frac{\frac{\Phi(\hat{Y}_i^*(a_i))}{\phi_i(\hat{Y}_i^*(a_i))}\hat{\sigma} + a_i\hat{\gamma}_i}{\hat{\gamma}_i}, \tag{36}$$

wherein $\hat{Y}_i^*(a_i) = \frac{\hat{Y}_i + a_i\hat{\gamma}_i}{\hat{\sigma}}$ is the expected perceived return to college for individual i accounting for the financial aid offer and observables. I assume the government is constrained to use subsidies and not taxes ($a_i \geq 0\ \forall i$), which leads to the solution (given b) being the set $\{a_1\}_i$ that satisfies[43]

$$b_{1i} = \begin{cases} b & \text{if } b_{0i} < b, \\ b_{0i} & \text{if } b_{0i} \geq b. \end{cases} \tag{37}$$

where b_{1i} gives expected marginal cost per expected attendance for person i in the presence of the policy and b_{0i} is the same in the absence of the policy:

$$b_{0i} = \frac{\frac{\Phi(\hat{Y}_i^*(0))}{\phi_i(\hat{Y}_i^*(0))}\hat{\sigma}}{\hat{\gamma}_i}. \tag{38}$$

Essentially, this condition is that aid will be extended to those who respond most per cost for any attendance target above the initial attendance proportion. Note that b_{1i} is monotonically increasing in a_i, which implies that the single cutoff b will define marginal costs per marginal attendance for the treated group.[44]

The above gives the cost-minimizing idiosyncratic aid for each individual, $a_i(b)$, given an arbitrary cutoff value b. To reach attendance target A, all that remains is to find the value b^* that satisfies

$$\mathbb{E}\left[\Phi\left(\frac{\hat{Y}_i + a_i(b^*)\hat{\gamma}_i}{\hat{\sigma}}\right)\right] = A. \tag{39}$$

Then the cost-minimizing idiosyncratic aid is given by the a_i that solves (37).

[42] This formulation of the problem will generalize nicely to the case where the government has an idiosyncratic benefit, b_i, from individual i attending college, obtained for instance from estimates of lifetime returns to college.

[43] If the government can use taxes, the solution is is given by $b_{i1} = b\ \forall_i$.

[44] The condition that b_{1i} is monotonically increasing in a_i will be satisfied for any symmetric, log-concave distribution (such as the normal). This is a sufficient condition but not a necessary one, as $a_i\hat{\gamma}_i$ is increasing in a_i and will contribute to b_{1i} increasing in a_i.

The cost-minimizing financial aid solution has several interesting features. First, it focuses aid on individuals with low perceived returns. This happens because marginal increases in aid increase attendance by $\phi(\hat{Y}^*(a_i))$ while costing $\Phi(\hat{Y}^*(a_i))$, and the latter is large for large values of $\hat{Y}$ while the former is not. In other words, tuition subsidies for individuals with low perceived returns cause the government to spend less money on subsidies for people who would have attended college anyway. Secondly, it focuses aid on individuals who pay high percentages of their schooling costs, $\hat{\gamma}_i$. This is because the government must spend a_i to increase perceived returns by $a_i\hat{\gamma}_i$, which will be higher for high values of $\hat{\gamma}_i$. Thirdly, I note that individuals who have low perceived returns also tend to pay a high fraction of their educational costs, so these two types of people are really only one type of person. Many of these individuals will not respond to financial aid (because their perceived return is still below zero even in the presence of aid), keeping costs low for the government. Finally, such individuals that do respond will do so because they have high draws from the error term in their perceived returns (selection) equation. Carneiro, Heckman, and Vytlacil (2011) find that such individuals with high unobserved preferences for college also have relatively high real returns. Because this policy targets low socioeconomic status individuals who are likely to have relatively high returns while minimizing costs, it can likely serve as a useful heuristic for the government if it seeks to both reduce inequality and induce selection on gains. I conclude discussion of this policy by noting that its solution can easily be modified to provide optimal idiosyncratic financial aid conditional on known actual returns to college or to provide optimal aid conditional on a binding total financial aid budget constraint for the government.

7 Conclusions

I obtain estimates of beliefs about returns to college based on observed selection into college. Importantly, I am able to obtain these estimates without assuming that agents perfectly observe any data object that is known to the econometrician, and I only assume agent knowledge of the effect of tuition on returns. Prior research has made stronger assumptions about the information held by agents. The results suggest that 2.6% of individuals would be induced to attend to college with an annual tuition subsidy of only \$1000, which is consistent with the results from a host of studies of natural experiments.[45] Past estimates of the distribution of perceived returns

[45] It is common in this literature to provide effects of \$1,000 annual subsidies in year 2000 terms. This effect is 4.2%, while effects from 0%-6% are commonly found in studies of natural experiments, with the 0% estimates

such as those in Cunha, Heckman, and Navarro (2005) that are identified from assumptions on agents beliefs about real returns exhibit substantially higher variance and would be unable to predict the effects of tuition subsidies, though it is important to note that these authors do not claim to estimate compensating variation and do not claim to predict any such effects.

The methodology employed in this paper is especially well-suited to counterfactual policy analysis for multiple reasons. First, it avoids assuming that agents have rational expectations over returns to college. Second, it naturally identifies perceived returns in terms of compensating variation, which is directly applicable to policy questions. Third, if credit constraints are a factor, they will be seamlessly incorporated into the compensating variation specifically because they, like compensating variation, are linear in dollars where they exist. The effects of a tuition subsidy would then be to not only increase perceived returns at a constant marginal rate, but to reduce credit-constraints at a constant marginal rate. The predictions about which and how many individuals will be induced to attend college in the presence of any such policy will be identical whether we explicitly account for perceived credit constraints or not.

