WORKING PAPERS

N° 1234

July 2021

"Estimating Coherency between Survey Data and Incentivized Experimental Data"

Christian Belzil, Julie Pernaudet and François Poinas

Toulouse School of Economics

Estimating Coherency between Survey Data and Incentivized Experimental Data*

Christian Belzil
CREST, CNRS, Paris Polytechnic Institute
and IZA and CIRANO

Julie Pernaudet
University of Chicago

François Poinas
Toulouse School of Economics,
University of Toulouse Capitole

July 13, 2021

*Belzil acknowledges financial support from Investissement d'Avenir (ANR-11-IDEX-0003/ Labex Ecodec/ANR-11-LABX-0047) and CIRANO. Poinas acknowledges funding from the H2020-MSCA-RISE project GEMCLIME 2020 GA No. 681228 and from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program). We thank Claude Montmarquette (CIRANO) and Jean-Pierre Voyer (SRDC), who were at the inception of the "Willingness to Borrow" project, for introducing us to this experiment and allowing us to use the data. We thank participants to the Workshop on Subjective Expectations, 2020, and seminar participants at Arizona State University. We also acknowledge discussion and/or comments from Esteban Aucejo, Teodora Boneva, Adeline Delavande, Christophe Gaillac, Marc Henry, John Rust and Basit Zafar. All data used in this paper may be obtained at no cost from any of the authors.

Abstract

Imagine the situation in which an econometrician can infer the distribution of welfare gains induced by the provision of higher education financial aid using survey data obtained from a set of individuals, and can estimate the same distribution using a highly incentivized field experiment in which the same set of individuals participated. In the experimental setting relying on incentivized choices, making the wrong decision can be costly. In the survey, the stakes are null and reporting false intentions and expectations is costless. In this paper, we evaluate the extent to which the decomposition of the two welfare gain distributions into latent factors are coherent. We find that individuals often put a much different weight to a specific set of determinants in the experiment and in the survey and that the valuations of financial aid are rank incoherent. About 66% of Biased Incoherency (defined as the tendency to have a higher valuation rank in the experiment than in the survey) is explained by individual heterogeneity in subjective benefits, costs and other factors and about half of these factors affect the welfare gains of financial aid in the survey and in the experiment in opposite directions. Ex-ante policy evaluation of a potential expansion of the Canadian higher education financial aid system may therefore depend heavily on whether or not the data have been obtained in an "incentivized" context.

JEL Codes: I2, C91, C93, D12, D9, D91.

1 Introductory Comments

In this paper, we make use of a unique data set that combines rich survey data on Canadian high school students' perceptions of the costs and benefits to higher education, cognitive and non-cognitive skills as well as financial and informational constraints with a high-stake field experiment that elicits their willingness-to-pay for financial aid. Exploiting these two data sources, we can estimate the distribution of welfare gains induced by financial aid and assess to what extent they differ depending on the measurement tool being used.

To obtain the first distribution, we estimate the subjective ex-ante valuation of higher education from survey data and use the parameters to construct counterfactual welfare gains induced by the provision of higher education financial aid offers identical to those found in a field experiment. Our analysis distinguishes between university (requiring 4 years) or other types of post-secondary education such as community college and the like (requiring 2 years).

To construct the model, we assume that some primitive latent factors (cognitive and non-cognitive skills, knowledge about the labor market and financial aid opportunities) are mapped onto qualitative perceptions about the cost and the benefits (both pecuniary and non-pecuniary) of higher education through a recursive factor structure. In turn, these components are mapped onto structural elements of the intertemporal utility of attending higher education to be estimated. Using out of sample information about life-cycle consumption of Canadians who have not attended higher education, we first estimate the expected utility of attending either type of higher education in the future under the status-quo (outside the experiment) using survey and recover the welfare gains induced by accepting hypothetical financial aid.

To obtain the second distribution, we use data taken from the field experiment to estimate directly the welfare gains of actual generous financial aid offers disclosed by a large number of binary decisions between immediate cash payments and financial aid for the same individuals.[1] It is then possible to measure the coherency between two candidate distributions originating from the same individuals.

In our opinion, this is a crucial undertaking which may deliver important implications for education policies. Financial aid expansions are regularly debated as a tool to promote access to higher education for all. However, little is known about the welfare gains induced by the provision of financial aid for students. One important challenge is the measurement of such gains. Suppose that a policy analyst is interested in evaluating the impact of an expansion in higher education financial aid on university vs. other post-secondary enrollments, and assume that he/she can access data from the survey or from the field experiment. Among other things, the analyst would want to know if those who are endowed with a high utility (or high return) to two-year college investment have a different valuation of financial aid than those with a propensity to choose a university curriculum, or to what extent individual valuations of financial aid may depend on individual latent factors. This could be achieved with the survey as well as with the experiment.[2] Ideally, conclusions drawn from one source of data should also be valid with the other as they

[1] In the terminology of Harrison and List (2004), the experiment can be categorized as an artefactual field experiment.

[2] As far as we know, economists have never estimated formally how financial aid generosity impacts on the odds ratio of choosing 2-year college vs. 4-year curriculum. While this is not the objective of our paper, decomposition of individual valuations in terms of subjective costs and benefits could be a useful step toward understanding who benefits from financial aid.

contain information and choices of the same individuals.

One major challenge that we face is to design an approach that circumvents potential scale differences in cardinal utilities between the survey and the experiment because valuations in the experiment are benchmarked with respect to the utility of an immediate cash payment and therefore depend on a reference consumption level which must be estimated. On the other hand, valuations obtained from the survey are about hypothetical future payoffs regardless of immediate consumption and are expressed in reference to a given lifecycle consumption profile which cannot be introduced explicitly in the model of individual choices taking place in the experiment.

One implication of this contextual difference between the survey and the experiment is that it is unrealistic to compare point estimates of the intertemporal valuation of additional financial aid measuring welfare gains. Instead of performing a simple evaluation strategy based on summary statistics such as the average of individual welfare gains, we search for the determinants of individual financial aid valuations and try to detect incoherencies in the mapping from type specific factors onto type specific valuations.

To achieve our goals, we must work with the mapping from individual ranks in costs, benefits and other factors onto financial aid valuation ranks. Despite the contextual differences aforementioned, if valuations obtained from the survey and the experiment are coherent, the mapping from factor ranks onto valuation ranks obtained from the survey should be the same as that obtained from the experiment.

As will become clear later, with our approach it is possible to distinguish between "Absolute" incoherency and "Biased" incoherency. The former is defined as the tendency for individual valuations to be ranked differently in the experiment than in the survey and is of little interest to us. The latter is defined as the propensity for individual valuations to be systematically ranked higher in the experiment than in the survey and is the measure of incoherency that we focus upon.

As we use latent factor techniques to isolate the key determinants of the valuation of financial aid in the survey and in the experiment, evaluating individual (or type) coherency requires lots of data. Survey data disclose information about student cognitive and non-cognitive skills, financial situations (liquidity constraint expectations), information about the labor market, and other relevant pieces of information. On top of this, students are asked to provide qualitative subjective valuations of monetary and non-monetary benefits and costs of education, as well as higher education attendance intentions (university or other types of post-secondary education). With these data, we can decompose the lifetime ex-ante utility of higher education into 6 dimensions: monetary costs, non-monetary (psychic) costs, monetary benefits, non-monetary benefits, a reference consumption level (depending on financial constraint and information about financial aid) and the consumption value of attending either type of higher education and easily characterize the ex-ante valuation of higher education under the status-quo.

The financial aid experiment consists of a sequence of 12 choices between a cash payment to be paid within one week and generous higher education financial aid conditional on future enrollment. It therefore varies both the cost of education by providing financial aid (loan or grant) ahead of enrollment and the opportunity cost of accepting financial aid. Choices exerted in the financial aid experiment identify the intertemporal gain of accepting financial aid (the difference between the futures components of two value functions), which depends itself on the same parameters characterizing the subjective returns to higher education.

One key distinction between survey data and the field experiment is the amount of incentives involved. Survey data contain self-reported elements obtained without providing any incentive as is commonly done in the literature.[3] On the contrary, the financial aid experiment is highly incentivized. Many decisions involve a choice between a high cash payment and a generous financial aid offer. For instance, some of the financial aid offers can cover between 25% and 50% of total 4-year university tuition at top universities in Canada, or close to 100% of the total 2-year community college tuition, while cash offers may be as high as $700. In the experimental setting, any decision made on the basis of false or unreliable parameters could turn out to be very costly as the immediate cash payments act as shadow prices. In the survey, disclosing abnormally high (or low) cost or benefit of higher education, or stating erroneous intentions would be costless.[4]

Although our primary objective is to evaluate how economic parameters estimated in a context where stakes are very large may differ from those obtained in an incentive-free setting (such as in survey data), our paper also contributes to various branches of the literature on education decisions and in particular, to the more recent literature on higher education financial aid. Those are reviewed in the next section.

The main results may be summarized as follows. Overall, we find relatively strong evidence of individual (or type) incoherency between information disclosed in the survey and choices exerted in the experiment when choices have a high opportunity cost. Beliefs and intentions revealed in the survey imply a very high valuation of higher education (especially university), very high hypothetical attendance probabilities, and a particularly important role played by non-financial dimensions. This implies that providing generous grants (reducing the cost of higher education) or generous loans remains ineffective at raising enrollments but raises welfare mostly by raising consumption during education. Indeed, an econometrician estimating a behavioral model from the survey could conclude that expanding financial aid programs would generate modest benefit over the lifecycle.

However, when estimating the valuation of financial aid from the experiment, we find a disconnection with individual valuations obtained from the survey. This is exemplified by the rank correlations between individual valuations of a given type of financial aid, which are generally low (between 0.3 and 0.4). This leads us to investigate the reasons explaining this low level of coherency.

Our investigation of the relationships between valuation ranks and factor ranks in the survey and in the experiment indicates that incoherency is not driven by purely stochastic components such as random errors independent from individual heterogeneity. Instead, incoherency is strongly correlated with heterogeneity measured by various latent factors. About 66% of Biased Incoherency (the tendency to have a higher valuation rank in the experiment than in the survey) is explained by individual heterogeneity in subjective benefits, costs and other factors. A precise decomposition of the individual valuations using rank-rank relationships indicates that about half of these factors affect the welfare gains of financial aid in the survey and in the experiment in opposite directions. This is particularly true about university monetary benefits, university monetary costs, post-secondary monetary benefits, post-secondary non-monetary costs, which appear to

[3]In the next section, we review the relevant literature.

[4]Another approach would be to treat the survey and the field experiment as complementary sources of information enabling identification of parameters otherwise unindentifiable. In recent years, a relatively large number of papers have made used of experimental data in order to perform structural estimation. As documented in Todd and Wolpin (2021), many of these papers exploit the advantages and ameliorate the disadvantages of each approach.

constitute incoherent determinants of both loan and grant valuations.

By grouping determinants according to different dimensions such as (i) monetary vs. non-monetary determinants, (ii) university costs and benefits vs. post-secondary costs and benefits, (iii) all benefits and costs vs other factors, and finally (iv) all benefits vs. all costs, it becomes relatively clear that individuals often put a much different weight to one set of determinants in the experiment and in the survey. For instance, we find that individual beliefs about university benefits and costs play a much higher role in the valuations inferred from the experiment than those obtained in the survey.

To the extent that one is willing to attach more reliability to data obtained in a high-stake context than in an incentive-free setting, our findings cast serious doubts about the validity of inference drawn from an incentive-free setting and indicate that ex-ante policy evaluation coming from it should also be questioned. To see this, reconsider the hypothetical case where an economist is interested in evaluating ex-ante the impact of an expansion in higher education financial aid on university vs other post-secondary enrollments such as 2-year colleges, but can only access data from the survey or from the field experiment. Estimating the determinants of individual valuations (essentially the willingness to pay for financial aid), could constitute a natural step as it should disclose if those who are willing to pay more are those with high ex-ante probabilities of attending 2-year colleges or with high ex-ante probability to enroll in a university.

When using the survey, our estimates indicate that those endowed with higher monetary benefits and costs of a university curriculum would benefit more from financial aid expansion. However, estimates obtained in the experiment point to the exact opposite, as those with higher cost and benefits of university attendance are those benefiting the less. While we recognize that predicting the impact of a new policy may require more information than the valuation decompositions into costs and benefits only, it is clear that the latter constitute valuable information. However, in this specific context, it would be impossible to reconcile the results obtained from the survey with those obtained in the experiment. As will become clear later, this sort of problem affects other determinants as well, and depending on which one the analyst is focusing upon, policy conclusions would depend on whether or not the data have been obtained in an "incentivized" context.

The remaining portion of the paper is organized as follows. Section 2 is devoted to the background literature and our contributions. A description of survey data and the field experiment are found in Section 3. The statistical model used for measuring individual factors is found in Section 4. In Section 5, we discuss the identification and estimation of the structural parameters of the model of the ex-ante value of higher education estimated from the survey, which is used to generate the distribution of financial aid valuations in the survey. In Section 6, we present the model explaining decisions exerted within the experiment and discuss the distribution of financial aid valuations. Section 7 is devoted to the confrontation of the two candidate distributions of the value of financial aid. Concluding remarks are found in Section 8.

2 Contribution and Background Literature

Our paper contributes to various branches of economics which study the reliability of economic preferences (or skills) obtained in contexts that differ according to the level of incentives involved. Although the literature is in its infancies, it already incorporates diverse applications and approaches which are difficult to unify. We now briefly review some of the most important.

First, there are instances where potential incoherencies may be explained by the conflict between disclosing attitudes (or opinions) that are in line with mainstream views (usually in surveys) and taking individual decisions that may be costly. In other words, individuals may not behave coherently when they do not have to back-up self-reported choices with actual actions. Many studies in environmental economics detected differences between stated valuations of prospective regulatory changes and actual individual behavior.[5] In our paper, we refer to this sort of incoherency, arising when hypothetical actions and actual choices are systematically different, as "Biased Incoherency".

The existence of individual incoherencies across contexts is also particularly acute when estimating deep preference parameters, and in particular, when estimating risk aversion. As many studies use answers to hypothetical situations involving risk (thereby potentially creating a potential "Hypothetical Bias") while others infer risk aversion from individual choices involving actual payoffs (mostly within a lab experiment), it is important to assess the extent to which both approaches lead to similar answers. Reviewing the literature on the estimation of risk aversion is beyond the scope of this paper, but Heckman, Jagelka, and Kautz (2019) discuss the notion of "hypothetical bias" and review various approaches and their link to quantitative psychology. As of now, there seems to be no consensus about the sign of the bias induced by choices exerted within hypothetical situations.

In consumer choice theory, individual incoherencies may be explained by erratic behavior caused by endogenous levels of attention in problem solving. Stochastic behavior has attracted attention in a recent literature concerned with the theoretical foundations of discrete choice models and economists have introduced concepts such as Rational Inattention (Matějka and McKay, 2015), Rational Imprecision (Steverson, Brandenburger, and Glimcher, 2019), or Limited Attention (Barseghyan, Molinari, and Thirkettle, 2021).[6] In line with these concepts, econometricians concerned with the structural modeling of the sources of stochastic behavior in the lab (such as reverting choices within a Multiple price List setting) and how it relates to economic incentives, have pointed out the need to separate parameters capturing noise and effort from true preferences.[7]

Finally, and as pointed out in Heckman, Jagelka, and Kautz (2019), a substantial number of researchers in Psychology and Economics have documented the existence of a strong dependence between achievements and incentives through the impact of incentive on effort. In a recent paper, Gneezy et al. (2019) show that US students improve performance on standardized tests substantially in response to incentives, while Chinese students do not seem to be affected. Their findings therefore strongly question the validity of international comparisons of education systems based on standardized test achievements.

In the current paper, we build on ideas found in these various segments of the literature, but use structural econometric techniques to uncover the nature and the determinants of incoherencies (if any) between the distribution of parameters obtained from the survey and from the experiment. To our knowledge, our paper is the first that uses structural methods in order to estimate the degree of coherency between economic parameters estimated from survey data obtained in an incentive-free environment and the same parameters obtained

[5]See Arrow et al. (1993) for a discussion of environmental policy regulations.

[6]The literature on behavioral inattention is summarized in Gabaix (2019).

[7]Many of these issues are detailed in Jagelka (2020), Belzil and Jagelka (2020) and Andersson et al. (2020).

from a highly incentivized field experiment.

Because our paper measures individual coherency from data on schooling subjective valuations and financial aid decisions, it also contributes to the literature on modeling education decisions. Our paper is the first to estimate subjective valuation of higher education using data in which both monetary and non-monetary costs and benefits are elicited using qualitative (ordered) measures. Indeed, the decomposition of the lifetime ex-ante utility of higher education into 6 components has no pendant in the literature since no study has ever been able to decompose the intertemporal utility of higher education into that many separate components.

While our approach makes use of data that are normally used by those using the elicitation approach, it differs substantially from papers found in the literature.[8] For instance, in Arcidiacono, Hotz, and Kang (2012) and Wiswall and Zafar (2015), monetary returns are elicited and non-monetary returns are inferred as residual parameters. In others, such as Boneva and Rauh (2019), non-pecuniary returns are elicited but the distinction between benefits and costs is ignored. The data used in this paper allow us to make use of a more disaggregated monetary and non-monetary motive measurements than is usually done in the literature as the measures are sub-divided in their cost and benefit dimension. In turn, those allow us to distinguish between higher education amenities reaped beyond graduation and amenities captured during education (the consumption value).[9]

As far as we know, our paper is also one of the first that treats self-reported qualitative valuations as noisy objects, thereby separating true individual beliefs from statistical noise, and maps these factors into structural parameters that admit a monetary interpretation. Until now, virtually all papers estimating models from elicited beliefs have abstracted from measurement errors.[10]

Finally, our paper contributes indirectly and to a lesser extent to the growing literature on the evaluation of higher education financial aid policies. As far as we know, the only structural paper concerned with higher education financial aid is Belzil, Maurel, and

[8]The first generation of schooling models (assuming Rational Expectation) were based on the premise that agents' subjective beliefs about the economic return to schooling coincide with the parameters that may be identified from observational data on schooling and earnings and the estimation therefore requires to use those parameters as input to the solution of the dynamic programming problem of the agent. This approach, followed in a seminal paper by Keane and Wolpin (1997), but also in several other contributions such as Eckstein and Wolpin (1999), Keane and Wolpin (2001), Belzil and Hansen (2002), Heckman and Navarro (2007) and, more recently, Todd and Zhang (2020), is surveyed in Belzil (2007). An alternative approach advocating the use of expectation survey data to estimate behavioral models has gained substantial popularity in recent years (see Manski, 2004). In the education literature, Arcidiacono, Hotz, and Kang (2012), Stinebrickner and Stinebrickner (2014) and Wiswall and Zafar (2015) have used subjective expectations in order to study college major choices while Kaufmann (2014), Delavande and Zafar (2019) and Boneva and Rauh (2019) investigated differences in motives (pecuniary vs non-pecuniary) for education.