Past estimates of heterogeneous lifetime income returns to college commonly produce distributions of returns that have much higher mean and variance than the perceived return distribution that I estimate (See Cunha and Heckman (2007) for a survey of papers that estimate heterogeneous lifetime income returns). Cunha and Heckman (2016) more recently provides similar results for earnings from age 22 to 36 which are consistent with lifetime earnings that substantially exceed the perceived returns I estimate in both mean and variance. Average treatment effect estimates of wage returns to college are generally consistent with these estimates of lifetime earnings when making standard assumptions about hours worked per year and years worked.[46] The qualitative takeaway from this result is that individuals at best dramatically underestimate their returns to college while still making the attendance decision that will maximize their earnings (this will occur anytime the sign of an individual's actual return matches the sign of their perceived return, and at worst that they make a suboptimal decision due to underestimating the value of college relative to returns). Another way of describing the results is that individuals appear to dramatically overweight tuition costs relative to the other components of returns to college, an interpretation that appears consistent with reports in the popular

commonly attributed to administrative costs and/or information frictions associated with the policy. See Deming and Dynarski (2010) for a broad survey.

[46]See, for instance, Card (2001), Carneiro, Heckman, and Vytlacil (2011) and Heckman, Humphries, and Veramendi (2018)

press relating to concerns that the costs of college are considered prohibitively high for many individuals.

The methodology employed in this paper is well-suited for extensions in a variety of education decisions. These estimates are of potential interest for comparison to the analogous model of actual lifetime returns to college. Using a compatible specification for estimation of actual returns, with the same controls, the same imputation of tuition, and the same instruments, will produce the same estimates as those of perceived returns if agents have perfect foresight of their actual returns conditional on these variables. A test of perfect foresight in such a model is a joint test of all parameters being equal in the actual returns equation and the perceived returns equation. Similarly, upon performing such an analysis, it would be possible to identify predictors of misinformation as those variables with more divergent coefficients across the two equations. A comparison of perceived returns and actual returns is beyond the scope of this paper and is left for future work. The difference in estimates of the marginal effects of determinants of returns on actual returns and perceived returns will describe optimal schedules of tuition subsidies to induce selection on financial gains, with the caveat that such an exercise would ignore nonpecuniary private returns, externalities from education, and general equilibrium effects. Additional fruitful areas for future research include extensions of the methods above to college major choice (in a multinomial choice setting) or years of education (in an ordered choice setting).

In addition to education applications, the method described in this paper is well-suited to the estimation of perceived benefits for any purchase in which there are information frictions in pricing. One potential example is fertility decisions, in which pecuniary medical costs associated with childbirth are one of many components of the net benefits to childbearing, and could be used to identify the perceived valuation of having children despite not likely being perfectly forecast at the time of the childbearing decision. Another potential application is the perceived value of home ownership, especially in the context of adjustable rate mortgages, wherein the tuition ultimately paid for the home is again unforecastable at the time of purchase. Finally, as evidenced by the use of similar methodology in Dickstein and Morales (2018), it is clear that this method can be used to determine profit expectations of firms for a wide variety of potential investments such as export decisions, R&D, plant openings, and others.

References

ANDREWS, D. W., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

CARD, D. (1993): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," Discussion paper, National Bureau of Economic Research.

——— (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127–1160.

CARNEIRO, P., K. T. HANSEN, AND J. J. HECKMAN (2001): *Educational Attainment and Labor Market Outcomes: Estimating Distributions of the Returns to Educational Interventions.* Paper Presented As Klein Lecture, University of Pennsylvania, Philadelphia.

——— (2003): *Estimating Distributions of Treatment Effects with An Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College. No. W9546.* National Bureau of Economic Research.

CARNEIRO, P., J. J. HECKMAN, AND E. J. VYTLACIL (2011): "Estimating Marginal Returns to Education," *American Economic Review*, 101(6), 2754–81.

CUNHA, F., AND J. HECKMAN (2006): *The Evolution of Labor Earnings Risk in the Us Economy.* in 2006 Meeting Papers, No. 665. Society for Economic Dynamics.

——— (2016): "Decomposing Trends in Inequality in Earnings Into Forecastable and Uncertain Components," *Journal of Labor Economics*, 34(S2), S31–S65.

CUNHA, F., J. HECKMAN, AND S. NAVARRO (2005): "Separating Uncertainty from Heterogeneity in Life Cycle Earnings," *Oxford Economic Papers*, 57(2), 191–261.

CUNHA, F., AND J. J. HECKMAN (2007): "Identifying and Estimating the Distributions of Ex Post and Ex Ante Returns to Schooling," *Labour Economics*, 14(6), 870–893.

CUNHA, F., J. J. HECKMAN, AND S. NAVARRO (2006): "Counterfactual Analysis of Inequality and Social Mobility," *Mobility and Inequality: Frontiers of Research in Sociology and Economics :*, pp. 290–348.

DEMING, D., AND S. DYNARSKI (2010): "College Aid," in *Targeting Investments in Children: Fighting Poverty When Resources Are Limited*, pp. 283–302. University of Chicago Press.

Dickstein, M. J., and E. Morales (2018): "What do exporters know?," *The Quarterly Journal of Economics*, 133(4), 1753–1801.

Dynarski, S. M. (2003): "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion," *American Economic Review*, 93(1), 279–288.

Grodsky, E., and M. T. Jones (2007): "Real and Imagined Barriers to College Entry: Perceptions of Cost," *Social Science Research*, 36(2), 745–766.

Hansen, K. T., J. J. Heckman, and K. J. Mullen (2004): "The Effect of Schooling and Ability on Achievement Test Scores," *Journal of Econometrics*, 121(1-2), 39–98.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2018): "Returns to Education: the Causal Effects of Education on Earnings, Health, and Smoking," *Journal of Political Economy*, 126(S1), S197–S246.

Heckman, J. J., and S. Navarro (2007): "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*, 136(2), 341–396.