[9]The determinants and expectations of US high school students is attracting much attention in the US. For instance, the University of Texas system and the Census Bureau expanded a feature the College Scorecard website showing the earnings payoffs of 37,459 majors at 4,434 colleges and universities in order to show how much money graduates earn, broken down by major and campus, in order to help future students make good choices. As of now, evidence suggests that individuals are not responsive to this sort of detailed information (Washington Post, December 25, 2020).

[10]Manski and Molinari (2010) use interval data modeling to characterize the potential consequences of rounding behavior in survey data. However, their approach does not seem to have been applied in empirical research. Cunha, Elo, and Culhane (2020) elicit mother's probabilities or age ranges that children will learn how to do specific developmental tasks and represent those measures with a latent factor structure and Attanasio, Cunha, and Jervis (2019) study the importance of maternal beliefs.

Sidibé (2021), who estimated semi-structurally the value of higher education financing aid from the same financial aid experiment that we use in the current paper. However, their focus is totally different. The authors are primarily concerned with measuring the impact of discount factors and risk aversion on financial aid decisions and they ignore most information contained in the student and parental surveys (they do not use measures of liquidity constraints, information, and cognitive and non-cognitive abilities). They do not relate the value of financial aid to structural parameters of the subjective valuation of higher education and they also ignore information on ex-ante enrollment probabilities.

3 The Field Experiment and Pre-Experiment Surveys

The experiment was funded by the Canada Millennium Foundation (a public enterprise created by the Canadian federal government) and was carried jointly by The Social Research and Demonstration Corporation (Ottawa) and the Centre Interuniversitaire de Recherche en Analyse des Organisations (Montreal) between October 2008 and March 2009. The sample, which consists of 1,248 full time Canadian students aged from 16-18 years, was drawn from both urban and rural sites across Canada, although our analysis is restricted to Ontarian and Manitoban high school students because the school curriculum is different in Québec (there is an intermediate step between high school and university).[11]

Our model is estimated from two distinct sources of data. The first element comprises two surveys: one administered to the students and another one to the parents. Both of them were administered before the experiment, which is the second source of data. One positive aspect of the experiment is that those recruited had no knowledge of the choices to be undertaken subsequently. This implies that answers found in the survey (especially those pertaining to higher education valuation and intentions) were not contaminated by potential (expected) changes in higher education costs.

We now describe both the survey and the experiment.

3.1 Student and Parental Surveys

At least one week before the experimental session at school, students were given a unique identifying number to complete an online survey. At the end of the survey, each student could select the session to which they were to participate. The aim of the survey questions was to obtain comprehensive profile of the participants and their family context. Many survey questions were adapted from some of the most well-known Canadian data sets: Youth in Transition Survey (YITS), Post Secondary Education Survey (PEPS) and Survey of Labour and Income Dynamics (SLID). They include measures of educational ambitions, expectations with regards to ambitions, perceived obstacles to pursuing post-secondary education, financial means at students' disposal, debt aversion, parents' education and parents' economic status. In addition, several other measurements were included to assess other attitudes and behaviors like inter-temporal orientation (planning ability), attitudes towards risk, aspiration level, engagement while in high school, perceptions of labour market conditions and perceptions of the cost of, and returns to post-secondary education.

[11]For an in-depth description of the project, see Johnson and Montmarquette (2015).

Parents were contacted by telephone through a telephone survey company, EKOS. School sign-up sheets were transmitted to EKOS from the schools. These included student name and parental phone numbers. They requested consent and conducted a very short survey. Parents were interviewed by telephone for basic income information, educational background and expectations concerning their child's educational achievement. The basic income and parents' educational background questions were also asked in the student survey.

3.1.1 Measurements of Unobserved Factors, Costs and Benefits of Higher Education and Higher Education Enrollment Intentions

The vast majority of the measurements that we use to identify the relevant factors are taken from the Student Survey. The questionnaires, along with the financial choice booklet, may be found in a file containing supplementary material, Section A, in which we provide a detailed description of each measurement used in order to estimate our model. Virtually all questions had to be answered within a discrete ordered choice structure such as : poor, good, very good, excellent, or strongly disagree, disagree, agree, or strongly agree. All other questions are coded as binary outcomes. We now present a synthetic summary.

- Primitive factors:
  - a cognitive skill factor identified from 8 measures and denoted CO,
  - a non-cognitive factor measuring motivation while in high school identified from 7 measures and denoted MS,
  - a non-cognitive factor measuring individual attitudes (the Locus of Control) identified from 7 measures and denoted LC,
  - a factor measuring information about the labor market identified from 5 measures and denoted MI,
  - a factor measuring the extent to which the student feels financially constrained, identified from 12 measures, and denoted FC.
- Subjective qualitative costs and benefits of higher education:
  - the monetary benefits of higher education identified from 4 measures for university and 8 measures for post-secondary and denoted MB^u and MB^{ps},
  - the non-monetary benefits of higher education identified from 8 measures for university and 16 for post-secondary, and denoted NMB^u and NMB^{ps},
  - the monetary costs of higher education identified from 4 measures for university and 8 for post-secondary, and denoted MC^u and MC^{ps},
  - the non-monetary costs of higher education identified from 7 measures for university and 14 measures for post-secondary, and denoted NMC^u and NMC^{ps}.

To ease notation, we use the following notation for each group of factors:

$$\begin{aligned} F &= \{CO, MS, LC, MI, FC\} \\ X^u &= \{MB^u, NMB^u, MC^u, NMC^u\} \\ X^{ps} &= \{MB^{ps}, NMB^{ps}, MC^{ps}, NMC^{ps}\} \\ X &= \{F, X^u, X^{ps}\}. \end{aligned}$$

In total, the primitive factors and the subjective costs and benefits are estimated from a total of 108 measurements, decomposed into 39 measurements for the primitive factors and 69 measurements of benefits and costs.

Finally, when estimating the behavioral model, we also make use of information relaying intentions to enroll in higher education. Precisely, we use answers to two questions measuring enrollment intentions:[12]

- Measure 1: "As things stand now, what is the highest level of education you think you will get?" The responses can take 3 modalities: university degree, other post-secondary education, and no further education.

- Measure 2: "What is the highest level of education you would like to get?" The responses can take 3 modalities: university degree, other post-secondary education, and no further education.

3.2 Choices between Financial Aid and Cash Payments

The experiment was conducted using pen and paper choice booklets and simple random sampling devices like bingo balls and dice. Urban and rural school districts were selected in each of the four provinces and the implementation team was able to carry out work in urban and rural schools.

All subjects were presented with the full set of decisions and are paid for one, randomly selected, at the end of the session. The subjects were informed that they would be paid for one decision, but they did not know which one at the beginning of the session. The questions can be split into two main groups.

First, young individuals must answer a set of questions aimed at measuring their preferences for risk and time. The design of the questions is based on standard multiple-list approaches commonly used in experimental economics.[13] The second portion, which constitutes the main originality of the experiment, is entirely devoted to individual decisions between cash payments and higher education financial aid.

The financial aid packages are offered conditional on entering post-secondary education and comprise various combinations of loans offered at the prevailing market rate of interest and grants. Financial aid packages offered over the experiment were very generous. They varied between \$500 and \$4000 and represented a potentially high fraction of a yearly tuition at any Canadian university. Over the period considered, a grant of \$2,000 in 2008 would have covered 65% of total fees at University of Western Ontario, University

[12] In the student survey, those questions are D19 and D20.

[13] Belzil, Maurel, and Sidibé (2021) use the experiment to estimate the value of higher education financial aid and to measure the relative importance of deep preference parameters (preferences for risk and time) in explaining the value of higher education.

of Toronto and Queen's University. Those universities represent the most prestigious universities in Ontario.[14]

In the paper, we make use of 12 choices between a financial aid offer and an immediate cash payment. Cash payments varied between $25 and $700 although 9 of the 12 choices that we model incorporated either a $300 or a $700 payment. Before the start of the decision process, students were presented with the following definitions:

Grant: "Educational grants will be disbursed if a participant enrolls in an institution for learning or training full time within two years from the date of experiment participation. The grant will cover direct and indirect costs related to the learning activity. For tuition fees, payments will be made directly to the education institution. Receipts will be required for the reimbursement of other costs."

Loan: "Educational loans will be disbursed if a participant enrolls in an institution for learning or training full time. These loans will be available up to two years from the date of the experiment. The loans are repayable upon the completion of the study or if the participant drops out of the program of study. The interest rate, which is the same as the one offered by the Canadian Federal Student Assistance program, is floating and is set at the prime rate (3.2% on average over the period of interest) plus 2.5%."

It is important to note that interest rates attached to loans are also practically identical to those provided by charter banks within private loans. Indeed, over the period considered, private loan interest rates were within a 1% range of those offered within the public system. However, private loans accounted for about 7% of total resources devoted to higher education financing.[15]

All 12 decisions are detailed in Table 1. In total, there are 3 different types of offers considered in our analysis.

- Choices 1 to 5: The first 5 choices are between cash payments and classical loans. The loans vary between $1000 and $4000 while the cash payments vary between $25 and $700.
- Choices 6 to 12: In the last 7 choices, students must choose between cash payments and grants equal to $500 (choice 10), $1000 (choices 6, 7, 8 and 9), $2000 (choice 11) and $4000 (choice 12).

The last column of Table 1 documents the fraction of Ontarian and Manitobans high school students who accepted financial aid for each question (the take-up rates). First, it should be noted that except for the case where the cash payment is minuscule ($25), standard loans are chosen by only a minority of students. For grants, we observed the opposite. Except for choices 9 and 10, where cash payments ($700 and $300) are close to the amounts of the grants ($1,000 and $500), a majority of young individuals tend to choose grants.

[Table 1 here]

[14] The average tuitions was equal to $5667 in Ontario over the period covered by the experiment.

[15] For more details on higher education financing, see Berger, Motte, and Parkin (2007).

4 Measuring Primitive Factors and Motives for Higher Education

In this section, we provide details about the estimation of factors. First, we assume the existence of individual primitive factors which include cognitive ability, motivation in school, the locus of control, one factor measuring knowledge about the labor market and the financial aid system, and last one measuring liquidity constraints perceptions. These 5 individual factors are then mapped onto four major components that summarize individual perceptions about the likelihood to attend higher education in the future. Those components, which are monetary benefits, non-monetary benefits, monetary costs and non-monetary costs, represent different motives for attending higher education.

4.1 Student Skills, Liquidity Constraint Perceptions and Market Information

The vector of primitive factors is denoted F_i and incorporates CO_i, MS_i, LC_i, MI_i and FC_i. Those factors are meant to represent individual skills, information and liquidity constraints perceptions. To ease notation, we use f_i as a generic term denoting any single factor belonging to F_i.

For each of the $f's$, we observe a specific set of dedicated measurements taken from the student survey and the parental survey. The jth measurement (for individual i) devoted to factor f is denoted M_i^{fj}. The probability that the measure M_i^{fj} takes the modality m_i^{fj} depends on the nature of the recording (continuous or discrete). We now detail the contribution to the likelihood for each specific type of recording.

4.1.1 Case 1: Discrete Measurements

For discrete (binary) measurements, we assume the existence of a latent variable, $\tilde{M}^{fj}$, such that

$$\tilde{M}_i^{fj} = m_0^{fj} + m_1^{fj} \cdot f_i + \zeta_i^{fj} ,$$

where the $\zeta_i^{fj\prime}s$ represent measurement error shocks and are assumed to follow a Normal distribution with mean 0 and standard deviation 1. Measurement errors are independent across individuals and measures and m_0^{fj} and m_1^{fj} are parameters to be estimated. Their identification is addressed below.

The contribution to the likelihood of measurement M_i^{fj} (conditional on the factor) is

$$\Pr(M_i^{fj} = m_i^{fj}) = \Phi(m_0^{fj} + m_1^{fj} \cdot f_i)^{m_i^{fj}} \cdot (1 - \Phi(m_0^{jf} + m_1^{jf} \cdot f_i))^{1-m_i^{fj}} ,$$

where $\Phi(\cdot)$ is the standard Normal cumulative distribution function. When measurements take more than 2 values, we treat them as continuous measurements.

4.1.2 Case 2: Continuous Measurements

Some of the measurements used in our analysis (for instance, the Numeracy Test score) are recorded as continuous variables. To avoid an unreasonable number of parameters, we also treat multinomial qualitative measurements as continuous. In this case, the measurement equation is linear:

$$m_i^{fj} = m_0^{fj} + m_1^{fj} \cdot f_i + \zeta_i^{fj} ,$$

where m_0^{fj} and m_1^{fj} are parameters to be estimated and ζ_i^{fj} is the measurement error term which is assumed Normal with mean 0 and standard deviation σ_{fj}. The contribution to the likelihood is simply

$$\Pr(M_i^{fj} = m_i^{fj}) = \frac{1}{\sigma_{fj}} \cdot \phi\left(\frac{m_i^{fj} - m_0^{fj} - m_1^{fj} \cdot f_i}{\sigma_{fj}}\right),$$

where $\phi(\cdot)$ denotes the standard Normal density.

4.2 Subjective Monetary and Non-monetary Benefits and Costs of Higher Education

Cognitive and non-cognitive skills, as well as information about the labor market are mapped onto individual beliefs about both monetary and non-monetary benefits (denoted MB^e and NMB^e respectively) and monetary and non-monetary costs (denoted MC^e and NMC^e) of higher education of type e. We express them as follows:

$$\begin{aligned} MB_i^e &= \delta_0^{MB,e} + \delta_{CO}^{MB,e} \cdot CO_i + \delta_{MS}^{MB,e} \cdot MS_i + \delta_{LC}^{MB,e} \cdot LC_i + \\ &\quad \delta_{MI}^{MB,e} \cdot MI_i + \xi_i^{MB,e} \quad \text{for } e = u, ps \end{aligned}$$

$$\begin{aligned} NMB_i^e &= \delta_0^{NMB,e} + \delta_{CO}^{NMB,e} \cdot CO_i + \delta_{MS}^{NMB,e} \cdot MS_i + \delta_{LC}^{NMB,e} \cdot LC_i + \\ &\quad \delta_{MI}^{NMB,e} \cdot MI_i + \xi_i^{NMB,e} \quad \text{for } e = u, ps \end{aligned}$$

$$\begin{aligned} MC_i^e &= \delta_0^{MC,e} + \delta_{CO}^{MC,e} \cdot CO_i + \delta_{MS}^{MC,e} \cdot MS_i + \delta_{LC}^{MC,e} \cdot LC_i + \\ &\quad \delta_{MI}^{MC,e} \cdot MI_i + \xi_i^{MC,e} \quad \text{for } e = u, ps \end{aligned}$$

$$\begin{aligned} NMC_i^e &= \delta_0^{NMC,e} + \delta_{CO}^{NMC,e} \cdot CO_i + \delta_{MS}^{NMC,e} \cdot MS_i + \delta_{LC}^{NMC,e} \cdot LC_i + \\ &\quad \delta_{MI}^{NMC,e} \cdot MI_i + \xi_i^{NMC,e} \quad \text{for } e = u, ps\,, \end{aligned}$$

where $\delta_0^{MB,e}, \delta_{CO}^{MB,e}, \ldots, \delta_{MI}^{NMC,e}$ are parameters measuring the effect of individual primitive factors on benefit and cost perceptions and $\xi_i^{MB,e}$, $\xi_i^{NMB,e}$, $\xi_i^{MC,e}$ and $\xi_i^{NMC,e}$ represent individual specific components of non-benefits and costs which may not be accounted for by primitive factors. This means that all four components of the intertemporal utility of attending school are arbitrarily correlated with each other through the recursive effects of the primitive factors. We exclude the financial constraint factor (FC) from the costs and benefits because it plays a direct role in the model as indicated below.

The measurements for subjective benefits and costs are denoted $m^{MB,e,j}$, $m^{NMB,e,j}$, $m^{MC,e,j}$ and $m^{NMC,e,j}$ and are all discrete ordered. As for the primitive factors, they are treated as continuous variables. The measurements equations are written as follows:

$$m_i^{MB,e,j} = m_0^{MB,e,j} + m_1^{MB,e,j} \cdot MB_i^e + \zeta_i^{MB,e,j}$$

$$m_i^{NMB,e,j} = m_0^{NMB,e,j} + m_1^{MB,e,j} \cdot NMB_i^e + \zeta_i^{NMB,e,j}$$

$$m_i^{MC,e,j} = m_0^{MC,e,j} + m_1^{MC,e,,j} \cdot MC_i^e + \zeta_i^{MC,e,j}$$

$$m_i^{NMC,e,j} = m_0^{NMC,e,j} + m_1^{NMC,e,j} \cdot NMC_i^e + \zeta_i^{NMC,e,j} ,$$

where the m_0's and m_1's are parameters to be estimated and $\zeta_i^{MB,e,j}$, $\zeta_i^{NMB,e,j}$, $\zeta_i^{MC,e,j}$ and $\zeta_i^{NMC,e,j}$ are measurement error shocks which follow a Normal distribution with mean 0 and standard deviation equal to $\sigma_{MB,e,j}$, $\sigma_{NMB,e,j}$, $\sigma_{MC,e,j}$ and $\sigma_{NMC,e,j}$ respectively.

More details regarding estimation are found below in conjunction with the estimation of the model in the next section. Empirical estimates will also be discussed in the next section but many results will be found in a supplementary file to ease presentation.

5 Modeling the Ex-Ante Value of Higher Education and Financial Aid from Survey Data

5.1 The Model

To fulfill our objectives, we need a model of higher education attendance under the status-quo (outside the experiment) which is meant to reflect individual intentions revealed in the survey. We refer to period 0 as the interval (about 6-month) between the time when individuals participate into both the survey and the experiment and the time when they would actually decide to enter higher education and assume that actual enrollment decisions would take place at the beginning of period 1. All periods subsequent to period 0 actually last one year. We set a terminal date at period 20 (corresponding to age 38 to 40 for most individuals).

It is understood that entering university (superscript u), entering other post-secondary institution (superscript ps) or taking the outside option (superscript a) can only happen in period 1 and for this reason, there is no need for time subscripts for the terms defining the following intertemporal utilities:

- $\bar{V}_{i,sq}^{u}$: expected lifetime utility of student i entering university in period 1 under the status-quo (subscript sq), using information available at the beginning of period 0 (survey time).
- $\bar{V}_{i,sq}^{ps}$: expected lifetime utility of student i entering other type of post-secondary institution in period 1 under the status-quo, using information available at the beginning of period 0 (survey time).
- $\bar{V}_{i,sq}^{a}$: expected lifetime utility of student i of choosing the alternative option (not to enter university) in period 1 under the status-quo using information available in period 0.

5.1.1 The Alternative Option

As neither the experiment nor the surveys provide data on life-cycle consumption, we set the consumption level of the alternative option (to leave education after high school completion) to an estimate of the average yearly net income of high school or post-high school certificate obtained from an official publication of the Analytical Studies Branch Research of Statistics Canada (Frenette, 2014). This leads us to a yearly consumption reference level set to \$34,000. The expected lifetime utility of the alternative option, $\bar{V}_{i,sq}^{a}$, is therefore interpreted as the discounted sum of an average flat consumption profile over that period. It is equal to

$$\bar{V}^a_{i,sq} = E\left\{\sum_{t=1}^{20} \beta^{t-1} \cdot (\log \tilde{c}_i + \varepsilon^a_{it})\right\}$$

where ε^a_{it} is an i.i.d. mean 0 disturbance term affecting the valuation of the alternative option (realized at survey time) and $\beta = \frac{1}{1+0.05}$.

5.1.2 The Value of Higher Education

For the expected lifetime utility of going to university, $\bar{V}^u_{i,sq}$, we split the forthcoming periods into two parts: one college episode lasting 4 years and a post-education episode made of 16 years of employment. For other post-secondary studies, $\bar{V}^{ps}_{i,sq}$, we assume 2 years of education (as is the case for most community college programs) and 18 years of labor market employment.

As we neither observe the consumption level achieved in school nor in the labor market, our estimate is a generic term that may incorporate individual beliefs about financial and non-financial rewards net of borrowing plans that are not revealed in the survey. For this reason, we cannot measure subjective returns to education as defined in standard Mincerian or Beckerian models.

If we could access data on pecuniary and non-pecuniary net benefits of education only, differentiating between consumption while in school and post-education consumption would be tenuous. However, as we are able to dissociate monetary costs and non-monetary costs from both monetary and non-monetary benefits, we can use this information (along with other restrictions) to separate post-education consumption from consumption while in school.

In order to reduce the burden of notation, we rely again on the generic superscript e which applies to university (u) and post-secondary education (ps), and present one set of equation only when possible. From now on, we use small letters to represent structural parameters representing benefits and costs in monetary units (or utility) in order to distinguish them from the qualitative costs and benefit factors identified from survey data (in Section 4). The structural components are therefore denoted mc^u, nmc^u, mb^u and nmb^u for university education and mc^{ps}, nmc^{ps}, mb^{ps} and nmb^{ps} for other types of post-secondary education.

To facilitate comprehension, it is convenient to distinguish between the levels of consumption that would prevail under the status-quo (characterizing beliefs when the survey takes place in period 0), which we denote by $c^e_{i,sq}$ and $c^{m,e}_{i,sq}$, from the level of consumption that would be affected by acceptance of counterfactual financial aid offers such as those involved in the experiment. The latter will be introduced in the next section. The term $c^e_{i,sq}$ refers to consumption while in education of type e while $c^{m,e}_{i,sq}$ refers to post-education consumption generated by education of type e when in the labor market (superscript m).

The per-period utility flow of attending higher education depends on a consumption profile and a disutility level (driven by non-monetary costs). The utility flow when working in the labor market depends on the consumption level and non-monetary benefits (nmb^e_i). By allowing $c^e_{i,sq}$ to be inferior to $c^{me}_{i,sq}$, we therefore assume that the student cannot smooth consumption fully over periods of education and schooling.

One element of the utility of attending higher education which is not accounted for in the student survey is the non-monetary benefit of attending education as all questions related to non-monetary benefits are referring to amenities perceived while in the labor market.[16] For this reason, we need to introduce a supplementary element, denoted cv_i^e, and which corresponds to the consumption value of attending higher education.

The per-period utility of attending education of type e under the status quo, denoted $U_{it}^e(\cdot)$, and the period utility of working after graduation, denoted $U_{it}^{me}(\cdot)$ are then equal to

$$U_{it}^e(\cdot) = \ln(c_{i,sq}^e) - nmc_i^e + cv_i^e + \varepsilon_{it}^e$$
$$U_{it}^{me}(\cdot) = \ln(c_{i,sq}^{me}) + nmb_i^e + \varepsilon_{it}^{me} ,$$

where ε_{it}^e and ε_{it}^{me} are i.i.d. random shocks (with mean 0) and where the subjective consumption level during education, $c_{i,sq}^e$, is given by

$$c_{i,sq}^e = \tilde{c}_i^e - mc_i^e ,$$

with $\tilde{c}_i^e$ representing a reference consumption level to be detailed below.

For post-education consumption, denoted $c_{i,sq}^{m,e}$, we assume that

$$c_{i,sq}^{me} = \tilde{c} + mb_i^{me}$$

where mb_i^{me} measures the monetary equivalent of self-reported financial benefits of higher education (measured in deviation from the alternative consumption level $\tilde{c}$).

To estimate the model, we need to choose a functional form for the mapping from subjective benefit and cost factors (in capital letters), which are purely qualitative and have no monetary scale, onto relevant structural components of the ex-ante utility of higher education. The parameterization chosen is discussed below, along with identification issues. At this stage, it is sufficient to regard structural parameters as depending deterministically on the distribution of all factors.

Putting all these elements together, the expected lifetime utility of going to university, $\bar{V}_{i,sq}^u$, and its pendant for the post-secondary option , $\bar{V}_{i,sq}^{ps}$, are equal to

$$\begin{aligned} \bar{V}_{i,sq}^u &= \sum_{t=1}^{4} \beta^{t-1} \cdot (\ln\{\tilde{c}_i^u - mc_i^u\} - nmc_i^u + cv_i^u) + \\ & \quad \sum_{t=5}^{20} \beta^{t-1} \cdot (\ln\{\tilde{c} + mb_i^{m,u}\} + nmb_i^{m,u}) \end{aligned}$$

$$\begin{aligned} \bar{V}_{i,sq}^{ps} &= \sum_{t=1}^{2} \beta^{t-1} \cdot (\ln\{\tilde{c}_i^{ps} - mc_i^{ps}\} - nmc_i^{ps} + cv_i^{ps}) + \\ & \quad \sum_{t=3}^{20} \beta^{t-1} \cdot (\ln\{\tilde{c} + mb_i^{m,ps}\} + nmb_i^{m,ps}) . \end{aligned}$$

We interpret the utility functions as reduced-form representations of more structural objects which may incorporate savings decisions or any other choices exerted in the future.

[16] In the structural literature, this form of non-monetary benefit is referred to as the consumption value (Keane and Wolpin, 1997; Gong et al., 2019).

In order to estimate the model and obtain identification, we need to impose restrictions on some key parameters. Those are discussed in the next section.

In the survey, individuals must state their intention about higher education attendance in period 0. To interpret those intentions and avoid degenerate probabilities, we need to treat them as random variables. We assume that the relevant valuations, denoted $V_i^u(0)$, $V_i^{ps}(0)$ and $V_i^a(0)$, are equal to the sum of their expected lifetime utility plus a random disturbance realized in period 0. We thereby obtain the following equations:

$$\begin{aligned} V_i^u(0) &= \bar{V}_{i,sq}^u + \varepsilon_{i0}^u \\ V_i^{ps}(0) &= \bar{V}_{i,sq}^{ps} + \varepsilon_{i0}^{ps} \\ V_i^a(0) &= \bar{V}_{i,sq}^a + \varepsilon_{i0}^a \,, \end{aligned}$$

where ε_{i0}^a, ε_{i0}^u and ε_{i0}^{ps} are random shocks affecting the valuation of entering higher education in the future (when reaching period 1) but realized in period 0. When assumed to follow an Extreme Value distributions, those shocks generate standard logistic probabilities.

5.2 Identification and Estimation

To estimate the model, various restrictions (or normalizations) need to be imposed. One set relates to the measurement of individual factors. A second one relates to the identification of individual beliefs about consumption and utility levels reached while in higher education and once in the market.

5.2.1 The Distribution of Primitive Factors, Benefits and Costs

A first set of restrictions concerns the intercept terms and the loading coefficients of some of the measurement equations. A second set incorporates exclusion restrictions within the measurement system. We assume that each primitive factor has its own dedicated measurements. This means that a total of 39 measurements are used to infer 5 primitive factors. The dedication of specific measurements allow us to use the correlation between measures to infer the correlation between factors, and thereby allows us to avoid assuming orthogonality.

To be more precise, we choose one question for each set of measurement and set its loading parameter to 1 and its intercept to 0. Details are provided in Section B of the supplementary file.

When estimating the distribution of factors, we use a discrete approximation of the joint distribution of the primitive factors. We approximate the joint distribution of the 5 primitive factors by a multi-variate discrete distribution. In line with Bajari, Fox, and Ryan (2007) and Train (2008), we adopt a fixed (and known) mass points approach by choosing grid points covering the entire range of possible values and estimate all associated frequencies (type probabilities).[17] First, we normalize the measurements so to help us set

[17]While most recent papers using factor structures assume that the factors are continuous random variables, their empirical implementation require to use approximation methods which translate invariably into a form of discretization. For instance, Heckman, Humphries, and Veramendi (2018), Humphries, Joensen, and Veramendi (2019) and Ashworth et al. (2020) all use polynomial approximation methods in conjunction with Normality of the factors.

up support points for the distribution of each factor. We then assume that each factor can take one of the following values, -1.5, -0.5, 0.0, 0.5 and 1.5, except the market information factor which takes values, -1.0, 0, 0.5, 1 and 2. This generates $5^5 = 3125$ different combinations.

Obtaining identification of the qualitative monetary and non-monetary benefits and costs is relatively straightforward, given identification of the primitive factors. To do so, we make use of 69 measurements on monetary and non-monetary components (costs and benefits) of the ex-ante returns to university and other post-secondary education. All of them depend on 4 primitive factors (cognitive ability, motivation in school, locus of control and market information). The 5th factor, namely the financial constraint indicator, is used as a determinant of monetary consumption $c^e_{i,sq}$.

When estimating the distribution of benefits and costs, we fix the intercept terms (δ_0's) to 0 and estimate the ξ_i's for each type.[18] This means that each marginal distribution of costs and benefits has 3125 points.

5.2.2 Ex-ante Beliefs

In order to translate individual perceptions about benefits and costs into estimable components of the ex-ante utility of attending higher education, we need to impose restrictions on the mappings between qualitative factors (costs and benefits) and their monetary equivalents. To obtain a mapping from the factors on structural parameters, we express the structural parameters as a function of the rank of its corresponding factor.

We obtain the following set of equations:

$$\begin{aligned}
mc^e_i &= mc^e_0 + mc^e_1 \cdot Rank(MC^e_i) + mc^e_2 \cdot Rank^2(MC^e_i) + \ldots + \\
&\quad mc^e_5 \cdot Rank^5(MC^e_i) \quad \text{for } e = u, ps \\[1em]
nmc^e_i &= nmc^e_0 + nmc^e_1 \cdot Rank(NMC^e_i) + nmc^e_2 \cdot Rank^2(NMC^e_i) + \ldots + \\
&\quad nmc^e_5 \cdot Rank^5(NMC^e_i) \quad \text{for } e = u, ps \\[1em]
mb^e_i &= mb^e_0 + mb^e_1 \cdot Rank(MB^e_i) + mb^e_2 \cdot Rank^2(MB^e_i) + \ldots + \\
&\quad mb^e_5 \cdot Rank^5(MB^e_i) \quad \text{for } e = u, ps \\[1em]
nmb^e_i &= nmb^e_0 + nmb^e_1 \cdot Rank(NMB^e_i) + nmb^e_2 \cdot Rank^2(NMB^e_i) + \ldots + \\
&\quad nmb^e_5 \cdot Rank^5(NMB^e_i) \quad \text{for } e = u, ps\,.
\end{aligned}$$

We proceed similarly with the individual consumption while in school ($\tilde{c}^e_i$) and the consumption value (cv^e_i). We assume that $\tilde{c}^e_i$ depend on the financial constraint factor and the market information factor (MI_i and FC_i) and that the consumption value depends on cognitive and non-cognitive factors (CO, MS and LC). In both cases, we use polynomials of order 2 as each of the primitive factor has 5 points of support only and the ranks can only take 5 values. To limit space, we do not reproduce their equations here.

[18] Williams (2020) discusses identification of the classical linear factor model. He shows that a variety of models used in micro-econometrics tend to impose more restrictions than actually needed in order to identify their model. To provide an intuitive reasoning to identification, one has to note that, even ignoring higher order moments, the availability of 108 measurements implies nearly 6000 first and second order moments (means, variances and covariances).

As both consumption while in school ($c^u_{i,sq}$ and $c^{ps}_{i,sq}$) and post-higher education consumption ($c^{mu}_{i,sq}$ and $c^{mps}_{i,sq}$) are identified from self-reported intentions of attending higher education, it is clear that there must be an infinite number of combinations that can rationalize observed choices.

In practice, this means that the intercept terms of the expression relating individual consumption while in school ($\tilde{c}^e_i$) and the consumption value (cv^e_i) to the ranks of the primitive factors cannot be distinguish from the location parameters of the monetary and non-monetary costs and the consumption value (cv^e_0) and therefore renders the separation of consumption while in school and post-schooling consumption impossible without restrictions.

To obtain identification, we set the minimal consumption level to be the poverty level, as defined by Canadian authorities and associate it to types that are endowed to the minimum of the information factor and the maximum for both the financial constraint and the monetary cost factors.[19]

Altogether, these restrictions allows us to achieve the following. First, we can separate the monetary components of consumption while in school from the non-pecuniary dimensions of the utility of attending education, although we cannot separate the location parameters of the consumption value from that of the non-monetary costs. However, as will become clear when discussing empirical results, identification of the scale of each structural component will allow us to quantify the relative importance of each factor.

Second, with respect to post-education benefits, knowledge of the minimum level of consumption achievable (equal to that of the outside option) along with the normalization of mb^e_0 to 0, allows us to estimate the location parameter of the non-monetary benefits of education and therefore separate the per-period utility of attending education from the utility of post-education labor market experience.[20] As was the case with consumption while in school, the relative importance of monetary and non-monetary benefits may be evaluated using the scale of each component.

5.2.3 Likelihood and Estimation Method

As indicated earlier, we separate the estimation of the model from the estimation of the distribution of the primitive factors and benefits and costs of education. We do so because one of our objectives is to evaluate the explanatory power of the subjective valuations of the related benefits and costs of education. We now present each component. To ease notation, we represent all of them after conditioning on relevant factors, namely F_i and all ξ_i's.

The first component is the contribution to the likelihood of primitive factor measurements (denoted L^F_i). It is equal to

$$L^F_i(F_i) = \prod_{j=1}^{39} \Pr(M^{Fj}_i = m^{FJ}_i \mid F_i)\,.$$

[19]According to living standards found in documentation titled "Low Income Lines, 2008-2009" from the Income Statistics Division of Statistics Canada, a person living with about \$8,000 would be judged as at the poverty level.

[20]The reader will note that an alternative approach would have been to normalize the location parameter of the non-monetary benefits and estimate mb^e_0 freely. We chose to estimate the non-monetary benefits (and normalize mb^e_0) mostly because official statistics provide more guidance about the minimum consumption level than about the minimum level of non-monetary benefits induced by education.

The contribution to the likelihood of benefits and costs measurements (denoted $L_i^{BC,u}$ and $L_i^{BC,ps}$), which are used to identify the vector $\xi_i^u = \{\xi_i^{MB,u}, \xi_i^{NMB,u}, \xi_i^{MC,u}, \xi_i^{NMC,u}\}$ and the vector $\xi_i^{ps} = \{\xi_i^{MB,ps}, \xi_i^{NMB,ps}, \xi_i^{MC,ps}, \xi_i^{NMC,ps}\}$ are respectively equal to

$$\begin{aligned}
&L_i^{BC,e}(F_i, \xi_i^e) = \\
&\Pr(M^{MB,e,1} = m^{MB,e,1} \mid F_i, \xi_i^{MB,e}) \cdot \ldots \cdot \Pr(M^{MB,e,4} = m^{MB,e,4} \mid F_i, \xi_i^{MB,e}) \cdot \\
&\Pr(M^{NMB,e,1} = m^{NMB,e,1} \mid F_i, \xi_i^{NMB,e}) \cdot \ldots \cdot \Pr(M^{NMB,e,8} = m^{NMB,e,8} \mid F_i, \xi^{NMB,e}) \cdot \\
&\Pr(M^{MC,e,1} = m^{MC,e,1} \mid F_i, \xi_i^{MC,e}) \cdot \ldots \cdot \Pr(M^{MC,e,4} = m^{MC,e,4} \mid F_i, \xi_i^{MC,e}) \cdot \\
&\Pr(M^{NMC,e,1} = m^{NMC,e,1} \mid F_i, \xi_i^{NMC,e}) \cdot \ldots \cdot \Pr(M^{NMC,e,7} = m^{NMC,e,7} \mid F_i, \xi_i^{NMC,e}),
\end{aligned}$$

for $e = u$ and $e = ps$ and 69 measurements for each individual. Those are used to form a product, $L_i^{BC}(F_i, \xi_i) = L_i^{BCu}(F_i, \xi_i^u) \cdot L_i^{BCps}(F_i, \xi_i^{ps})$.