Iranzo, S., and G. Peri (2009): "Schooling Externalities, Technology, and Productivity: Theory and Evidence from Us States," *The Review of Economics and Statistics*, 91(2), 420–431.

Jensen, R. (2010): "The (Perceived) Returns to Education and the Demand for Schooling," *The Quarterly Journal of Economics*, 125(2), 515–548.

Kane, T. J., and C. E. Rouse (1995): "Labor-Market Returns to Two-and Four-Year College," *The American Economic Review*, 85(3), 600–614.

Manski, C. F. (1993): "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?," in *Studies of Supply and Demand in Higher Education, . University of Press*, pp. 43–60.

Manski, C. F. (2004): "Measuring Expectations," *Econometrica*, 72(5), 1329–1376.

Navarro, S. (2005): "Understanding Schooling: Using Observed Choices to Infer Agent's Information in a Dynamic Model of Schooling Choice When Consumption Allocation is Subject to Borrowing Constraints.'," Ph.D. thesis, University of Chicago Ph. D. Thesis.

Oreopoulos, P., and K. G. Salvanes (2011): "Priceless: The Nonpecuniary Benefits of Schooling," *Journal of Economic Perspectives*, 25(1), 159–84.

Rivers, D., and Q. H. Vuong (1988): "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models," *Journal of Econometrics*, 39(3), 347–366.

Roy, A. D. (1951): "Some Thoughts on the Distribution of Earnings," *Oxford Economic Papers*, 3(2), 135–146.

Sanders, C. (2015): "Reading Skills and Earnings: Why Do Doing Words Good Hurt You're Wages," Discussion paper, Working Paper.

Trautmann, S. T., and G. Van De Kuilen (2015): "Belief Elicitation: a Horse Race Among Truth Serums," *The Economic Journal*, 125(589), 2116–2135.

Wiswall, M., and B. Zafar (2015): "How Do College Students Respond to Public Information about Earnings?," *Journal of Human Capital*, 9(2), 117–169.

Zafar, B. (2011): "How Do College Students Form Expectations?," *Journal of Labor Economics*, 29(2), 301–348.

Appendix A: Moment Inequality Intuition

In this section I provide more discussion of the intuition of the moment inequalities used in this paper in the context of a simple example with simulated data. Suppose we have a simple discrete choice model wherein individuals select under decision rule (3)

$$S = 1 \iff Y \geq 0; \text{ else } S = 0 \tag{40}$$

and as in section 2 the perceived return Y is generated by a linear-in-parameters production function

$$Y = -2.5 + \tilde{x}_1\beta_1 + \epsilon_i, \tag{41}$$

which we assume for simplicity is a function of agent beliefs about a single variable x_1. With simulated data, we can clearly investigate what a standard Probit is able to achieve, what its limitations are, and how moment inequalities can improve upon it when there is imperfect information.

I generate data as follows, wherein $\beta_1 = -0.1$:

$$\begin{bmatrix} \tilde{x}_1 \\ z_1 \\ \epsilon_i \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ .5 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 & 3.8 & 0 \\ 3.8 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} \right).$$

Importantly, I generate the variable observed by the econometrician x_1 as a mean-preserving spread of the agents' expectations of it:

$$x_1 = \tilde{x}_1 + \mathcal{N}(0, 1). \tag{42}$$

Finally, I normalize the data to ensure that each variable has positive support.

First, we consider the realistically impossible setting in which we have access to the agent's information set such that we can make use of $\tilde{x}_1$ in our estimation procedure. When we run a simple Probit using

$$Pr(S_i = 1) = \Phi(\sigma^{-1}\tilde{x}_1\beta_1). \tag{43}$$

in which we include the agent's belief about x_1 in the estimation procedure, we unsurprisingly obtain an unbiased estimate $\widehat{\mathbb{E}[\beta_1} = 0.1957$. No normalization is required because the standard deviation of the error term has been set to 1. In this case the values of β_1 for which both inequalities are satisfied will be the exact parameter value identified by MLE, as seen in Figure 8.

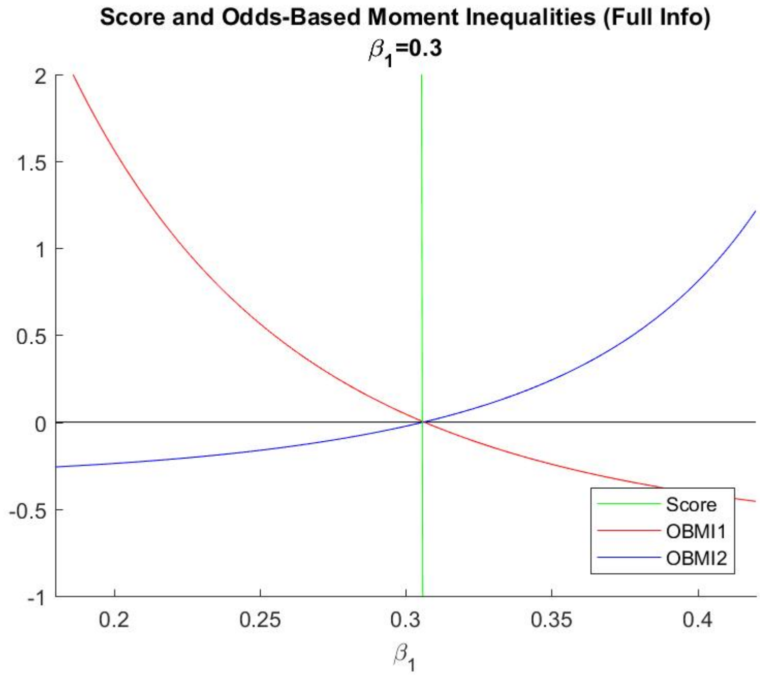

Figure 7: The score function from the MLE estimation perfectly intersects both of the odds-based moment inequalities, which are simply algebraic transformations of the score when we assume agents have perfect information on x_1. The point identified by MLE such that the score is equal to zero corresponds to the knife edge region in which both odds-based moment inequalities are satisfied only in the case of full information.

In the realistic setting in which we only observe x_1 and not $\tilde{x}_1$, the Probit will fail even in the absence of endogeneity. As discussed in section 2, we will essentially underestimate how elastic agents are to changes in x_1 because we will assume they are reacting to all of its variation when in reality they may only be reacting (more strongly) to only some of its variation. Erroneously performing a simple Probit

$$Pr(S_i = 1) = \Phi(\sigma^{-1} x_1 \beta_1) \tag{44}$$

produces the biased estimate $\hat{\beta}_1 = 0.0936$. If we use the moment inequality functions without conditioning on Z (such that they are equivalent to the score function) they still provide no additional benefit. However, when we condition on z_1, we essentially see both inequality curves from figure 1 shift upward such that a purple region emerges in which both inequalities are satisfied. This purple region contains the true parameter value $\beta_1 = 0.3$. This occurs because the distribution chosen is log-concave such that it has a convex odds-ratio. In other words, the

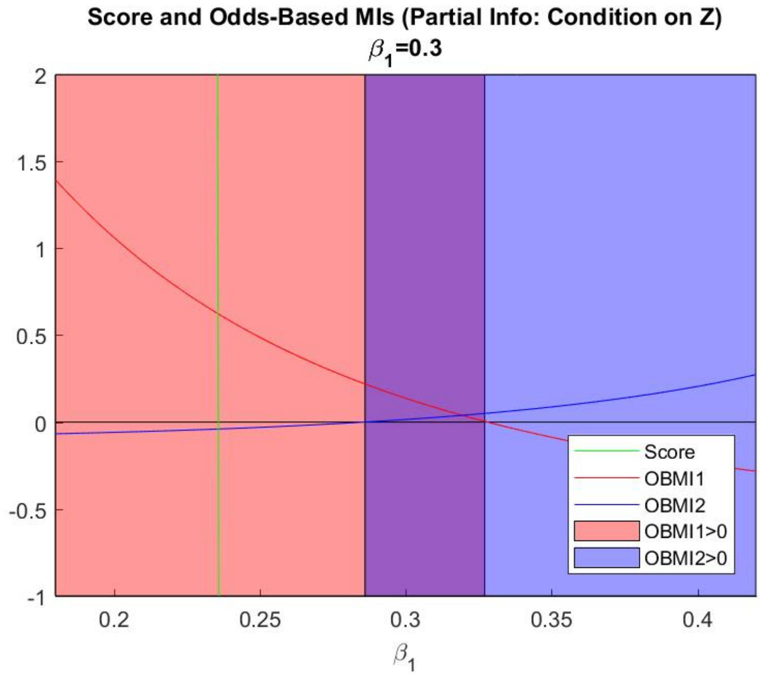

Figure 8: The estimates from MLE with a full information assumption (where the score function is equal to zero) suffers from attenuation bias. The red region designates the area where the first moment inequality is satisfied while the blue designates the area where the second inequality is satisfied. The overlapping region (purple) shows the area where both are satisfied and contains the true parameter value $\beta_1 = 0.3$.

noise in x_1 that is not contained in $\tilde{x}_1$ causes half of the $x_1 > \tilde{x}_1$ and the other half $x_1 < \tilde{x}_1$. Because of the shape of the normal distribution, the mean of $\frac{\Phi(x_1\beta_1)}{1-|phi(x_1\beta_1)}$ will be dominated by the values of x_1 that are greater than $\tilde{x}_1$.

Appendix B: Estimation of Moment Inequality Confidence Sets

I primarily follow Appendix A.5 and A.7 in Dickstein and Morales (2018) to estimate the moment inequality model. My method of evaluating a given point, also described in Andrews and Soares (2010), is the same that of Dickstein and Morales (2018). The primary difference arises in the grid search. I begin by briefly describing the intuition of the evaluation of parameters when estimating the moment inequality confidence sets.

Defining an error in this context as the deviation of a data moment from satisfaction of its inequality, the essential goal is to compare the sum of squared errors of the unconditional sample moments to what it would be under the null hypothesis that a given parameter vector is asymptotically consistent with the set of moment inequalities. This yields an intuitive test statistic that measures the the degree of violation of the ℓ moment inequalities for a given parameter vector:

$$Q(\psi_p^*) = \sum_\ell [min(\sqrt{N}\frac{\bar{m}_\ell(\psi_p^*)}{\hat{\sigma}_\ell}, 0)]^2, \tag{45}$$

where $\bar{m}_\ell(\psi_p^*)$ is the sample mean of the ℓth unconditional moment evaluated at ψ_p^*, and $\hat{\sigma}_\ell/\sqrt{N}$ is the estimated standard deviation of the ℓth unconditional moment.

Define $Q_a^n(\psi_p^*)$ as the asymptotic distribution of $Q(\psi_p^*)$ under the null hypothesis that $\mathbb{E}[m_\ell] = 0 \; \forall \; \ell$. If the value of $Q(\psi_p^*)$ obtained is less than the critical value defined at the αth percentile of $Q_a^n(\psi_p^*)$, then we will fail to reject that $\psi_p^* \in \Psi_0^*$. The distribution $Q_a^n(\psi_p^*)$ is thus sufficient to test this hypothesis. It is clear that distribution of the normalized moments at a given parameter vector is a multivariate normal with mean $\sqrt{N}\frac{\mathbb{E}[m(\psi_0^*)]}{\sigma_\ell}$ and variance $\Sigma_\psi(\psi_p^*)$ by central limit theorem. However, because the distribution of $Q_a^n(\psi_p^*)$ is that of the sum of ℓ squared truncated normals, it does not follow a known distribution. We can however obtain a simulated distribution $\hat{Q}_a^n(\psi_p^*)$ by generating R draws from the null distribution of normal moments with mean 0 and variance $\hat{\Sigma}_\psi(\psi_p^*)$. Each draw from this simulated distribution of moments provides a test statistic $Q_{ar}^n(\psi_p^*)$ resulting in a simulated distribution $\hat{Q}_a^n(\psi_p^*)$. Define the critical value at confidence level α $cv_\alpha(\psi_p^*)$ as the αth percentile of the simulated distribution $\hat{Q}_a^n(\psi_p^*)$. If the calculated test statistic in our sample $Q(\psi_p^*)$ is less than the critical value $cv_\alpha(\psi_p^*)$, the we fail to reject that the parameter vector ψ_p^* is within Ψ_0.