To introduce the contribution to the likelihood of higher education enrollment beliefs (under the status-quo), we need to define indicators obtained from the enrollment intentions described in Section 3.1.1. For each measure, we obtain binary indicators U_{1i} (for university), PS_{1i} (for post-secondary), A_{1i} (for no higher education) from the first measure and U_{2i}, PS_{2i}, A_{2i} (defined similarly) from the second measure. The contribution to the likelihood, denoted L_i^{edu}, is given by

$$\begin{aligned}
L_i^{edu}(\cdot) \;=\;& \Lambda_{i,w}^{u*}(.; F_i, \xi_i)^{U_{1i}} \cdot \Lambda_{i,w}^{ps*}(.; F_i, \xi_i))^{PS_{1i}} \cdot \Lambda_{i,w}^{a*}(.; F_i, \xi_i))^{A_{1i}} \cdot \\
& \Lambda_{i,w}^{u*}(.; F_i, \xi_i)^{U_{2i}} \cdot \Lambda_{i,w}^{ps*}(.; F_i, \xi_i))^{PS_{2i}} \cdot \Lambda_{i,w}^{a*}(.; F_i, \xi_i))^{A_{2ii}},
\end{aligned}$$

where

$$\begin{aligned}
\Lambda_{i,w}^{u}(\cdot) &= \frac{\exp(\bar{V}_{i,sq}^{u})}{\exp(\bar{V}_{i,sq}^{u}) + \exp(\bar{V}_{i,sq}^{ps}) + \exp(\bar{V}_{i,sq}^{a})} \\
\Lambda_{i,w}^{ps}(\cdot) &= \frac{\exp(\bar{V}_{i,sq}^{ps})}{\exp(\bar{V}_{i,sq}^{u}) + \exp(\bar{V}_{i,sq}^{ps}) + \exp(\bar{V}_{i,sq}^{a})} \\
\Lambda_{i,w}^{a}(\cdot) &= \frac{\exp(\bar{V}_{i,sq}^{a})}{\exp(\bar{V}_{i,sq}^{u}) + \exp(\bar{V}_{i,sq}^{ps}) + \exp(\bar{V}_{i,sq}^{a})},
\end{aligned}$$

The total likelihood, for a given individual i, and denoted L_i, is constructed by taking the weighted average (over types) of the product of all contributions:

$$L_i = \sum_{m=1}^{3125} p_m \cdot L_i^F(F_{im}) \cdot L_i^{BC}(F_m, \xi_m) \cdot L_i^{edu}(.; F_m, \xi_m).$$

In total, the likelihood is composed of 77,550 contributions (705 individuals times 110 measurements and choices).

Estimation Method To estimate the model, we proceed in 3 steps. They are defined as follows.

Step 1

We estimate the probabilities of all possible combinations of the primitive factors where each p_m are estimated as a Logistic transform. From the first stage likelihood, we obtain an estimate of the type probabilities, $\hat{p}_m$, by maximizing the likelihood $\sum_{m=1}^{3125} p_m \cdot L_i^F(F_m)$ where the F_m's are composed of the grid points mentioned earlier. There are 3125 of them.[21]

Step 2

With the joint distribution of primitive factors in hand, we then construct the conditional distribution of the costs and benefits factors introduced in sub-section 4.2. In the second step, we maximize the likelihood given by the following expression

$$\sum_{m=1}^{M} \hat{p}_m \cdot \hat{L}_i^F(F_m) \cdot L_i^{BC}(F_m, \xi_m^{MB}, \xi_m^{NMB}, \xi_m^{MC}, \xi_m^{NMC}),$$

where M denotes the number of elements in the support of the primitive factors which have density above the threshold and obtain $\hat{\xi}_m = \{\hat{\xi}_m^{MB}, \hat{\xi}_m^{NMB}, \hat{\xi}_m^{MC}, \hat{\xi}_m^{NMC}\}$.

Step 3

Finally, with the distribution of costs, benefits, and primitive factors, we estimate the structural parameters of the ex-ante utility of attending higher education using data on higher education self-reported intentions and the field experiment by forming the following likelihood:

$$\sum_{m=1}^{M} \hat{p}_m \cdot \hat{L}_i^F(F_m) \cdot \hat{L}_i^{BC}(F_m, \hat{\xi}_m) \cdot L_i^{edu}(\,.; F_m, \hat{\xi}_m)$$

and obtain estimates that allow us to reconstruct the population distribution of all structural parameters determining schooling intentions under the status-quo. We estimate the parameters of the model by maximizing the logarithm of the product of all individual contributions. Estimation is performed using Fortran IMSL routines.[22]

5.3 Estimates of the Ex-ante Value of Higher Education from Survey Data

Before discussing estimates of the behavioral model, we summarize the distribution of the unobserved primitive and costs and benefits factors.

5.3.1 The Distribution of Factors, Costs and Benefits

To limit the length of the paper, we only summarize some key findings and present a more detailed discussion in Section B of the supplementary file.

Not surprisingly, we find that the cognitive skill factor (CO) is positively correlated with motivation in school (MS) and with the locus of control (LC). The financial constraint factor (FC) tends to be relatively weakly correlated with other factors. The market

[21]In practice, many of these combinations are close to 0, so when maximizing the likelihood, we only retain those types with population proportion exceeding 0.0001.

[22]While splitting estimation into 3 steps allows us to obtain consistent estimates of all parameters, it does not provide consistent estimates of the standard errors. To obtain those, one method would be to use the full information likelihood function so to capture the effect of the first stage parameters and compute the resulting variance-covariance matrix using the Hessian (Arcidiacono, 2005). However, the total number of parameters needed over the 3 steps renders this approach hardly feasible.

information factor (MI) factor is positively correlated with both the cognitive factor and the motivation factor indicating that those more likely to attend higher education tend to be more informed about financial aid opportunities.

As normally expected, benefits and costs are negatively correlated. This is the case for both university and post-secondary education. At the same time, those endowed with high monetary benefit expectations also tend to be endowed with high non-monetary benefit expectations.

We find that the primitive factors explain at most 50% of the total variation across types for each benefit or cost factor. In some cases such as university and post-secondary monetary benefits, the primitive factors account for at most 25%. In many others, they account for one third only.[23] This validates our approach which consisted in allowing each cost and benefit factor to be explained by its own unobserved heterogeneity.

5.3.2 Structural Estimates

With the distribution of benefits and costs in hand, it is now possible to obtain the structural parameters governing higher education decisions by maximizing the total likelihood of reported attendance intentions. As the parameters mapping ranks onto utilities are not directly interpretable, they are presented in a supplementary file. It is however more informative to examine the relationship between the ranks of the costs and benefit factors and the relevant per-period utilities (Figure 1).

[Figure 1 here]

In general, the functions mapping factor ranks onto utilities reproduced in Figure 1 reveal non-trivial relationships. While self-reported monetary and non-monetary costs of university appear to have a monotonically decreasing impact on utility, this is far from being the case for all other components. The effect of the university monetary benefit factor appears to be particularly flat while the non-monetary benefit factor discloses a weakly increasing effect in the initial phase but reaches a flat spot at higher ranks.

The presence of non-monotonicities appear to be particularly problematic with post-secondary monetary and non-monetary costs. In both cases, and at least in the initial ranges, higher costs imply higher per-period utilities. On a more positive side, both post-secondary monetary and non-monetary benefits seem to have a monotonically increasing effect on per-period utility. However, the relationship seems weak thereby indicating that post-secondary education subjective benefits may not be informative.

[Table 2 here]

As indicated in Table 2, our model predicts stated education intentions quite well as it predicts that 66% will attend university, 31% will attend other post-secondary institutions and only 3% would not go beyond high school. These are very close to the average frequencies observed for the first measure (question D19), which are equal to 64%, 32% and 4% respectively.

[23] In a supplementary file, we also report the structural parameters of the function mapping factors onto benefits and costs. Overall, the results are coherent with pure intuition. For instance, cognitive skills and motivation in school are found to have a positive impact on both monetary and non-monetary benefits and a negative effect on non-monetary costs.

The average lifetime utilities obtained for the 3 options (50.6 for university, 48.8 for other post-secondary) are also reported in Table 2. There is a substantial amount of heterogeneity as indicated by the high enrollment probability inter-quartile ranges, equal to 0.59 ($0.98 - 0.39$) for university and 0.51 ($0.53 - 0.02$) for other post-secondary.

To ease comprehension and first obtain a crude comparison with respect to the outside option, it is useful to translate the intertemporal utilities into a monetary indicator and evaluate the flat consumption profile that would provide the level of intertemporal utilities of the two education options (university and other post-secondary institution) since the alternative option is itself defined as a flat consumption profile of \$34,000 per year. Those are reported in Table 2 as well. They indicate that for the average individual in the population, choosing university provides a welfare level that is equivalent to a flat consumption profile equal to \$63,500 per year (for 20 years) while choosing other post-secondary is equivalent to a total yearly consumption of \$53,173. These numbers obviously reflect the totality of financial and non-monetary rewards and ignore the split between education years and post-education years, but they also illustrate the important role of heterogeneity as the inter-quartile differentials are close to \$30,000 ($\$74,651 - \$45,121$) for university and \$20,000 ($\$61,002 - \$43,152$) for post-secondary.

To separate consumption from non-monetary utilities and to distinguish between utility flow during and after education, it is interesting to gauge the importance of the 4 main monetary components. Three of them, namely monetary costs, market information and financial constraint perceptions, affect expected consumption while in school while the 4th one, monetary benefits, affects expected post-education consumption. As stated earlier, it is not possible to separate the location parameters of the monetary and non-monetary components, but as their scale depend directly on individual factors, it is still possible to assess their relative importance.

To evaluate it, we computed the impact of shutting down those specific factors on lifetime consumption. The contribution of each model component is calculated using the difference between the estimated type-specific consumption level and the consumption level obtained when all types are set at the minimum level.

[Table 3 here]

The results, found in Table 3, indicate that individual differences in liquidity constraints and differences in monetary costs perceptions are a more important determinant of the decision to attend other post-secondary institutions than university. Precisely, eliminating monetary costs would raise consumption by \$15,603 while attending university and by \$20,299 for other post- secondary. Differences in financial constraint perceptions appear to have a small impact on consumption in university (\$6770) but a large one on consumption while attending post-secondary (\$18,910). Interestingly, the factor measuring individual differences in information has its largest effect on post-secondary attendance as eliminating information would reduce consumption by \$34,841. This may indicate that those who report being well-informed are also those enjoying a higher consumption level either because they are receiving larger transfers from their parents or because they expect more financial aid.

As will become clear below, non-monetary benefits are a much more important determinant of revealed higher education intentions than monetary benefits for either university or post-secondary. However, we find monetary benefits to be a much more important determinant of post-secondary institution attendance than university. Putting every type of individuals at the lowest level, would reduce post-education consumption by $121,257 over the early life-cycle when considering post-secondary, as opposed to $33,557 for university. This indicates that young individuals have much more dispersed beliefs about the financial rewards of post-secondary education compared to their beliefs about university education.

[Table 4 here]

We now turn to the quantification of the overall differences between monetary and non-monetary components, and put them in perspective. By shutting down each possible component (both monetary and non-monetary), we can evaluate the extent to which individual intentions are explained primarily by non-monetary dimensions. In Table 4, we report the % changes in lifetime utility attributable to each dimension, and also their impact on expected enrollment probabilities.

The results show ample evidence that the intention to enroll in university is motivated by non-pecuniary elements. Non-monetary benefits are by far the most important as shutting down their effects would reduce lifetime utility by 6.2% and reduce enrollment proportion by 0.34. Non-monetary costs appear to be the second most important element, as it would reduce life-cycle well being by 3.6% and reduce enrollment probability by 0.21. The factors measuring cognitive ability and motivation in school, which presumably approximate the consumption value of education, are the 3rd and 4th most important determinants. Putting everyone at the lowest level of the liquidity constraint factor would raise the lifetime utility of attending university by less than 1% and affect enrollment only marginally.

The results are somewhat similar for other post-secondary institutions. Again, non-monetary benefit is the most important determinant with welfare effects equal to -6.5% and an enrollment impact equal to -0.27, although monetary benefits is the second determinant with a 3.8% loss when removing it and an effect of 0.20 on enrollment probability. As for university enrollment, non-monetary costs are an important determinant, as eliminating them would raise welfare by 3.7% and raise enrollment probability by 0.18. As was the case with university, both financial constraint perceptions and information are relatively unimportant.

To summarize, beliefs and intentions disclosed by the students participating in the survey imply a very high valuation of higher education (especially university), very high hypothetical attendance probabilities, and a particularly important role played by non-financial dimensions.

5.3.3 Robustness and Reduced-Form Estimates

In order to assess the robustness of the results obtained from the structural model, we also evaluated the impact of all factors, using the reduced-form representation of the ex-ante attendance probabilities from survey data. We therefore use a reduced-form representations of the intertemporal utilities of attending university and other post-secondary institutions ($\bar{V}^{u}_{i,sq}$ and $\bar{V}^{ps}_{i,sq}$) in reference to $\bar{V}^{a}_{i,sq}$ by maximum likelihood assuming Extreme-Value random shocks. As we did when estimating the structural model on the survey, we set $\bar{V}^{a}_{i,sq}$ to $\sum_{t=1}^{20} \beta^{t-1} \cdot \log(\tilde{c})$ where $\tilde{c} = \$34,000$. We specify $\bar{V}^{u}_{i,sq}$ and $\bar{V}^{ps}_{i,sq}$ as functions of

X_i^u and X_i^{ps} (defined in Section 3.1)

$$\begin{aligned}\bar{V}_{i,sq}^{u} &= \gamma' X_i^u + \gamma_u' F_i \\ \bar{V}_{i,sq}^{ps} &= \gamma' X_i^{ps} + \gamma_{ps}' F_i\,,\end{aligned}$$

and therefore ignore the distinction between consumption while in school and post-education consumption. In Table 5, we report the marginal effects obtained for a one-standard deviation change in each determinant.

[Table 5 here]

Overall, the results are consistent with those obtained when estimating the structural model. First, the probability of attending university decreases with its own costs and increases with its own benefits. This is also the case with post-secondary education.

Second, as was noted when estimating the structural model, individual choices appear to be based primarily on non-monetary components as opposed to monetary benefits and costs. The effect of a one-standard deviation increase in university non-monetary benefits on university attendance (equal to 0.11) is twice as large as the effect of a similar increase on monetary benefits (0.05). The effect of non-monetary costs (-0.09) is about 5 times larger than the effect of monetary costs (-0.02). The results are practically identical for the demand for post-secondary institution, as the impact of non-monetary benefits are 2 times as large as the impact of monetary benefits and the non-monetary costs effects are 4 times as important as monetary costs.

Finally, among individual factors, the cognitive skill factor is by far the most important as it raises university attendance by 0.09 and reduces post-secondary attendance by 0.07. It is interesting to note that the financial constraint factor has no significant impact on either university or other post-secondary attendance. The estimates indicate that it reduces university attendance probability by 0.01 but its effect is insignificant and that the impact on post-secondary attendance (0.003) is very small. The factor measuring information seems to have no impact whatsoever.

As both structural and reduced-form estimates point toward the domination of non-financial components of higher education intentions, it will now be interesting to evaluate their implications for the valuation of counterfactual financial aid policies.

5.3.4 Implications for Financial Aid

Returning to the main objective of the paper, which is to obtain two separate distributions of the welfare gains of financial aid offers in the experiment, we now use the structural parameters obtained from survey data to predict the welfare gain of the specific offers found in the experiment (found in Table 1). Any offer q may be represented by a vector $A_q = \{G_q, L_q\}$ where G denotes a grant and L a loan, and where either G or L must be equal to 0 and let's introduce an indicator, d_{iq}, equal to 1 when financial aid offer q is accepted and 0 if not.

To obtain the distribution of the marginal utilities of financial aid from the surveys, we increase the consumption while in school under the status quo by the amount of financial aid accepted and the level prevailing after conditioning on financial aid offer acceptance and now write the new consumption level, $c_{iq}^e(\cdot)$, as a function of financial aid decisions about offer q:

$$c_{iq}^e(d_{iq}, A_q) = c_{i,sq}^e + c_q^{e,fa}(d_{iq}, A_q) \quad \text{for } e = u, ps,$$

where $c_q^{e,fa}(\cdot)$ is a function that maps financial aid offers and period 0 decisions onto consumption during university. Note that $c_q^{e,fa}(\cdot)$ is allowed to depend on the "e" superscript only because we assume that the agent divides the amount of financial aid equally across years required by a standard curriculum which is not the same at a university than at 2-year colleges. It also depends on q as each offer is potentially different. They are given by the following equations:

$$\begin{aligned} c_q^{u,fa}(\cdot) &= (0.25 \cdot G_q + 0.25 \cdot L_q) \cdot d_{iq} \\ c_q^{ps,fa}(\cdot) &= (0.50 \cdot G_q + 0.50 \cdot L_q) \cdot d_{iq} \,. \end{aligned}$$

Second, we must model the effect of loan expected repayment on post-university consumption. Denoting post-university consumption by $c_{iq}^{m,e}(\cdot)$, we define it as follows

$$c_{iq}^{m,e}(d_{iq}, A_q) = c_{i,sq}^{m,e} - c_q^{m,fa}(d_{iq}, A_q)\,,$$

where $c_q^{m,fa}(\cdot)$ denotes the yearly repayment that is induced by accepting a loan. We assume that individuals anticipate repayment over 10 years (maximum under Canadian regulations) and start repaying 1 year after college completion. Because repayment only applies for loans (that is when $L_q > 0$), it follows that

$$c_q^{m,fa}(d_{iq}, A_q) = L_q^r \cdot d_{iq}$$

where L_q^r denotes the yearly repayment associated to L_q. We define it as

$$L_q^r = L_q \cdot \left(r + \frac{r}{(1+r)^{10} - 1} \right) ,$$

where r is the interest rate.

Finally, after replacing the status-quo consumption levels estimated in the survey ($c_{i,sq}^e$ and $c_{i,sq}^{m,e}$) by their values reflecting the impact of accepting financial aid ($c_{iq}^e(\cdot)$ and $c_{iq}^{m,e}(\cdot)$), it is then easy to evaluate the marginal lifetime utilities of accepting any grant or loan, including all those that are offered in the experiment. These counterfactual marginal lifetime utilities will then be confronted to those estimated in the experiment in Section 7.

Before doing so, and in order to illustrate the main features of the model estimated from the survey, we first use the structural parameters obtained from survey data to predict the welfare gain of providing generous higher education financial aid and its related impact on counterfactual higher education attendance probabilities. To do so, we simulate the implementation of a generous \$16,000 grant (\$4,000 for each year of university) which would make practically university education free.[24] We proceed similarly for post-secondary education. Finally, we examine the impact of an annual \$4,000 subsidy for either type of education. The results are in Panel A of Table 6.