Regarding the algorithm for determining which points are within the confidence set, I will focus primary on distinctions between my estimation algorithm and those of DM. DM perform a brute force grid search on 3 parameters with a grid fineness of 40, producing $40^3 = 64,000$ points to evaluate. Because I evaluate 4 parameters when estimating the moment inequalities, I would need to evaluate $40^4 = 2,560,000$ points, dramatically more than DM. Initial attempts to achieve convergence with this method were unsuccessful. I augment the grid search algorithm in two ways. First, after making an initial grid to search in (by following the method DM use), I order these points in terms of their distance from the analogous Control function method estimates. Because the intuition for these two methods is very similar, I expect them to produce similar results. Second, once I find a feasible point, I abandon the grid search of all points and search locally around the successful point.

This second alteration essentially makes use of the continuity of the moment inequalities to avoid checking points that will not succeed. For instance, ceteris paribus, if the moment inequalities are satisfied at $\beta_0^* = 1$ and are not satisfied at $\beta_0^* = 2$, then this algorithm avoids checking $\beta_0^* = 3$. This essentially turns one extremely large grid search into a set of very small grid searches. When in the course of performing the grid search described by DM, a point in k-dimensional space space is found that cannot be rejected as satisfying the inequalities, I abandon the initial grid and instead check the k-dimensional hyper-rectangle defined by grid points that are 1 unit away from the unrejected point. I then repeat this procedure for all points that I fail to reject, and not for rejected points. This procedure allows me to find all unrejected points that are adjacent to other unrejected points, in a fraction of the time of searching the entire grid.

Appendix C: Principal Component Analysis

Here I provide estimates related to the principal component analysis mentioned in Section 4. The purpose of the principal component analysis is to reduce the parameter space sufficiently for the estimation algorithm described in Appendix B to converge in a timely fashion. I condense the controls listed in Table 2 into principal components according to the categorization in Table 6.

Table 6: List of Variables Included and Excluded in Principal Component Analysis

Variable Name	PC1 (Ability)	PC2 (Location)
ASVAB (All Tests)	✓	·
Mother's Education	✓	·
Mother's Education Squared	✓	·
Father's Education	✓	·
Father's Education Squared	✓	·
Number of Siblings	✓	·
Number of Siblings Squared	✓	·
High School GPA	✓	·
High School GPA Squared	✓	·
Bio Parents Home	✓	·
Urban at Age 14	·	✓
Average County Wage, Age 17	·	✓
State Unemployment, Age 17	·	✓

The loadings from the first principal component of each set of controls are provided in Table 7.

Table 7: Principal Component Loadings

	(1) PC1 (Ability)	(2) PC2 (Location)
Mother Education	0.192	
Mother Education Squared	0.193	
Father Education	0.206	
Father Education Squared	0.203	
Number of Siblings	-0.159	
Number of Siblings Squared	-0.149	
ASVAB Score Subtest 3	0.288	
ASVAB Score Subtest 3 Squared	0.021	
ASVAB Score Subtest 4	0.286	
ASVAB Score Subtest 4 Squared	0.088	
ASVAB Score Subtest 5	0.291	
ASVAB Score Subtest 5 Squared	-0.081	
ASVAB Score Subtest 6	0.264	
ASVAB Score Subtest 6 Squared	-0.089	
ASVAB Score Subtest 7	0.203	
ASVAB Score Subtest 7 Squared	-0.048	
ASVAB Score Subtest 8	0.175	
ASVAB Score Subtest 8 Squared	-0.039	
ASVAB Score Subtest 9	0.230	
ASVAB Score Subtest 9 Squared	0.061	
ASVAB Score Subtest 10	0.279	
ASVAB Score Subtest 10 Squared	0.103	
ASVAB Score Subtest 11	0.262	
ASVAB Score Subtest 11 Squared	0.093	
ASVAB Score Subtest 12	0.264	
ASVAB Score Subtest 12 Squared	0.058	
High School GPA	0.200	
High School GPA Squared	0.205	
Broken Home	-0.027	
Average County Wage at Age 17		0.707
State Unemp Rate at Age 17		0.566
Urban Residence at Age 14		0.424
Observations	5492	5492

Notes: Estimates are for the full NLSY79 sample. I use the first principal component from each set of variables in Table 6 to construct a measure of ability and local geographic characteristics.

Appendix D: Estimation of γ

In order to estimate $\hat{\gamma}(X)$ to obtain the perceived returns scaled in dollars, I use data from the NLSY79 on the percentage of college costs that students pay themselves. This information is only available in 1979. The raw data for observed values of γ are provide in Figure 9. Individual responses take one of four values. Students may report that they pay all, over half, less than half, or none of their educational expenses. I assign a value of 0.25% to those who report paying less than half, and a value of 0.75% to those who report paying more than half.

I estimate the following regression:

$$\gamma(X_i) = \frac{Tuition\_Paid_i}{Tuition_i} = X\beta_\gamma + \frac{X}{Tuition_i}\beta_{\gamma T}. \tag{46}$$

The terms divided by $Tuition_i$ will provide the effect of that component of X on the percentage of tuition paid, $\gamma(X)$. The terms that are not divided by $Tuition_i$ will provide the effect of that component of X on raw tuition. If for instance an individual's parents contribute $\$A + \$Tuition_i B$, the A will be caught by the terms not divided by zero, and should not be included in γ. Results from this regression are shown in Table 8. The imputed values for $\gamma(X)$ across the full sample are provided in Figure 10. Note that a few of these values exceed 1, which is conceptually interpretable as parents paying more than 100% of marginal tuition costs (parents provide in-kind benefits in excess of tuition, potentially as a reward for choosing a high quality, expensive college).

Appendix E: Imputation of Tuition

I impute tuition as described in section 4. I impute sticker tuition at college and scholarships separately and then combine them to produce net tuition. The results from the imputation are presented in Tables 9 and 10. I impute across all time periods in which I observe individuals because I use multiple years of tuition for single individuals to impute tuition conditional on observed characteristics.

Appendix F: Auxiliary Results

The unscaled results with all controls are provided in Table 11. The first stage for the Control function method is provided in Table 12. Recall that the F-stat on local tuition at age 17 should

Table 8: Effect on Percentage of Tuition Paid

	(1) Coef.	
inv_tuition	-68.58	(-1.35)
Mother Education	-0.00855	(-0.34)
Mother Education Squared	-0.000486	(-0.48)
Father Education	0.00793	(0.44)
Father Education Squared	-0.000827	(-1.18)
Number of Siblings	0.0821***	(4.84)
Number of Siblings Squared	-0.00427*	(-2.54)
ASVAB Score Subtest 3	-0.00553	(-0.21)
ASVAB Score Subtest 3 Squared	-0.0112	(-0.54)
ASVAB Score Subtest 4	-0.00822	(-0.35)
ASVAB Score Subtest 4 Squared	0.0149	(0.80)
ASVAB Score Subtest 5	-0.0614	(-1.72)
ASVAB Score Subtest 5 Squared	0.00409	(0.14)
ASVAB Score Subtest 6	0.0154	(0.48)
ASVAB Score Subtest 6 Squared	-0.0259	(-0.98)
ASVAB Score Subtest 7	0.0382	(1.75)
ASVAB Score Subtest 7 Squared	0.00391	(0.19)
ASVAB Score Subtest 8	-0.0240	(-1.28)
ASVAB Score Subtest 8 Squared	-0.00867	(-0.56)
ASVAB Score Subtest 9	0.00596	(0.28)
ASVAB Score Subtest 9 Squared	0.0127	(0.77)
ASVAB Score Subtest 10	0.00363	(0.16)
ASVAB Score Subtest 10 Squared	-0.0144	(-0.77)
ASVAB Score Subtest 11	-0.0306	(-1.38)
ASVAB Score Subtest 11 Squared	0.0121	(0.65)
ASVAB Score Subtest 12	0.0582*	(2.34)
ASVAB Score Subtest 12 Squared	0.0136	(0.75)
High School GPA	-0.0434	(-0.39)
High School GPA Squared	0.00926	(0.45)
Broken Home	0.179***	(5.72)
T_div_mhgc	5.655	(1.15)
T_div_mhgc2	-0.198	(-0.93)
T_div_fhgc	-0.0404	(-0.01)
T_div_fhgc2	0.00216	(0.01)
T_div_numsibs	-1.523	(-0.36)
T_div_numsibs_sq	-0.0777	(-0.18)
T_div_asvab3	2.681	(0.39)
T_div_asvab3_sq	-1.599	(-0.29)
T_div_asvab4	2.022	(0.56)
T_div_asvab4_sq	-1.125	(-0.22)
T_div_asvab5	8.127	(1.25)
T_div_asvab5_sq	6.543	(0.91)
T_div_asvab6	-6.804	(-1.27)
T_div_asvab6_sq	2.046	(0.57)
T_div_asvab7	0.841	(0.21)
T_div_asvab7_sq	-2.876	(-0.70)
T_div_asvab8	-2.670	(-0.72)
T_div_asvab8_sq	8.866*	(2.02)
T_div_asvab9	-5.246	(-1.35)
T_div_asvab9_sq	5.687	(1.89)
T_div_asvab10	-4.271	(-0.97)
T_div_asvab10_sq	-0.188	(-0.05)
T_div_asvab11	5.663	(1.23)
T_div_asvab11_sq	-3.355	(-0.84)
T_div_asvab12	0.613	(0.10)
T_div_asvab12_sq	-1.272	(-0.35)
T_div_GPA	29.60	(1.26)
T_div_GPA_sq	-5.105	(-1.29)
T_div_Broken_Home	-7.677	(-1.13)
Constant	0.568**	(2.85)
Observations	1113	

Notes: Standard errors are in parentheses. Estimates are for the part of the NLSY79 sample who attended college in 1979 and provided tuition and tuition paid information.