[Table 6 Panel A here]

The results provide some clear evidence on the behavioral implications of individual beliefs and intentions. Put generally, those imply that generous higher education financial aid would be welfare improving but would have little impact on total higher education enrollments.

[24]The impact of loan provisions will be analyzed in the context of the field experiment.

First, when considered separately, providing free post-secondary education and free university would increase ex-ante welfare by 0.13% (for post-secondary) and 1.12% for university. Providing free post-secondary education would raise post-secondary enrollments by 0.023, but this increase would be at the expense of university enrollments which would drop by 0.020. Providing free university would raise university enrollments by 0.098 but would reduce post-secondary enrollments by 0.089. In the case where both types of higher education would be subsidized, the increase in university enrollments would be slightly lower than when only university is subsidized (0.080) but the decrease in other post-secondary institution enrollments would also be more important (0.070). So, despite the welfare gains, the total of post-secondary and university enrollments would raise by 0.010.

[Table 6 Panel B here]

In Panel B of Table 6, we report the welfare gains induced by a $2,000 grant and a $2,000 loan, as these are representative of many financial aid package offered in the field experiment. Note that these offers are fully portable in any post-secondary institutions. In line with was noted already when commenting on Panel A, the welfare gain of a $2,000 student loan would represent less than a one tenth of 1% increase in lifecycle utility and would correspond to the difference between the modest gain in utility attributed to early consumption and the small loss caused by subsequent repayment.

The findings reported in both panels of Table 6 are easily explained. While individual consumption while in school matters, it seems that enrollment expectations are already very high under the status quo. As noted earlier, enrollments are primarily driven by non-financial parameters. This indicates that providing generous grants (reducing the cost of higher education) or generous loans remains ineffective at raising enrollments but raises welfare mostly by raising consumption during education.

6 Modeling Decisions in the Experiment

6.1 The Model

As choices exerted in the field experiment disclose information about the marginal utility of financial aid only, it is not possible to identify the same structural parameters estimated from the survey such as the expected value of following the optimal strategy (either university, post-secondary or alternative option) given rejection of financial aid (under the status-quo). This is however not needed as we can already obtain the distribution of the marginal utility of financial aid from the survey.

Recalling that in the experiment, individuals exert q binary decisions between immediate cash payments (in period 0) and financial aid, it follows that recovering the intertemporal marginal utilities of financial aid requires to identify the difference in period 0 utilities induced by the acceptance or the rejection of the cash payment.

One limitation of the experiment is that individuals disclose their preference between accepting cash or taking up financial aid but it is impossible to know if individuals who accept the cash payment also save a portion for the period when they decide to enter higher education or not. While saving behavior is unlikely to be relevant when cash payments are relatively modest, it may be more likely when individuals face decisions between financial aid and cash offers of \$700. As estimating a saving rate would be tenuous, we ignore saving and instead we estimate the model presented below on all decisions, and then re-estimate it after excluding decisions in which the cash payment is equal to \$700. If a large share of the experiment's subjects choose cash in order to save, this would automatically tend to reduce the valuation of financial aid and affect our results. We shall return to this point when discussing empirical estimates.

To retain symmetry with the model estimated from survey data, we assume a logarithmic utility function. We interpret choices as involving a comparison between two intertemporal utilities. Dropping the individual subscript for convenience, the value of accepting the cash payment at decision q, denoted C_q, is equal to

$$V(\text{accept } C_q) = \ln(c_r + C_q) + \beta_1 E \max(V^u, V^{ps}, V^a \mid \text{status-quo})\,,$$

where c_r is a reference consumption level that would prevail if the cash payment is rejected and $\beta_1 E \max(V^u, V^{ps}, V^a \mid \text{status-quo})$ is the intertemporal utility achievable beyond high school graduation in absence of any additional financial aid opportunities. The discount factor, denoted β_1, is equal to $\frac{1}{1+0.05\cdot\Delta t_i}$, where Δt_i is the fraction of the year that separates the experiment date to September 1st, 2009. Note that $E \max(V^u, V^{ps}, V^a \mid \text{status-quo})$ is identified from the survey as well.

The value of accepting financial aid A_q is given by

$$V(\text{accept } A_q) = \ln(c_r) + \beta_1 E \max(V^u, V^{ps}, V^a \mid A_q)\,,$$

and the acceptance probability, $\Pr(\text{accept } A_q)$, is therefore equal to

$$\begin{aligned}\Pr(\text{accept } A_q) \;\; = \;\; & \Pr\{\ln(c_r) + \beta_1 E \max(V^u, V^{ps}, V^a \mid A_q) \geq \\ & \ln(c_r + C_q) + \beta_1 E \max(V^u, V^{ps}, V^a \mid \text{status-quo})\}\end{aligned}$$

for $q = 1, 2, \ldots, 12$. The latter may be identified given knowledge of the difference between $\ln(c_r)$ and $\ln(c_r + C_q)$, which we address in Section 6.2 below.

In what follows, we distinguish between the marginal intertemporal utility of accepting a grant G_q (which reduces the cost of education) and the marginal intertemporal utility of accessing a loan L_q. These quantities are formally defined as follows:

$$MV_{g,q} = \beta_1 E \max(V^u, V^{ps}, V^a \mid A_q = \ G_q) - \beta_1 E \max(V^u, V^{ps}, V^a \mid \text{status-quo})$$

$$MV_{l,q} = \beta_1 E \max(V^u, V^{ps}, V^a \mid A_q = L_q) - \beta_1 E \max(V^u, V^{ps}, V^a \mid \text{status-quo})\,.$$

and their parameterization are presented below.

6.2 Estimating the Marginal Utility of Consumption

In order to obtain the marginal utilities of financial aid, we need to obtain (and identify) the marginal utility of consumption induced by accepting (and consuming) the cash payment. To achieve this we make use of a large number of binary choices (also incentivized) between pairs of lotteries which all individuals needed to exert in the first portion of the experiment. These have been used in recent papers (Belzil, Maurel, and Sidibé, 2021; Jagelka, 2020).

We make use of 30 binary choices following the popular Holt and Laury design with narrow brackets. These may be described as follows. Each decision, indexed by l, requires to choose between two lotteries X_l and Y_l. In each case, the second lottery Y_l is unambiguously more risky then the second one.

The first lottery is characterized by a low payoff, x_{1l} and a high payoff, denoted x_{2l}, while the second lottery entails a low payoff, y_{1l} and a high payoff, denoted y_{2l}. For each lottery, the probability of the high outcome is equal to p_{2l} and the probability of the low outcome is $p_{1l} = 1 - p_{2l}$.

In absence of any stochastic element, the individual selects the lottery that entails the highest expected utility. That is the individual compares $EU(X_l)$ with $EU(Y_l)$ where $EU(X_l)$ and $EU(Y_l)$ are given by the following

$$\begin{aligned} EU(X_l;\theta) &= p_{1l} \cdot \ln(c^r + x_{1l}) + p_{2l} \cdot \ln(c^r + x_{2l}) \\ EU(Y_l;\theta) &= p_{1l} \cdot \ln(c^r + y_{1l}) + p_{2l} \cdot \ln(c^r + y_{2l}) \,, \end{aligned}$$

where c^r is the background (reference) level of consumption. As the payoffs range from \$25 to a maximum of \$96, we assume that the entire payment would be consumed before higher education enrollment (over the period between the experiment and the start of period 1) and ignore saving. In Section C of the supplementary file, we describe all payoffs and probabilities.

We follow the literature and assume that choices are stochastic. Assuming logistic errors, we obtain the following probability:

$$\Pr(EU(X_l;\theta) > EU(Y_l;\theta)) = \frac{\exp(p_{1l} \cdot \ln(c^r + x_{1l}) + p_{2l} \cdot \ln(c^r + x_{2l}))}{\exp(p_{1l} \cdot \ln(c^r + x_{1l}) + p_{2l} \cdot \ln(c^r + x_{2l})) + \exp(p_{1l} \cdot \ln(c^r + y_{1l}) + p_{2l} \cdot \ln(c^r + y_{2l}))} \,.$$

To capture heterogeneity in the marginal utility of accepting the cash payment across individuals, we allow the reference consumption level to depend on the same heterogeneity components of our model, namely type specific benefits and costs and write the reference consumption level as follows:

$$c_i^r = r_0 + r_1' \cdot X^u + r_2' \cdot X^{ps} \,,$$

where r_0 is an intercept term and r_1 and r_2 are column vectors of parameters.

With these estimates in hand, we can now recover the difference in period 0 utilities of consumption. Ignoring potential savings in period 0, the difference in period 0 utilities of consumption, denoted $MU_{0,q}$, is equal to

$$MU_{0,q} = \ln(c_r + c_q) - \ln(c^r) \,.$$

The results, summarized in the first column of Table 7, indicate a relatively low level of the reference consumption. The average and the median reference consumption levels are around \$10 and point toward a reasonable level of heterogeneity in the marginal utility of consumption in period 0 as half of the population have a reference level between \$7 and \$13. In column 2, we report the distribution of $MU_{i,0}$ corresponding to \$300, as it represents a cash payment representative of many decisions in the experiment. As should be clear now, $MU_{i,0}$ plays the role of the benchmark utility (marginal) level used to anchor the lifecycle marginal utility of accepting financial aid and is the component of the model that introduces the opportunity cost which acts as a discipline device for agents' decisions in the experiment.

[TABLE 7 HERE]

6.3 Specification and Estimation of the Value of Financial Aid

Our objective is now to estimate the reduced-form of the marginal valuations of financial aid in the experiment and decompose the ex-ante valuations obtained from the survey into the same components so to investigate coherency. As documented in Section 5.3.4, it is easy to compute marginal utilities of any financial aid offer using structural parameters obtained from the survey. So, from now on, we introduce a superscript (s for survey and fe for the field experiment), in order to distinguish MV_g and MV_l obtained from the experiment and the survey. MU_0 has no pendant in the survey so it does not need any superscript.

In order to parameterize the marginal values of a grant (MV^{fe}_{i,g_q}, MV^{s}_{i,g_q}) and a loan (MV^{fe}_{i,l_q}, MV^{s}_{i,l_q}) we use 3 main elements: the estimated ex-ante probabilities of enrolling in university and post-secondary education ($\Lambda^u_i(\cdot)$ and $\Lambda^{ps}_i(\cdot)$) which both map all factors onto a scalar expression, the monetary amount of the offer (either the grant or the loan) and the set of all factors (the X vector). As financial aid offers apply either to university or post-secondary, it is informative to consider the effect of the sum of both attendance probabilities ($\Lambda^u_i + \Lambda^{ps}_i$) as well as the effect of a binary indicator of whether or not an individual has a higher probability of attending university than post-secondary ($\mathbb{1}(\Lambda^u_i > \Lambda^{ps}_i)$).

We thereby obtain the following representations of the intertemporal marginal utilities:

$$\begin{aligned} MV^e_{i,g_q} &= g'^e_1 X_i + g^e_2 \cdot (\Lambda^u_i + \Lambda^{ps}_i) + g^e_3 \cdot \mathbb{1}(\Lambda^u_i > \Lambda^{ps}_i) + \\ &\quad g^e_4 \cdot \mathbb{1}(G_q = \$1000) + g^e_5 \cdot \mathbb{1}(G_q = \$2000) + g^e_6 \cdot \mathbb{1}(G_q = \$4000) \text{ for } e = s, fe \end{aligned}$$

$$\begin{aligned} MV^e_{i,l_q} &= l'^e_1 X_i + l^e_2 \cdot (\Lambda^u_i + \Lambda^{ps}_i) + l^e_3 \cdot \mathbb{1}(\Lambda^u_i > \Lambda^{ps}_i) + \\ &\quad l^e_4 \cdot \mathbb{1}(L_q = \$2000) + l^e_5 \cdot \mathbb{1}(G_q = \$4000) \quad \text{for } e = s, fe \end{aligned}$$

where g^e_1 and l^e_1 are both column vectors of parameters (including an intercept), and $g^e_2, \ldots, g^e_6$, $l^e_2, \ldots, l^e_5$ are parameters to be estimated. The reference grant level is \$500 and the reference loan level is \$1000.

To perform estimation of MV^{fe}_{i,g_q} and MV^{fe}_{i,l_q}, we first introduce an indicator, d_{iq}, which is equal to 1 when individual i accepts financial aid offer q and 0 if not, and assume the existence of a logistic random shock affecting additively the utility of accepting the cash. We thereby obtain the following expression (to take the example where offer q consists of a grant):

$$\begin{aligned} \Pr\{d_{iq} = 1\} &= \Lambda(\beta_1[E\max(\cdot \mid A_q) - E\max(\cdot \mid \text{status-quo})] - MU_{i,0,q}) \\ &= \Lambda(MV^{fe}_{i,g_q} - MU_{i,0,q})\,, \end{aligned}$$

where $MU_{i,0,q}$ is defined above, and where $\Lambda(\cdot)$ denotes the Logistic distribution function. To obtain the parameters, we maximize the following likelihood:

$$L^{FE}_i = \sum_{m=1}^{M} \hat{p}_m \left\{ \prod_{q=1}^{12} \Pr\{d_{iq} = 1 \mid F_m, \hat{\xi}_m\}^{d_{iq}} \cdot \Pr\{d_{iq} = 0 \mid F_m, \hat{\xi}_m\}^{(1-d_{iq})} \right\}.$$

6.4 Estimates of the Value of Financial Aid

The distributions of the intertemporal marginal utilities obtained from the experiment and the survey are summarized in Table 8. Individual valuations of grants disclosed within the experiment (equal to 5.27 on average), are much higher than in the survey (equal to 0.09 on average) and the marginal utility of loans is also higher in the experiment (1.31) than in the survey (0.05). It is however difficult to say more about these differences at this stage since the location and scale of the marginal utilities estimated in the experiment are benchmarked according the utility of consumption of period 0, which does not last one full year period, and may therefore depend crucially on the reference consumption level. On the other hand, valuations obtained from the survey are inferred from hypothetical future payoffs which have no obvious connection to immediate (period 0) consumption and which are expressed in reference to lifecycle consumption profile. We return to this point in Section 7 but for the moment, these differences are not sufficient to conclude anything about coherency. At most, can we observe that the average marginal utility of a \$2000 loan in the experiment, is below the marginal utility of a \$300 cash payment (equal to 3.44 in Table 7), despite that about 10% of the sample attributes a negative value to a \$2,000 loan and that, as indicated in Table 1, many individuals are actually willing to sacrifice an immediate cash payment against a loan.

[Table 8 here]

In Table 9, we report the correlations between survey and experiment valuations for a given type of financial aid and the correlations between grant and loan valuations for a given setting (either the survey or the experiment). First, the correlations between grant and loan valuations are relatively high in both the experiment and the survey. Either in ranks or in level, they are between 0.9 and 1.0 and indicate that individuals with high grant valuations seem to also have high loan valuations. However, the correlations between survey and experiment valuations (either in rank or in level), and which are around 0.3 for grants and 0.4 for loans, are much lower and suggest a disconnection between individual valuations in the experiment and in the survey. This raise serious doubts about whether or not choices in the experiment are coherent with both intentions and valuations disclosed in the survey and leads us to investigate the reasons explaining this low level of coherency.

[Table 9 here]

The decompositions of grant and loan valuations into their determinants are found in Table 10. The estimates obtained when decomposing the experiment valuations are in columns 2 (grants) and 4 (loans). Those obtained from the survey are in columns 1 (grants) and 3 (loans).

[Table 10 here]

We first compare the effect of some determinant of a specific valuation obtained with one data source with the effect of the same factor on the same valuation but estimated with the alternative data source. For this sort of comparison, and for reasons aforementioned, it is much safer to focus our discussion on the signs of the parameters measuring the effect of factors, rather than on their levels.[25]

First, in both the survey and the experiment, grant valuations increase with the ex-ante probability of attending higher education and are higher for those who have a higher probability of attending university than other post-secondary institutions, although the magnitudes of the parameters are much higher in the experiment than in the survey. As normally expected, our estimates also indicate that both grant and loan valuations are increasing with the amount involved, but again, the impact is much higher in the experiment. These results are therefore not problematic.

When examining the impact of the individual factors, it is intriguing to note that the signs of many determinants (mostly costs and benefits) on valuations are not the same in the survey and the experiment. For instance, in the case of grants (columns 1 and 2), the effects of 5 determinants (reported in bold in Table 10) switch signs between the experiment and the survey. Those are the Monetary Benefits and Monetary Costs of University, the Monetary Benefits and Non-Monetary Costs of Post-Secondary, and the Information factor.

One major difference between the marginal utilities of loans and grants is the existence of negative valuations for loans (documented in Table 8), as we find that about 10% of the population attach a negative marginal valuations to loans. Nevertheless, the decomposition of the marginal values of loans into ex-ante probabilities and factors (reported in columns 3 and 4) disclose similar tendencies. More importantly, and as was the case with grant valuations, we also find that 5 factors switch signs between the survey and the experiment. Those are the Monetary Benefits, the Monetary Costs and Non-Monetary Costs of University, the Monetary Benefits and Non-Monetary Costs of Post-Secondary and are virtually the same that switched signs in the grant valuation (except for the Non-Monetary Costs of University). However, the effect of the Information factor on loans remains negative in the survey and the experiment, whereas it switched signs for grants.[26]

Overall, and to the extent that one is willing to attach more reliability to data obtained in a high-stake context than in an incentive-free setting, these findings raise serious doubts about the validity of self-reported components of the value of education and/or self-reported enrollment intentions, as their impact on both grant and loan valuations differ between the experiment and the survey. For this reason, we now turn to a more formal analysis of the degree of coherency (or lack thereof) that may be detected and we discuss some policy implications.

[25] In this paper, we ignore comparisons between loan and grant valuations (when obtained with same type of data) as this sort of analysis would lie beyond the scope of the paper. As pointed out in the theoretical literature, financial aid programs are likely to be subject to severe Moral Hazard problems (Lochner and Monge-Naranjo, 2016) and there may be individual factors affecting loan and grant valuations differently. As noted in Section 2, the valuations of loans and grants have practically never been estimated structurally. The only published paper that we are aware of is Belzil, Maurel, and Sidibé (2021).

[26] We also estimated the same model after excluding two decisions involving \$700 because those are more likely to be affected by savings than decisions involving smaller cash amounts. Most parameter estimates were found to be very close. However, incoherency appeared to be as important as we found an additional factor (the monetary costs of post-secondary) that switched sign between the survey and the experiment.