Table 9: Tuition Imputation

	(1) Coef.	
sticker_		
mhgc	-1936.1***	(-28.45)
mhgc2	116.4***	(35.40)
fhgc	-257.7***	(-4.93)
fhgc2	34.65***	(15.72)
numsibs	-1013.1***	(-19.83)
numsibs Squared	67.54***	(13.90)
asvab3	276.3***	(15.78)
asvab3 Squared	42.57***	(19.05)
asvab4	-84.28***	(-6.81)
asvab4 Squared	32.72***	(26.39)
asvab5	219.6***	(17.14)
asvab5 Squared	-14.40***	(-12.26)
asvab6	291.4***	(12.01)
asvab6 Squared	-37.87***	(-7.07)
asvab7	13.35*	(2.09)
asvab7 Squared	-9.771***	(-21.23)
asvab8	-11.90**	(-2.85)
asvab8 Squared	1.147***	(6.88)
asvab9	-304.5***	(-21.06)
asvab9 Squared	5.887***	(3.48)
asvab10	346.9***	(23.35)
asvab10 Squared	65.26***	(38.42)
asvab11	-6.806	(-0.45)
asvab11 Squared	-22.31***	(-12.12)
asvab12	-109.0***	(-5.82)
asvab12 Squared	17.92***	(6.21)
urban	657.7***	(5.48)
GPA	4307.5***	(10.77)
GPA Squared	-125.3	(-1.90)
c_wage_per_employed_age_17	1256.9***	(76.41)
unemployment_age_17	-298.0***	(-12.94)
local_tuition_17	7.613***	(77.89)
Constant	-9589.1***	(-9.75)
select		
College_age_14	0.182***	(28.30)
mhgc	-0.165***	(-44.67)
mhgc2	0.0108***	(57.28)
fhgc	-0.0228***	(-8.03)
fhgc2	0.00401***	(29.84)
numsibs	-0.0293***	(-27.45)
asvab3	0.00167	(1.73)
asvab3 Squared	0.000838***	(6.72)
asvab4	-0.0129***	(-18.90)
asvab4 Squared	0.00240***	(35.15)
asvab5	0.00335***	(5.00)
asvab5 Squared	-0.00186***	(-33.55)
asvab6	0.0205***	(16.33)
asvab6 Squared	-0.00924***	(-37.46)
asvab7	0.00138***	(3.91)
asvab7 Squared	-0.000467***	(-19.78)
asvab8	-0.00664***	(-29.24)
asvab8 Squared	-0.000127***	(-13.05)
asvab9	-0.0169***	(-21.49)
asvab9 Squared	-0.00115***	(-12.02)
asvab10	0.0232***	(29.98)
asvab10 Squared	0.00398***	(42.66)
asvab11	-0.0126***	(-15.20)
asvab11 Squared	0.000729***	(6.91)
asvab12	-0.0124***	(-11.95)
asvab12 Squared	0.00429***	(26.30)
urban	0.191***	(31.29)
GPA	0.599***	(168.74)
c_wage_per_employed_age_17	-0.00919***	(-9.65)
unemployment_age_17	0.0101***	(7.69)
local_tuition_17	-0.000112***	(-21.17)
Constant	-1.182***	(-50.34)
/mills		
lambda	11064.7***	(25.71)
Observations	352933	

Notes: Standard errors in parentheses. Parameters are marginal effects of the variable on college sticker price. See section 4 for details.

Table 10: Scholarship Imputation

	(1) Coef.	
NPV_scholarship_		
mhgc	-2409.9***	(-15.64)
mhgc2	116.0***	(14.76)
fhgc	1541.9***	(16.84)
fhgc2	-49.95***	(-13.74)
numsibs	554.6***	(9.32)
numsibs Squared	-30.59***	(-5.83)
asvab3	104.0***	(3.99)
asvab3 Squared	49.81***	(14.79)
asvab4	-351.2***	(-20.00)
asvab4 Squared	21.55***	(10.53)
asvab5	88.77***	(5.31)
asvab5 Squared	-3.781	(-1.81)
asvab6	354.4***	(9.80)
asvab6 Squared	-169.1***	(-13.75)
asvab7	70.91***	(8.17)
asvab7 Squared	-11.43***	(-17.72)
asvab8	-121.6***	(-15.95)
asvab8 Squared	-2.422***	(-8.58)
asvab9	-441.4***	(-11.08)
asvab9 Squared	-40.76***	(-15.65)
asvab10	122.3***	(6.22)
asvab10 Squared	63.24***	(13.61)
asvab11	-172.1***	(-6.10)
asvab11 Squared	28.30***	(11.05)
asvab12	-27.14	(-1.07)
asvab12 Squared	68.04***	(9.41)
urban	2303.7***	(13.57)
GPA	2144.5*	(2.56)
GPA Squared	1100.3***	(14.84)
c_wage_per_employed_age_17	195.3***	(8.28)
unemployment_age_17	-30.34	(-0.98)
local_tuition_17	1.674***	(12.97)
Constant	-28077.8***	(-8.02)
select		
College_age_14	0.0536***	(8.49)
mhgc	-0.119***	(-35.29)
mhgc2	0.00630***	(39.25)
fhgc	0.0502***	(18.44)
fhgc2	-0.00170***	(-14.16)
numsibs	0.00761***	(7.17)
asvab3	0.00977***	(10.35)
asvab3 Squared	0.00161***	(13.47)
asvab4	-0.00527***	(-8.01)
asvab4 Squared	0.00120***	(18.80)
asvab5	0.000356	(0.54)
asvab5 Squared	-0.00137***	(-24.66)
asvab6	0.0143***	(11.48)
asvab6 Squared	-0.00868***	(-34.52)
asvab7	-0.00120***	(-3.52)
asvab7 Squared	-0.000209***	(-8.96)
asvab8	-0.00479***	(-22.00)
asvab8 Squared	-0.000170***	(-18.60)
asvab9	-0.0302***	(-39.72)
asvab9 Squared	-0.00112***	(-12.08)
asvab10	0.00449***	(5.95)
asvab10 Squared	0.00373***	(43.27)
asvab11	-0.0177***	(-22.14)
asvab11 Squared	0.000764***	(7.59)
asvab12	0.00603***	(6.02)
asvab12 Squared	0.00562***	(36.46)
urban	0.0683***	(11.37)
GPA	0.502***	(142.33)
c_wage_per_employed_age_17	-0.00962***	(-10.39)
unemployment_age_17	-0.000505	(-0.40)
local_tuition_17	0.00000553	(1.08)
Constant	-1.580***	(-68.24)
/mills		
lambda	22726.9***	(13.44)
Observations	352933	

Notes: Standard errors in parentheses. Parameters are marginal effects of the variable on college sticker price. See section 4 for details.

be viewed as a lower bound on the strength of the instrument, as explained in Section 5. This value is 4629.13.