7 Estimating Coherency using a Confrontation Approach

While there are huge differences between valuations inferred from the survey and those from the experiment, it is difficult to say to what extent those depend on the reasons stated in Section 6.4. Evaluating the degree of coherency requires to avoid methods that would depend too much on comparison of the location and scale of the cardinal utilities. This is the main issue that we now address.

To evaluate the degree of coherency, we propose a rank-rank confrontation approach based on regressing the rank of the valuations inferred from the survey and those from the experiment on the ranks of the factors. Using ranks instead of cardinal utilities allows us to circumvent the problems aforementioned. To achieve this, we use the following 4 regressions:

$$\begin{aligned}
Rank(MV_{i,g}^{fe}) &= R_g^{fe\prime} \cdot Rank(X_i) + error \\
Rank(MV_{i,g}^{s}) &= R_g^{s\prime} \cdot Rank(X_i) + error \\
Rank(MV_{i,l}^{fe}) &= R_l^{fe\prime} \cdot Rank(X_i) + error \\
Rank(MV_{i,l}^{s}) &= R_l^{s\prime} \cdot Rank(X_i) + error\,,
\end{aligned}$$

where $Rank(X_i)$ denotes a vector measuring the rank of each factor (benefit, cost and primitive factors) and R_g^{fe}, R_g^s, R_l^{fe} and R_l^s are vectors of parameters to be estimated. If individual valuations and intentions revealed in the survey are accurate, the mapping from individual ranks in costs, benefits and factors onto valuation ranks should be the same in the survey and the experiment. Our analysis is therefore based on comparing R_g^{fe} to R_g^s and R_l^{fe} to R_l^s. To facilitate comparisons, we estimate the experiment parameters (the R_g^{fe} and R_l^{fe}) in deviation from those obtained with the survey (the R_g^s and R_l^s).

[Table 11 here]

In Table 11, we report the parameter estimates obtained for both the survey and the experiment and indicate by a "reject" the determinants for which the null hypothesis that the deviation parameter is equal to 0 is actually rejected. A rejection of the null is essentially an indicator that the weight attributed to a given factor in the survey is different in the experiment. In order to take into account potential heteroskedasticity induced by conditioning on the distribution of factors estimated in the first stages, we compute robust standard errors.

The results are striking. Focusing first on grant valuations, we find that equality is strongly rejected for all parameters (excluding the intercept term). When turning to loan valuations, the results are also spectacular. Except for 2 factors (post-secondary monetary costs and financial constraint factor) equality is also rejected .

One interesting question is whether or not going from utility to ranks resolves the inconsistencies in sign that we noted earlier. The answer is no. As for valuations measured in utilities, the 5 factors that switch signs when regressing grant valuations (marginal utilities) on factors also switch signs when using rank relationships. The results are identical for loans. The same factors switching signs in Table 10 also switch signs when using rank-rank relationships. These results raise an important issue, which is how to obtain a measure of coherency.

Focusing on grants as an illustration, this suggests using the difference in model fit (sum of square residuals) between the reduced form model mapping factor ranks onto experiment valuation ranks, $\sum_{i=1}^{N}(Rank(MV_{i,g}^{fe}) - \hat{R}_g^{fe\prime} \cdot Rank(X_i))^2$, and the sum of square residuals obtained when restricting the mapping parameters to be those estimated from the survey, $\sum_{i=1}^{N}(Rank(MV_{i,g}^{fe}) - \hat{R}_g^{s\prime} \cdot Rank(X_i))^2$, in order to build the ratio of the former by the latter. Perfect coherency would arise when the ratio is equal to 1 and the overall degree of coherency decreases as the ratio approaches 0.

At the bottom of Table 11, we report our measure of coherency applied to experiment-survey comparisons for grants and loans, with the parameters reported in the same table. In both cases, the indicator is low. The ratio of 0.15 obtained for grants indicate that the sum of square residuals obtained when using the survey parameters is about 6 times higher than the unrestricted sum of squares. For loans, the coherency indicator, equal to 0.25, is slightly lower, but points again to a high degree of incoherence.

7.1 Absolute vs. Biased Incoherency

We now analyze the distinction between two types of incoherency: Absolute Incoherency and Biased Incoherency. To measure Absolute Incoherency, we use the absolute difference between the grant (or loan) valuation rank in the experiment and the corresponding valuation in the survey ($|\text{Rank}(MV_g^{fe}) - \text{Rank}(MV_g^{s})|$ in Table 12). Biased Incoherency is meant to measure the tendency to disclose a higher valuation of financial aid in the experiment than in the survey (or vice-versa). To measure Biased Incoherency, we use an indicator equal to 1 when the grant (or loan) valuation rank in the experiment exceeds the grant (or loan) valuation in the survey ($\mathbb{1}(\text{Rank}(MV_g^{fe}) - \text{Rank}(MV_g^{s})) > 0$ in Table 12). We then use OLS regressions to decompose each measure on all factors.

[Table 12 here]

In order to obtain a general feeling about which type of incoherency is more predictable, it is sufficient to compare standard goodness of fit measures obtained for Absolute and Biased Incoherency. There is overwhelming evidence that individual factors, especially subjective costs and subjective benefits, have a much higher explanatory power on Biased Incoherency than on Absolute Incoherency as the R squares of the Biased Incoherency regressions (0.62 for grants and 0.60 for loans) are more than 6 times higher than the Absolute Incoherency R squares (0.16 for grants and 0.09 for loans). Because the notion of absolute incoherency does not allow us to uncover interesting relationships with individual factors, we focus our discussion on Biased Incoherency.

This finding generates a first important remark about the nature of incoherency. The tendency to be biased-incoherent cannot be represented by purely stochastic behavior only since it appears to be strongly correlated with individual heterogeneity measured by various latent factors.

We now take a deeper look and discuss the role of subjective costs and benefits. University non-monetary benefits seem to have the strongest impact on the likelihood to disclose a higher valuation of both loans and grants in the experiment than in the survey, as it is found to increase Biased Incoherency (the estimate is equal to 0.2775 for grants and 0.3388 for loans). To a lesser extent, those reporting high benefits to post-secondary education are also likely to disclose incoherent choices. On the other hand, some of the costs and benefit factors are found to reduce incoherency. This is the case with the factor measuring the monetary benefit of university which has a strong negative impact on Biased Incoherency for both loans and grants.

It is also important to note that the incidence of incoherency may depend on the type of financial aid. That is there are some factors which affect biased incoherency only with respect to one type of financial aid. For instance, post-secondary monetary costs decrease incoherency when choosing between grants and cash (the estimate is equal to -0.0573) but increase it when loans are concerned (the estimate is 0.0217).

Let's now consider the role of ex-ante enrollment probability. There is overwhelming evidence that those with a higher ex-ante probability of attending higher education also tend to display more Biased Incoherency. This is particularly true for loans as indicated by the estimate equal to 0.7700. Put differently, this implies that those who report high ex-ante attendance probabilities (who would normally need less financial aid) tend to be those who disclose a higher valuation when faced with real financial incentives.

Among the 5 primitive factors, the information factor is the only one found to be strongly correlated with the propensity to be incoherent. The estimates, equal to 0.2716 for loans and 0.3022 for grants, indicate that those who claim to be more informed also tend to disclose a higher loan valuation in the experiment than in the survey. Abstracting from the information factor, other primitive factors are much less correlated with incoherence. For instance, the factors measuring cognitive skills and motivation in school reduced biased Incoherency (after conditioning on costs and benefits). This is also true about the locus of control factor which reduces slightly the likelihood of Biased Incoherency. Finally, and after conditioning on costs and benefits, the financial constraint factor is found to have no impact.

One convenient method to summarize and illustrate the incidence of incoherencies is to compare the explanatory power of different groups of variables in the experiment and the survey valuations using R squares obtained when the regression of Table 11 is performed on specific subsets of determinants. This is achieved in Table 13 for different grouping methods: (i) monetary vs. non-monetary determinants, (ii) university costs and benefits vs. post-secondary costs and benefits, (iii) all benefits and costs vs. other factors, and finally (iv) all benefits vs. all costs. Among all those groupings, the university vs. post-secondary dimension is the most revealing one. The difference in R squares between survey and experiment for both loans and grants implies much more weight on university costs and benefits in the experiment valuations than in the survey, as the experiment R squares are about 4 to 5 times those obtained for the survey. Similarly, there is strong evidence that benefits and costs (when taken totally) are much more important in the experiment valuations than in the survey. When taken separately, we also find a much higher weight put to benefits in the experiment than in the survey. All those findings are fully applicable to both grants and loans.

[Table 13 here]

8 Policy Implications and Conclusion

Modern micro-econometric research makes use of a wide variety of data but for the most part, each study uses one of the following types: observational data, experimental (lab or field) or survey data. One major difference between various data sources is the amount of economic incentive involved. Because it is rarely possible to estimate the same parameters (or distribution of parameters) with different sources of data, economists are not capable of asserting whether or not structural parameters estimated with one data source would be coherent with those estimated from another source.

As the quality of individual decision making may vary with the presence of economic incentives involved, there are good reasons to question the validity of policy recommendations obtained in an incentive-free context. To see this, let's reconsider the hypothetical case where an economist is interested in evaluating ex-ante the impact of an expansion in higher education financial aid on university vs. other post-secondary enrollments such as 2-year colleges, but can only access data from the survey or from the field experiment. Estimating the determinants of individual valuations (essentially the willingness to pay for financial aid) could constitute a natural step as it should reveal if those who are willing to pay more have a high ex-ante probabilities of attending 2-year colleges or a high ex-ante probability to enroll in a university.

To achieve this, one way to proceed is to decompose individual valuations into subjective costs and benefits of both university and other post-secondary curriculums, as we did in this paper. When doing it using the survey, our estimates show that those endowed with higher monetary benefits and costs of a university curriculum would benefit more from financial aid expansion. However, estimates obtained in the experiment point to the exact opposite, as those with higher cost and benefits of university attendance are those benefiting less. While we agree that predicting what would be the impact of a new policy may require more information than the valuation decompositions into costs and benefits only, it is clear that the latter constitute valuable information. However, in this specific context, it is impossible to reconcile the results obtained from the survey with those in the experiment. Ex-ante policy evaluation of a potential expansion of the Canadian higher education financial aid system may depend heavily on whether or not the data have been obtained in an "incentivized" context.

Our findings raise an even more fundamental question. Given the existence of important incoherencies between survey data and field experiment data generated within at most 2 weeks, which type of data would be more able to predict actual post-high school economic outcomes? This is a research agenda that we plan to undertake in a near future.

References

Andersson, O., H.J. Holm, J.R. Tyran, and E. Wengström. 2020. "Robust Inference in Risk Elicitation Tasks." *Journal of Risk and Uncertainty* 61:195–209.

Arcidiacono, P. 2005. "Affirmative Action in Higher Education: How Do Admission and Financial Aid Rules Affect Future Earnings?" *Econometrica* 73:1477–1524.

Arcidiacono, P., V.J. Hotz, and S. Kang. 2012. "Modeling College Major Choices using Elicited Measures of Expectations and Counterfactuals." *Journal of Econometrics* 166:3–16.

Arrow, K., R. Solow, P.R. Portney, E.E. Leamer, R. Radner, and H. Schuman. 1993. "Report of the NOAA Panel on Contingent Valuation." *Federal Register* 58:4602–4614.

Ashworth, J., V.J. Hotz, A. Maurel, and T. Ransom. 2020. "Changes across Cohorts in Wage Returns to Schooling and Early Work Experiences." *Journal of Labor Economics* Forthcoming.

Attanasio, O., F. Cunha, and P. Jervis. 2019. "Subjective Parental Beliefs. Their Measurement and Role." Working Paper No. 26516, National Bureau of Economic Research, November.

Bajari, P., J.T. Fox, and S.P. Ryan. 2007. "Linear Regression Estimation of Discrete Choice Models with Nonparametric Distributions of Random Coefficients." *American Economic Review* 97:459–463.

Barseghyan, L., F. Molinari, and M. Thirkettle. 2021. "Discrete Choice under Risk with Limited Consideration." *American Economic Review* 111:1972–2006.

Belzil, C. 2007. "The return to schooling in structural dynamic models: a survey." *European Economic Review* 51:1059–1105.

Belzil, C., and J. Hansen. 2002. "Unobserved Ability and the Return to Schooling." *Econometrica* 70:2075–2091.

Belzil, C., and T. Jagelka. 2020. "Separating True Preferences from Noise and Endogenous Effort." Unpublished.

Belzil, C., A. Maurel, and M. Sidibé. 2021. "Estimating the Value of Higher Education Financial Aid: Evidence from a Field Experiment." *Journal of Labor Economics* 39:361–395.

Berger, J., A. Motte, and A. Parkin. 2007. *The Price of Knowledge: Access and Student Finance in Canada*, 3rd ed. The Canada Millennium Scholarship Foundation.

Boneva, T., and C. Rauh. 2019. "Socio-Economic Gaps in University Enrollment: The Role of Perceived Pecuniary and Non-Pecuniary Returns." Unpublished.

Cunha, F., I. Elo, and J. Culhane. 2020. "Maternal Subjective Expectations about the Technology of Skill Formation Predict Investments in Children one Year Later." *Journal of Econometrics* Forthcoming.

Delavande, A., and B. Zafar. 2019. "University Choice: The Role of Expected Earnings, Non-pecuniary Outcomes and Financial Constraints." *Journal of Political Economy* 127.

Eckstein, Z., and K.I. Wolpin. 1999. "Why Youths Drop out of High School: The Impact of Preferences, Opportunities, and Abilities." *Econometrica* 67:1295–1339.

Frenette, M. 2014. "An Investment of a Lifetime? The Long-term Labour Market Premiums Associated with a Postsecondary Education." *Analytical Studies Branch Research Paper Series* 359.

Gabaix, X. 2019. "Chapter 4 - Behavioral inattention." In B. D. Bernheim, S. DellaVigna, and D. Laibson, eds. *Handbook of Behavioral Economics - Foundations and Applications 2*. North-Holland, vol. 2 of *Handbook of Behavioral Economics: Applications and Foundations 1*, pp. 261–343.

Gneezy, U., J.A. List, J.A. Livingston, X. Qin, S. Sadoff, and Y. Xu. 2019. "Measuring Success in Education: The Role of Effort on the Test Itself." *American Economic Review: Insights* 1:291–308.

Gong, Y., L. Lochner, R. Stinebrickner, and T.R. Stinebrickner. 2019. "The Consumption Value of College." Working Paper No. 26335, National Bureau of Economic Research, October.

Harrison, G.W., and J.A. List. 2004. "Field Experiments." *Journal of Economic Literature* 42:1009–1055.

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:S197–S246.

Heckman, J.J., T. Jagelka, and T.D. Kautz. 2019. "Some Contributions of Economics to the Study of Personality." Working Paper No. 26459, National Bureau of Economic Research, November.

Heckman, J.J., and S. Navarro. 2007. "Dynamic Discrete Coice and Dynamic Teatment Efects." *Journal of Econometrics* 136:341–396, The interface between econometrics and economic theory.

Humphries, J.E., J.S. Joensen, and G.F. Veramendi. 2019. "Complementarities in High School and College Investments." Unpublished.

Jagelka, T. 2020. "Are Economists' Preferences Psychologists' Personality Traits? A Structural Approach." IZA Discussion Paper No. 13303, May.

Johnson, C., and C. Montmarquette. 2015. "The lack of Loan Aversion among Canadian High School Students." *Canadian Journal of Economics* 48:585–611.

Kaufmann, K.M. 2014. "Understanding the Income Gadient in College Attendance in Mexico: The Role of Heterogeneity in Expected Returns." *Quantitative Economics* 5:583–630.

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

—. 2001. "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment." *International Economic Review* 42:1051 – 1103.

Lochner, L.J., and A. Monge-Naranjo. 2016. "Student Loans and Repayment: Theory, Evidence, and Policy." Elsevier, vol. 5 of *Handbook of the Economics of Education*, chap. 8, pp. 397–478.

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

Manski, C.F., and F. Molinari. 2010. "Rounding Probabilistic Expectations in Surveys." *Journal of Business & Economic Statistics* 28:219–231, PMID: 20368764.

Matějka, F., and A. McKay. 2015. "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *American Economic Review* 105:272–98.

Steverson, K., A. Brandenburger, and P. Glimcher. 2019. "Choice-Theoretic Foundations of the Divisive Normalization Model." *Journal of Economic Behavior & Organization* 164:148–165.

Stinebrickner, R., and T.R. Stinebrickner. 2014. "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout." *The Review of Economic Studies* 81:426 – 472.

Todd, P.E., and K.I. Wolpin. 2021. "The Best of Both Worlds: Combining RCTs with Structural Modeling." *Journal of Economic Literature* Forthcoming.

Todd, P.E., and W. Zhang. 2020. "A Dnamic Model of Personality, Schooling, and Occupational Choice." *Quantitative Economics* 11:231–275.

Train, K.E. 2008. "EM Algorithms for Nonparametric Estimation of Mixing Distributions." *Journal of Choice Modelling* 1:40–69.

Williams, B. 2020. "Identification of the Linear Factor Model." *Econometric Reviews* 39:92–109.

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

Tables

Table 1: Financial Details of the Decisions between Cash and Higher Education Financing Packages

Choices	Cash (in $)	Loan (in $)	Grant (in $)	Financial aid take-up rates
1	25	2000	-	0.46
2	300	2000	-	0.19
3	700	2000	-	0.06
4	300	1000	-	0.12
5	300	4000	-	0.30
6	25	-	1000	0.90
7	100	-	1000	0.84
8	300	-	1000	0.72
9	700	-	1000	0.46
10	300	-	500	0.43
11	300	-	2000	0.79
12	300	-	4000	0.86
Sample size			705	
Ontario			361	
Manitoba			344	

Note: The take-up rates refer to the fraction of students who have chosen financial aid.

Table 2: Attendance Probabilities, Intertemporal Utilities and Model Fit

	University			Post-Secondary		
	Attendance probability	Lifetime utility	Consumption equivalence (in $ per year)	Attendance probability	Lifetime utility	Consumption equivalence (in $ per year)
Mean	0.66	50.6	63,500	0.31	48.8	53,173
St. Dev.	0.33	4.6	24,727	0.31	3.2	14,667
25th centile	0.39	47.1	45,121	0.02	46.5	43,152
Median	0.77	50.3	58,319	0.22	48.5	50,529
75th centile	0.98	53.4	74,651	0.53	50.9	61,002
Data (D19)	0.64	-	-	0.32	-	-
Data (D20)	0.74	-	-	0.24	-	-

Note 1: The utility of the outside option is normalized to the discounted sum of $34,000 over 20 years. The lifetime utility of the outside option is equal to 43.7.