Table 11: Perceived Returns Estimates, Unscaled, All Controls

	(1) Probit	(2) Std. Error	(3) IV Probit	(4) Std. Error
Constant	0.475	(0.063)	0.565	(0.080)
Mother Education	-0.193	(0.031)	-0.191	(0.031)
Mother Education Squared	0.012	(0.002)	0.012	(0.002)
Father Education	-0.038	(0.024)	-0.031	(0.024)
Father Education Squared	0.005	(0.001)	0.005	(0.001)
Number of Siblings	-0.092	(0.023)	-0.082	(0.024)
Number of Siblings Squared	0.006	(0.002)	0.005	(0.002)
ASVAB Score Subtest 3	0.022	(0.035)	0.022	(0.035)
ASVAB Score Subtest 3 Squared	0.031	(0.026)	0.031	(0.026)
ASVAB Score Subtest 4	-0.027	(0.034)	-0.027	(0.034)
ASVAB Score Subtest 4 Squared	0.122	(0.027)	0.133	(0.028)
ASVAB Score Subtest 5	0.040	(0.038)	0.041	(0.038)
ASVAB Score Subtest 5 Squared	-0.144	(0.029)	-0.153	(0.030)
ASVAB Score Subtest 6	0.055	(0.032)	0.063	(0.032)
ASVAB Score Subtest 6 Squared	-0.122	(0.025)	-0.127	(0.025)
ASVAB Score Subtest 7	-0.016	(0.028)	-0.011	(0.028)
ASVAB Score Subtest 7 Squared	-0.077	(0.025)	-0.079	(0.025)
ASVAB Score Subtest 8	-0.083	(0.027)	-0.085	(0.027)
ASVAB Score Subtest 8 Squared	-0.050	(0.024)	-0.048	(0.024)
ASVAB Score Subtest 9	-0.123	(0.033)	-0.130	(0.034)
ASVAB Score Subtest 9 Squared	-0.045	(0.025)	-0.039	(0.025)
ASVAB Score Subtest 10	0.155	(0.033)	0.168	(0.034)
ASVAB Score Subtest 10 Squared	0.187	(0.026)	0.189	(0.026)
ASVAB Score Subtest 11	-0.096	(0.033)	-0.102	(0.033)
ASVAB Score Subtest 11 Squared	-0.010	(0.025)	-0.013	(0.025)
ASVAB Score Subtest 12	-0.050	(0.033)	-0.040	(0.034)
ASVAB Score Subtest 12 Squared	0.113	(0.025)	0.118	(0.025)
High School GPA	0.654	(0.117)	0.716	(0.122)
High School GPA Squared	-0.012	(0.026)	-0.022	(0.027)
Broken Home	-0.012	(0.044)	0.018	(0.047)
Urban Residence at Age 14	0.179	(0.048)	0.181	(0.048)
Average County Wage at Age 17	0.021	(0.009)	0.028	(0.010)
State Unemp Rate at Age 17	0.002	(0.010)	0.003	(0.010)
Tuition	-0.027	(0.008)	-0.040	(0.011)
Observations	5492	5492	5492	5492

Notes: Standard errors in parentheses. Parameters are marginal effects of the variable on perceived returns to college in thousands of dollars. The value for σ is assumed to be 1. Estimates are from equation (11) with the details varying by estimation method. See text for details.

Table 12: First Stage Estimates, Effect of Instruments on Tuition

	(1)	
	Coef.	
Mother Education	-0.529***	(-13.37)
Mother Education Squared	0.0225***	(11.88)
Father Education	0.255***	(8.28)
Father Education Squared	-0.00964***	(-6.94)
Number of Siblings	0.578***	(18.45)
Number of Siblings Squared	-0.0304***	(-11.05)
ASVAB Score Subtest 3	0.348***	(7.17)
ASVAB Score Subtest 3 Squared	0.193***	(5.47)
ASVAB Score Subtest 4	-0.184***	(-3.85)
ASVAB Score Subtest 4 Squared	0.913***	(26.46)
ASVAB Score Subtest 5	0.133*	(2.48)
ASVAB Score Subtest 5 Squared	-0.669***	(-16.63)
ASVAB Score Subtest 6	0.587***	(13.19)
ASVAB Score Subtest 6 Squared	-0.363***	(-10.36)
ASVAB Score Subtest 7	0.476***	(12.20)
ASVAB Score Subtest 7 Squared	-0.322***	(-9.62)
ASVAB Score Subtest 8	-0.272***	(-7.37)
ASVAB Score Subtest 8 Squared	0.0794*	(2.51)
ASVAB Score Subtest 9	-0.571***	(-12.60)
ASVAB Score Subtest 9 Squared	0.403***	(11.81)
ASVAB Score Subtest 10	0.911***	(20.27)
ASVAB Score Subtest 10 Squared	0.600***	(17.66)
ASVAB Score Subtest 11	-0.323***	(-7.10)
ASVAB Score Subtest 11 Squared	-0.292***	(-8.64)
ASVAB Score Subtest 12	0.583***	(12.75)
ASVAB Score Subtest 12 Squared	0.433***	(12.96)
High School GPA	3.649***	(26.38)
High School GPA Squared	-0.430***	(-13.86)
Broken Home	2.296***	(39.64)
Average County Wage at Age 17	0.672***	(66.25)
State Unemp Rate at Age 17	-0.176***	(-12.17)
Urban Residence at Age 14	0.212**	(3.13)
Local Tuition	0.00405***	(68.04)
Constant	-15.39***	(-48.53)
Observations	5492	

Notes: Standard errors in parentheses. Parameters are marginal effects of the variable on thousands of dollars of effective tuition ($Tuition_i\hat{\gamma}(X)\lambda$).