Note 2: The consumption equivalence is measured at the amount such that its discounted sum over 20 years would equate the intertemporal utilities of university and post-secondary education.

Table 3: The Importance of Monetary Costs, Financial Constraint, Market Information, and Monetary Benefits on Life-cycle Consumption

	Changes in lifetime consumption (in \$)			
	During education		After education	
	$c^{u}_{i,sq}$	$c^{ps}_{i,sq}$	$c^{m,u}_{i,sq}$	$c^{m,ps}_{i,sq}$
Monetary Costs	15,603	20,299	-	-
Financial Constraint	6,770	18,910	-	-
Market Information	-10,242	-34,841	-	-
Monetary Benefits	-	-	-33,557	-121,257

Note: Monetary values for each component are calculated using the difference between the average consumption level in the population and the average consumption level obtained at the minimum (or maximum) level of each specific factor when shutting down the impact of each determinant.

Table 4: The Effects of Monetary and Non-Monetary Components on Intertemporal Utilities and Enrollment Probabilities

	Intertemporal utilities		Enrollment probabilities	
	University	Post-Secondary	University	Post-Secondary
Monetary Benefits	-1.0%	-3.8%	-0.0523	-0.1978***
Non-Monetary Benefits	-6.2%	-6.5%	-0.3371***	-0.2720***
Monetary Costs	2.0%	1.1%	0.1185***	0.0573***
Non-Monetary Costs	3.6%	3.7%	0.2110***	0.1847***
Financial Constraint	0.8%	1.1%	-0.0048	0.0180
Market Information	-1.0%	-1.8%	0.0313	-0.0449*
Cognitive Skills	-5.1%	-3.8%	-0.0899**	0.0672
Motivation in School	-3.0%	-1.6%	-0.0818***	0.0740***
Locus of Control	-0.7%	-0.8%	-0.0004	-0.0079

Note 1: Effects for each component are calculated using the average difference between the predicted utilities (resp. enrollment probabilities) and the utilities (resp. enrollment probabilities) obtained at the minimum (or maximum) level of each specific factor when shutting down the impact of each determinant.

Note 2: Significance levels: *** 1%; ** 5%; * 10%.

Table 5: Estimating the Reduced-Form Ex-Ante Enrollment Probabilities from Survey Data: Marginal Effects

	Potential Choices	
	University	Post-Secondary
	Marginal effects	Marginal effects
University:		
Monetary Costs (U)	-0.0209	0.0194
Non-Monetary Costs (U)	-0.0904***	0.0844***
Monetary Benefits (U)	0.0515**	-0.0479**
Non-Monetary Benefits (U)	0.1097***	-0.1019***
Post-Secondary:		
Monetary Costs (PS)	0.0229	-0.0250
Non-Monetary Costs (PS)	0.0918***	-0.1007***
Monetary Benefits (PS)	-0.0415**	0.0451**
Non-Monetary Benefits (PS)	-0.1248***	0.1347***
Individual Factors:		
Cognitive Skills	0.0853***	-0.0722***
Motivation in School	0.0362***	-0.0271**
Locus of Control	-0.0056	0.0088
Financial Constraint	-0.0115	0.0031
Market Information	0.0000	0.0078

Note 1: The marginal effect is the effects of a one standard deviation increase in the factor on the enrollment probability.

Note 2: Standard errors in parentheses. Significance levels: *** 1%; ** 5%; * 10%.

Table 6: The Welfare Gains of Providing Financial Aid

Panel A: Providing Free Higher-Education

	Policies		
	Free Post-Sec.	Free Univ.	Free education
Welfare gain	0.1258%	1.1229%	1.2106%
Δ in Post-Sec. enrollments	0.0231	-0.0894	-0.0695
Δ in University enrollments	-0.0200	0.0983	0.0801

Note 1: The welfare gain is computed from the % change in expected maximum utility reachable from period 1.

Note 2: Free post-secondary and free university refer to policies providing an annual $4,000 grant or loan for the tenure of each type of education. Free education refers to a policy that provides the same annual amount for both post-secondary and university.

Panel B: Providing $2,000 Financial Aid Packages

	Policies	
	$2,000 Grant	$2,000 Loan
Welfare gain	0.1752%	0.0881%
Δ in Post-Sec. enrollments	-0.0071	-0.0074
Δ in University enrollments	0.0091	0.0084

Note 1: The welfare gain is computed from the % change in expected maximum utility reachable from period 1.

Table 7: The Utility of Consumption in Period 0

	Reference consumption c_i^r (in \$)	Marginal utility (\$300) $MU_{i,0}$(\$300)
Mean	11.34	3.4425
St. Dev.	6.33	0.5034
25th centile	7.08	3.1679
Median	10.15	3.4200
75th centile	13.18	3.7694

Table 8: The Distribution of Intertemporal Marginal Utilities in the Survey and in the Field Experiment: a \$2,000 Grant and a \$2,000 Loan

	MV_g (Grant = \$2,000)		MV_l (Loan = \$2,000)	
	Survey	Field Experiment	Survey	Field Experiment
Mean	0.0937	5.2665	0.0472	1.3087
St. Dev.	0.0335	1.5992	0.0327	1.0541
5th centile	0.0467	2.6816	0.0019	-0.5110
25th centile	0.0705	4.2245	0.0239	0.6921
Median	0.0906	5.3419	0.0428	1.4986
75th centile	0.1126	6.3803	0.0642	2.0423
95th centile	0.1556	7.4715	0.1064	2.5655

Table 9: Correlations between Different Valuations

Correlations	Level	Rank
Between survey and field experiment:		
MV_g^{fe} and MV_g^s	0.3297	0.3374
MV_l^{fe} and MV_l^s	0.4100	0.4158
Between grants and loans:		
MV_g^{fe} and MV_l^{fe}	0.9189	0.8915
MV_g^s and MV_l^s	0.9948	0.9914

Table 10: The Determinants of the Valuations of Financial Aid in the Survey and in the Field Experiment

Dependent variable	MV_g (Grant)		MV_l (Loan)	
	Survey	Field Experiment	Survey	Field Experiment
	Parameters		Parameters	
Intercept	-0.0640***	-1.4712**	-0.0371***	-1.9907**
Monetary Costs (U)	**0.0112******	**-0.2165******	**0.0138******	**-0.1820****
Non-Monetary Costs (U)	-0.0049***	-0.1062***	**-0.0059******	**0.3415******
Monetary Benefits (U)	**0.0124******	**-2.1559******	**0.0185******	**-1.6261******
Non-Monetary Benefits (U)	0.0004	1.7625***	0.0016***	1.3191***
Monetary Costs (PS)	0.0032***	0.1149**	0.0040***	0.1302**
Non-Monetary Costs (PS)	**0.0051******	**-0.4541******	**0.0057******	**-0.3608******
Monetary Benefits (PS)	**-0.0098******	**0.0284**	**-0.0104******	**0.3675******
Non-Monetary Benefits (PS)	-0.0032***	-0.1508**	-0.0040***	-0.2297***
Cognitive skills	0.0022***	0.1989***	0.0030***	0.0098
Motivation in School	0.0026***	0.1552***	0.0031***	0.2572***
Locus of Control	-0.0014***	-0.0906**	-0.0017***	-0.1560***
Financial Constraint	0.0063***	0.2758***	0.0078***	0.2704***
Market Information	**-0.0186******	**0.1932******	-0.0231***	-0.0853**
Enrollment Prob. $(\Lambda_i^u + \Lambda_i^{ps})$	0.0839***	3.4275***	0.0569***	2.6716***
$\mathbb{1}(\Lambda_i^u > \Lambda_i^{ps})$	0.0274***	1.5634***	0.0319***	1.1041***
Grant = \$500 (reference)	-	-	-	-
Grant = \$1,000	0.0235***	1.4052***	-	-
Grant = \$2,000	0.0701***	2.2870***	-	-
Grant = \$4,000	0.1616***	3.0118***	-	-
Loan = \$1,000 (reference)	-	-	-	-
Loan = \$2,000	-	-	0.0233***	0.0883**
Loan = \$4,000	-	-	0.0678***	1.2972***

Note 1: Significance levels: *** 1%; ** 5%; * 10%.

Note 2: Determinants in bold are those for which the effect measured in the field experiment is of a different sign in the survey.

Table 11: Evaluating the Coherency between the Valuations of Financial Aid in the Survey and in the Field Experiment

Dependent variable	Rank of MV_g (Grant = \$2,000)			Rank of MV_l (Loan = \$2,000)		
	Survey	Field Experiment	Equality	Survey	Field Experiment	Equality
	Parameters			Parameters		
Intercept	42.9406***	45.4323***		40.6087***	40.9367***	
Rank Mon. Costs (U)	0.2540***	-0.0892***	reject	0.2519***	-0.1232***	reject
Rank Non-Mon. Costs (U)	-0.2138***	-0.1842***	reject	-0.1810***	0.0446***	reject
Rank Mon. Benefits (U)	0.2361***	-0.4299***	reject	0.2665***	-0.4555***	reject
Rank Non-Mon. Benefits (U)	0.2133***	0.8149***	reject	0.2180***	0.9034***	reject
Rank Mon. Costs (PS)	0.1398***	0.0625***	reject	0.1373***	0.1521***	
Rank Non-Mon. Costs (PS)	0.0181***	-0.2602***	reject	0.0422***	-0.3469***	reject
Rank Mon. Benefits (PS)	-0.0842***	-0.1236***	reject	-0.0770***	0.0039	reject
Rank Non-Mon. Benefits (PS)	-0.2420***	-0.1428***	reject	-0.2469***	-0.2199***	reject
Rank Cognitive	0.3201***	0.2303***	reject	0.2749***	0.1824***	reject
Rank Motivation in School	0.2437***	0.1488***	reject	0.1986***	0.3453***	reject
Rank Locus of Control	-0.0685***	-0.1227***	reject	-0.0608***	-0.2070***	reject
Rank Financial Constraint	0.1207***	0.0938***	reject	0.1448***	0.1573***	
Rank Market Information	-0.5740***	0.1833***	reject	-0.6156***	-0.0365***	reject
Rank ex-ante Prob. $(\Lambda_i^u + \Lambda_i^{ps})$	-0.2221***	-0.0893***	reject	-0.1652***	-0.2187***	reject
R square	0.80	0.81		0.83	0.73	
Overall Coherency Measure	0.15			0.25		

Note 1: The rank measures used for the left-hand side variable (MV) and the determinants (the factors) are the quantiles (ranging from 1 to 100) corresponding to each type.

Note 2: Significance levels: *** 1%; ** 5%; * 10%.

Note 3: Estimates with a “reject” are the determinants for which the null hypothesis that the deviation parameter is equal to 0 is actually rejected at a 1% level.

Table 12: The Determinants of Absolute and Biased Incoherency

Type of incoherency	Absolute $\lvert \text{Rank}(MV_g^{fe}) - \text{Rank}(MV_g^{s}) \rvert$		Biased $\mathbb{1}(MV_g^{fe} - MV_g^{s}) > 0$	
	Grants	Loans	Grants	Loans
Intercept	91.4869***	56.7085***	-0.1166**	-0.4185***
Monetary Costs (U)	-2.7003***	-3.1033***	-0.1059***	-0.0899***
Non-Monetary Costs (U)	6.2972***	5.2658***	0.0140*	0.1237***
Monetary Benefits (U)	-5.2649***	-7.6530***	-0.2424***	-0.3743***
Non-Monetary Benefits (U)	2.6017***	4.0483***	0.2775***	0.3388***
Monetary Costs (PS)	-2.6313***	-0.4757*	-0.0573***	0.0217***
Non-Monetary Costs (PS)	2.7227***	-0.9541***	-0.1441***	-0.2302***
Monetary Benefits (PS)	9.5488***	8.7160***	0.1431***	0.0867***
Non-Monetary Benefits (PS)	-0.8671**	-2.9328***	-0.0565***	-0.1097***
Cognitive	1.6973***	0.7038***	-0.0414***	-0.0895***
Motivation in School	0.6457***	0.4487**	-0.0115***	0.0587***
Locus of Control	0.1754	0.6278***	-0.0261***	-0.0295***
Financial Constraint	2.3981***	0.3758*	0.0079*	0.0208***
Market Information	2.4643***	0.5194***	0.3022***	0.2716***
Enrollment Prob. ($\Lambda_i^u + \Lambda_i^{ps}$)	-77.5547***	-42.3704***	0.5446***	0.7700***
$\mathbb{1}(\Lambda_i^u > \Lambda_i^{ps})$	8.4174***	6.1305***	-0.1366***	-0.0235**
R square	0.16	0.09	0.62	0.60

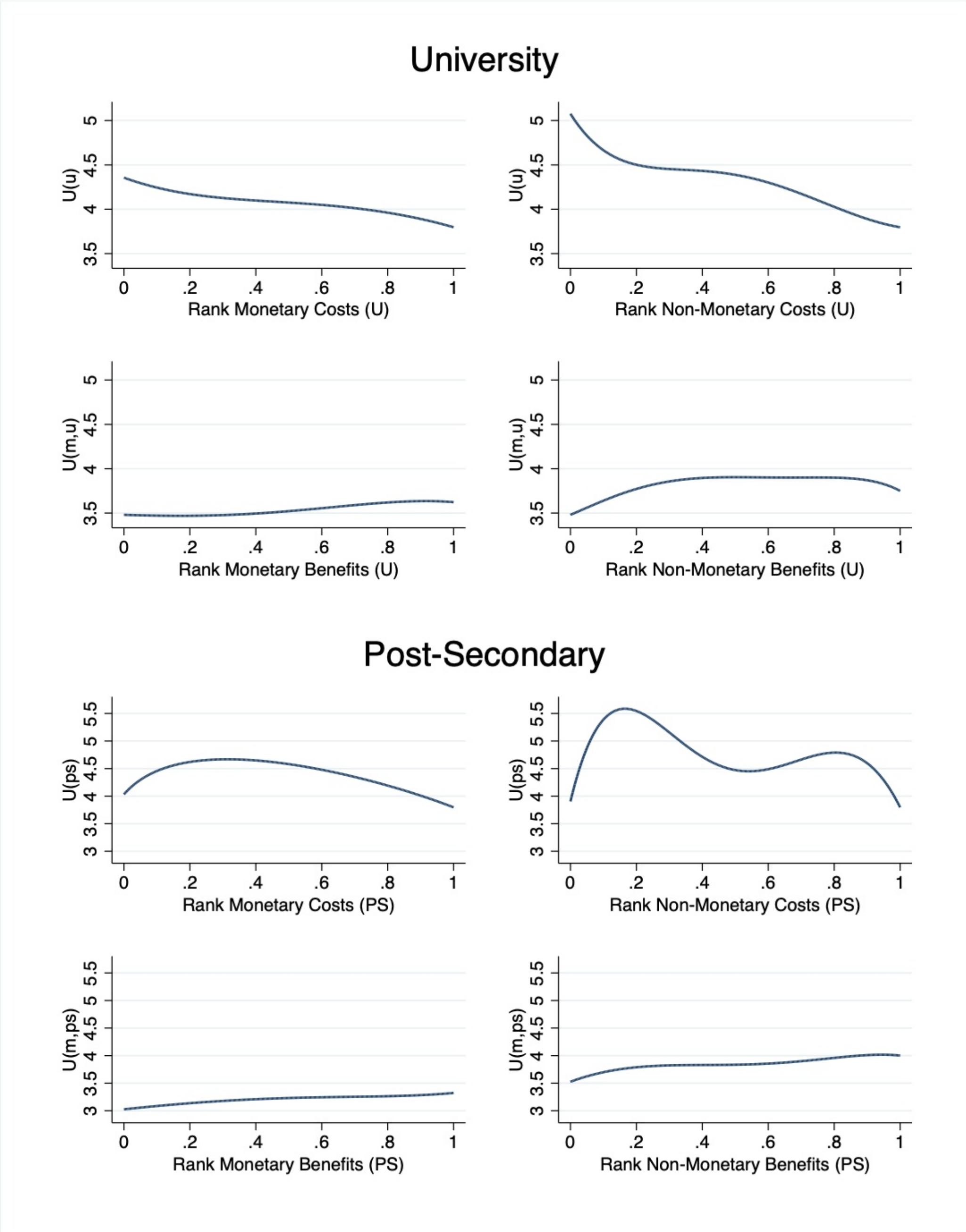
Note 2: Significance levels: *** 1%; ** 5%; * 10%.

Table 13: Relative Importance of Different Determinants in the Survey and the Field Experiment

	Grants		Loans	
	Survey	Field Experiment	Survey	Field Experiment
Groups of variables	R square	R square	R square	R square
All monetary	0.31	0.30	0.33	0.18
All non-monetary	0.35	0.62	0.36	0.47
All university	0.15	0.63	0.15	0.52
All post-secondary	0.24	0.22	0.24	0.14
All benefits and costs	0.44	0.73	0.44	0.62
All other factors	0.60	0.41	0.61	0.33
All benefits	0.36	0.58	0.36	0.56
All costs	0.18	0.28	0.17	0.10
All determinants	0.80	0.81	0.83	0.73

Figures

Figure 1: Effect of Costs and Benefits Factors Ranks



Supplementary Material for Online Publication

A Measurement of Factors and Higher Education Enrollment Intentions

The vast majority of the questions are contained in the Student Survey and a few of them are contained in the parental survey. The measurements are identified by the corresponding question in the student and parental booklets. A large number of questions had to be answered within an ordered discrete choice while others are coded as binary outcomes. We now provide the details for each question.

A.1 Primitive Factors

A.1.1 Cognitive Skill

Question D1: In your last year of high school, what was your overall grade average, as a percentage?

[90% to 100% / 80% to 89% / 70% to 79% / 60% to 69% / 55% to 59% / 50% to 54% / Less than 50%]

Question D2: How would you rate your ability to use a computer? For example, using software applications, programming, or using a computer to find or process information.

[Poor / Good / Very good / Excellent]

Question D3: How would you rate your writing abilities? For example, writing to get across information or ideas to others, or editing writing to improve it.

[Poor / Good / Very good / Excellent]

Question D4: How would you rate your reading abilities? For example, understanding what you read and identifying the most important issues, or using written material to find information.

[Poor / Good / Very good / Excellent]

Question D5: How would you rate your oral communication abilities? For example, explaining ideas to others, speaking to an audience, or participating in discussions.

[Poor / Good / Very good / Excellent]

Question D6: How would you rate your ability to solve new problems? For example, identifying problems and possible causes, planning strategies to solve problems, or thinking of new ways to solve problems.

[Poor / Good / Very good / Excellent]

Question D7: How would you rate your mathematical abilities? For example, using formulas to solve problems, interpreting graphs or tables, or using math to figure out practical things in everyday life.

[Poor / Good / Very good / Excellent]

Numeracy Test Score provided by Statistics Canada: Over the day of the experiment, a numeracy test provided by the Center for Education Statistics was administered to all students. The test was based on the numerical component of the International Adult Literacy and Skills Survey project undertaken by numerous OECD countries between 1995 and 2005. It is the same test used in PISA international comparisons. The questions are meant to capture the capacity to perform numerical calculations. Students received a score between 0 and 500, which is used as a cognitive ability measure. We use its standardized value.

A.1.2 Motivation in School

For the next questions, think only about your last year in high school, that is your most recent year there.

Question E1: During your last year in high school, about how many hours each week did you spend on homework outside class, during free periods and at home?

[Zero hours / Less than one hour per week / 1 to 3 hours / 4 to 7 hours / 8 to 14 hours / 15 hours or more]

Question E3: I did as little work as possible; I just wanted to get by.

[Never / Rarely / Some of the time / Most of the time / All of the time]

Question E5: I was interested in what I was learning in class.

[Never / Rarely / Some of the time / Most of the time / All of the time]

Question E7: I completed my homework on time.

[Never / Rarely / Some of the time / Most of the time / All of the time]

Question E8: I thought that many of the things we were learning in class were useless.

[Strongly Disagree / Disagree / Agree / Strongly Agree]

Question E11: School was often a waste of time.

[Strongly Disagree / Disagree / Agree / Strongly Agree]

Question E13: How often did you cut or skip a class without permission? Was it...?

[Never / Less than once a month / Once or twice a month / About once a week / More than once a week]

A.1.3 Level of Mastery (Locus of Control)

The next set of questions describes the way some people feel. After each statement please indicate whether you strongly disagree, disagree, agree or strongly agree.

Question G1a: You have little control over the things that happen to you.

[Strongly Disagree / Disagree / Agree / Strongly Agree]

Question G1b: There is really no way you can solve some of the problems you have.

[Strongly Disagree / Disagree / Agree / Strongly Agree]

Question G1c: There is little you can do to change many of the important things in your life.

[Strongly Disagree / Disagree / Agree / Strongly Agree]

Question G1d: You often feel helpless in dealing with the problems of life.

[Strongly Disagree / Disagree / Agree / Strongly Agree]

Question G1e: Sometimes you feel that you are being pushed around in life.

[Strongly Disagree / Disagree / Agree / Strongly Agree]

Question G1f: What happens to you in the future mostly depends on you.

[Strongly Disagree / Disagree / Agree / Strongly Agree]

Question G1g: You can do just about anything you really set your mind to do.

[Strongly Disagree / Disagree / Agree / Strongly Agree]

A.1.4 Information about Labor Market

Question D23: In the last two years, has anyone at school helped you find information about jobs you may be interested in when you finish all your schooling?

[Yes / No]

Question D24a: In the last two years at school, did you meet with a school counselor about your future education or work?

[Yes / No]

Question D24b: In the last two years at school, did you complete a questionnaire to find out about your interests or abilities?

[Yes / No]

Question D24c: In the last two years at school, did you obtain any information on student financing, such as student loans or grants?

[Yes / No]

Question D24d: In the last two years at school, did you attend a presentation by people working in different types of jobs?

[Yes / No]

A.1.5 Financial Constraint

Question K1: How important are low expenses (tuition, books, room and board) in choosing a college or university you would like to attend?

[Not important / Somewhat important / Very important]

Question K2: How important is availability of a scholarship, or grant in choosing a college or university you would like to attend?

[Not important / Somewhat important / Very important]

Question K12: In the past 12 months, would you say that your family spent:

[More than they earned / As much as they earned / Less than they earned / I don't know]

Question K13: Do you consider your personal level of debt to be a burden?

[Yes / No]

Question K14: Do you consider your family's level of debt to be a burden?

[Yes / No]

Question K15: Have you personally saved any money to finance your education after high school?

[Yes / No]

Question D10: I feel obligated to financially support my family while I am still in school.

[Strongly Disagree / Disagree / Agree / Strongly Agree]

Question D21: Is there anything standing in your way of going as far in school as you would like to go?

[Several options listed, up to 3 can be chosen. We coded it as binary, with value 1 if "financial situation" was chosen.]

Question Q5 (parental survey): Is your financial situation standing in your child's way of going that far?

[Several options listed. We coded it as binary, with value 1 if "financial situation" was chosen.]

Question Q6 (parental survey): Have you done anything specific to ensure that your child will have money for further education after high school?

[Yes / No]

Question F3: What is your best estimate of the total annual income before taxes and deductions for your entire family (including parents and other relatives living with you) for 2007?

Question Q9 (parental survey): Would you say the total income, before deductions, of all family members living in your household in 2007 was. . . ?

[The response is an interval of $10,000 (from less than $10,000 to $120,000 or more)]

A.2 Valuation of Costs and Benefits of Higher Education

For each of the following statements, respondents have to indicate how much they agree with the following statement, separately for three types of post-secondary education:

- University,
- Community college or CEGEP,
- Trade/vocational school or registered apprenticeship.

The "......" in the questions below are replaced by each type of post-secondary education. In our model, we use measures related to "University" to extract the costs and benefits factors related to university, and the measures related to "Community college or CEGEP" and "Trade/vocational school or registered apprenticeship" to extract factors related to other post-secondary education.

Answers are picked among the following options:

[Strongly Disagree / Disagree / Uncertain / Agree / Strongly Agree]

A.2.1 Monetary Benefits

Question H1a: "People who get a ...... education will make more money over their lifetime than those who just get a high school education".

Question H1i: "Although ...... can be costly, I believe that I would make more money in the long run".

Question H1l: "I think that if I were to put the time and effort into getting a good ...... education, I would make a lot more money in the long run".

Question H1p: "I am confident that a ...... education would lead me to a better paying job".

A.2.2 Monetary Costs

Question H1f: "I'm not sure that a ...... education would pay off even in the long run, given how costly it is these days".

Question H1r: "Given the high costs of a ...... education and the time it takes to complete it, you are really no further ahead financially than if you get a job right after high school".

Question H1t: "I am hesitant to undertake a ...... education because of the amount of debt I'm likely to accumulate by the time I graduate".

Question H1w: "The costs of a ...... education have become so high that they outweigh any future financial benefits".

A.2.3 Non-Monetary Benefits

Question H1d: "Getting a ...... education will lead me to find work that I really enjoy doing".

Question H1v: "The best way to get a prestigious job is through a ...... education".

Question H1g: "If you want a rewarding career these days, you need a ...... education".

Question H1h: "People who have a ...... education get jobs that are much more satisfying".

Question H1n: "I think I could find a rewarding job without a ...... education".

Question H1s: "Good jobs can be found without a . . . . . . education".

Question H1u: "You can learn enough about the real world without a . . . . . . education".

Question H1x: "I don't think I would ever find fulfilling work if I didn't get a . . . . . . education".

A.2.4 Non-Monetary Costs

Question H1b: "I don't feel that I am emotionally prepared to go to . . . . . . yet".

Question H1e: "If I were to pursue a . . . . . . education, my friends would think that I'm trying to be better than them".

Question H1j: "If I pursued . . . . . . education, I'm afraid that it would confuse me about "who I am"".

Question H1k: "I'm hesitant to pursue a . . . . . . education because it would create tensions with the people I grew up with".

Question H1m: "I'm hesitant to pursue a . . . . . . education because it would create tensions between my parents and me".

Question H1o: "I'm hesitant to pursue a . . . . . . education because I really don't know what I want to do with my life yet".

Question H1q: "I don't think that I have the correct mindset right now to tackle a . . . . . . program".

A.3 Higher Education Enrollment Intentions

Question D19: As things stand now, what is the highest level of education you think you will get?

[9 options listed. We grouped them in three categories: No post-secondary education / University / Other post-secondary]

Question D20: What is the highest level of education you would like to get?

[9 options listed. We grouped them in three categories: No post-secondary education / University / Other post-secondary]

B The Distribution of Primitive Factors and Subjective Benefits and Costs

B.1 Normalizations

In order to estimate the distribution of primitive factors, we use the following normalizations:

- For cognitive skill measurements, we set the loading parameter associated to the Numeracy Test to 1 and the intercept to 0.
- For motivation in school, we set the loading parameter of measure E1 to 1 and the intercept to 0.
- For locus of control, we set the loading parameter of measure G1a to 1 and the intercept to 0.
- For labor market information, we set the loading parameter of measure D23 to 1 and the intercept to 0.
- For financial constraint, we set the loading parameter of measure K2 to 1 and the intercept to 0.

The following normalizations are used to estimate the distribution of subjective benefits and costs factors:

- For monetary benefits, we set the loading parameter of measure H1a to 1 and the intercept to 0.
- For monetary costs, we set the loading parameter of measure H1w to 1 and the intercept to 0.
- For non-monetary benefits, we set the loading parameter of measure H1d to 1 and the intercept to 0.
- For non-monetary costs, we set the loading parameter of measure H1q to 1 and the intercept to 0.

B.2 Empirical Estimates

As a first step, we estimated the distribution of the unobserved primitive factors. Table B.1 discloses the correlation between the factors. Not surprisingly, we find that the cognitive skill factor (CO) is positively correlated with motivation in school (MS) and with the locus of control (LC), with correlations respectively equal to 0.448 and 0.335.

The liquidity constraint perception factor (FC) and the market information (MI) factors raise particular interest as they most likely never have been estimated using factor model techniques. In general, the financial constraints factor (FC) tends to be relatively weakly correlated with other factors. For instance, the correlation with motivation in school is only 0.120. It is also practically orthogonal to the cognitive factor (the correlation is -0.015).

Table B.1: The Distribution of Primitive Factors: Correlation Table

	Cognitive skills CO	Locus of control LC	Motivation in school MS	Labor market information MI	Financial constraints FC
CO	1.000				
LC	0.335	1.000			
MS	0.448	0.272	1.000		
MI	0.122	0.130	0.170	1.000	
FC	-0.015	-0.154	0.120	0.053	1.000

The Market Information factor (MI) factor is positively correlated with both the cognitive factor and the motivation factor (the correlations are 0.122 and 0.170). These indicate essentially that those more likely to attend higher education tend to be more informed about financial aid opportunities.

Table B.2 reports the correlation between the subjective benefits and costs factors, for university (Panel A) and other post-secondary education (Panel B).

As normally expected, benefits and costs are negatively correlated. This is the case for both university and post-secondary education.

The correlation between monetary benefits and costs, equal to -0.385 for university and -0.218 for other type of post-secondary institution, as well as the correlations between non-monetary benefits and costs, equal to -0.253 and -0.416, are found to be quite strong. Overall, those reporting high benefits therefore also tend to be those expecting low costs of education.

At the same time, those endowed with high monetary benefit expectations also tend to be endowed with high non-monetary benefit expectations, as indicated by correlations equal to 0.644 (university) and 0.561 (post-secondary). The correlations between monetary and non-monetary costs tend to be even stronger (with correlations equal to 0.697 and 0.512).

Because our model assumes that primitive factors are mapped onto benefits on costs, and because the latter are naturally seen as structural components of individual decisions, it is important to evaluate to what extent qualitative benefits and costs are explained by primitive factors.

In Table B.3, we report the standard deviation of each cost and benefit factor as well as the standard deviation of each factor's deterministic portion which is explained by the primitive factors. By comparing each standard deviation, it is easy to determine to what extent subjective costs and benefits may be viewed as separate factors.

The results validate our approach which consisted in allowing each cost and benefit factor to be explained by its own unobserved heterogeneity. For all 8 factors, we find that the primitive factors explain at most 50% of the total variation across types. In some cases such as university and post-secondary monetary benefits, the primitive factors account for at most 25%. In many others, they account for one third only.

In Table B.4, we also report the structural parameters of the function mapping factors onto benefits and costs. Overall, the results are coherent with pure intuition. For instance, motivation in school is found to have a positive impact on both monetary and non-

Table B.2: The Distribution of Benefits and Costs: Correlation Tables

Panel A: University

	Monetary Benefits MB (U)	Non-Monetary Benefits NMB (U)	Monetary Costs MC (U)	Non-Monetary Costs NMC (U)
MB (U)	1.000			
NMB (U)	0.644	1.000		
MC (U)	-0.385	-0.241	1.000	
NMC (U)	-0.251	-0.253	0.697	1.000

Panel B: Post-Secondary

	Monetary Benefits MB (PS)	Non-Monetary Benefits NMB (PS)	Monetary Costs MC (PS)	Non-Monetary Costs NMC (PS)
MB (PS)	1.000			
NMB (PS)	0.561	1.000		
MC (PS)	-0.218	-0.012	1.000	
NMC (PS)	-0.220	-0.416	0.512	1.000

Table B.3: The Distribution of Benefits and Costs: Importance of the Primitive Factors

	Sources of variation	
	All	Primitive factors
	St. dev.	St. dev.
University:		
Monetary Benefits (U)	0.4952	0.1320
Non-Monetary Benefits (U)	0.6245	0.2289
Monetary Costs (U)	0.8033	0.3407
Non-Monetary Costs (U)	0.6923	0.3797
Post-Secondary:		
Monetary Benefits (PS)	0.4057	0.1139
Non-Monetary Benefits (PS)	0.6678	0.1768
Monetary Costs (PS)	0.9744	0.2905
Non-Monetary Costs (PS)	0.8515	0.2401

monetary benefits and a negative effect on non-monetary costs.

Table B.4: The Mapping from Primitive Factors onto Perceived Benefits and Costs: Marginal Effects

Panel A: University

	Monetary Benefits MB (U)	Non-Monetary Benefits NMB (U)	Monetary Costs MC (U)	Non-Monetary Costs NMC (U)
Cognitive skills	0.0595***	-0.0301	-0.2293***	-0.2480***
	(0.0172)	(0.0205)	(0.0252)	(0.0225)
Locus of control	-0.0026	-0.0055	-0.2170***	-0.2672***
	(0.0158)	(0.0187)	(0.0234)	(0.0212)
Motivation in school	0.2283***	0.3769***	-0.1189***	-0.2260***
	(0.0173)	(0.0219)	(0.0247)	(0.0226)
Labor market information	0.0304**	0.0192	0.0585***	0.0830***
	(0.0154)	(0.0179)	(0.0206)	(0.0191)

Panel B: Post-Secondary

	Monetary Benefits MB (PS)	Non-Monetary Benefits NMB (PS)	Monetary Costs MC (PS)	Non-Monetary Costs NMC (PS)
Cognitive skills	-0.1777***	-0.2400***	-0.1398***	0.0012
	(0.0153)	(0.0134)	(0.0147)	(0.0108)
Locus of control	0.2235***	0.0959***	-0.1903***	-0.2753***
	(0.0143)	(0.0117)	(0.0131)	(0.0103)
Motivation in school	0.1483***	0.2279***	-0.0529***	-0.0108
	(0.0147)	(0.0126)	(0.0135)	(0.0104)
Labor market information	0.0608***	0.0447***	-0.0048	-0.0241***
	(0.0132)	(0.0109)	(0.0116)	(0.0090)

Note 1: The marginal effect is the effects of a one standard deviation increase in the primitive factor on the cost or benefit factor, normalized by the standard deviation of the cost or benefit factor.

Note 2: Standard errors in parentheses. Significance levels: *** 1%; ** 5%; * 10%.

C Estimating the Reference Consumption Level from Data on Lottery Binary Choices

Prior to exerting decisions between cash payments and financial aid, students were presented with a sequence of binary choices between two lotteries in which risk is objectively stated. The strategy used by the designers of the experiment was a standard Multiple Price List (MPL) design. It consists of choosing between a lottery with average payoff and another one with extreme payoff, and identify the cutoff point where an agent switches from the average to the extreme lottery. This approach, pioneered by Holt and Laury (2005), is standard in the experimental literature to measure risk aversion, but it may also be used to estimate individual reference consumption level.

Each risk aversion decision requires to choose between two lotteries, X_l and Y_l. In each case, the first lottery is unambiguously less risky then the second one. The first lottery is characterized by a low payoff, denoted x_{1l} and a high payoff, denoted x_{2l}, while the second lottery entails a low payoff, denoted y_{1l}, and a high payoff, denoted y_{2l}. For each lottery, the probability of the high outcome is equal to p_{2l} and the probability of the low outcome is $p_{1l} = 1 - p_{2l}$.

When estimating the benchmark consumption level, we make use of 27 decisions. Each decision is summarized in Table C.5. The decision identifier found in column 1 corresponds to the rank that each decision occupies in the original data file.

Table C.5: Lottery Payoffs and Probabilities

Decision #	x_{1l}	x_{2l}	y_{1l}	y_{2l}	p_{2l}
49	32	40	2	77	0.1
50	32	40	2	77	0.2
51	32	40	2	77	0.3
52	32	40	2	77	0.4
53	32	40	2	77	0.5
54	32	40	2	77	0.6
55	32	40	2	77	0.7
56	32	40	2	77	0.8
57	32	40	2	77	0.9
59	24	30	1.50	57.75	0.1
60	24	30	1.50	57.75	0.2
61	24	30	1.50	57.75	0.3
62	24	30	1.50	57.75	0.4
63	24	30	1.50	57.75	0.5
64	24	30	1.50	57.75	0.6
65	24	30	1.50	57.75	0.7
66	24	30	1.50	57.75	0.8
67	24	30	1.50	57.75	0.9
69	40	50	2.50	96.25	0.1
70	40	50	2.50	96.25	0.2
71	40	50	2.50	96.25	0.3
72	40	50	2.50	96.25	0.4
73	40	50	2.50	96.25	0.5
74	40	50	2.50	96.25	0.6
75	40	50	2.50	96.25	0.7
76	40	50	2.50	96.25	0.8
77	40	50	2.50	96.25	0.9