I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17174

Identification of Ex Ante Returns Using Elicited Choice Probabilities: An Application to Preferences for Public-Sector Jobs

Romuald Méango
Esther Mirjam Girsberger

JULY 2024

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17174

Identification of Ex Ante Returns Using Elicited Choice Probabilities: An Application to Preferences for Public-Sector Jobs

Romuald Méango
University of Oxford and CESifo

Esther Mirjam Girsberger
University of Technology Sydney and IZA

JULY 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Identification of Ex Ante Returns Using Elicited Choice Probabilities: An Application to Preferences for Public-Sector Jobs*

Ex ante returns, the net value that agents perceive before they take an investment decision, are understood as the main drivers of individual decisions. Hence, their distribution in a population is an important tool for counterfactual analysis and policy evaluation. This paper studies the identification of the population distribution of ex ante returns using stated choice experiments, in the context of binary investment decisions. The environment is characterised by uncertainty about future outcomes, with some uncertainty being resolved over time. In this context, each individual holds a probability distribution over different levels of returns. The paper provides novel, nonparametric identification results for the population distribution of returns, accounting for uncertainty. It complements these with a nonparametric/semiparametric estimation methodology, which is new to the stated-preference literature. Finally, it uses these results to study the preference of high ability students in Côte d'Ivoire for public-sector jobs and how the competition for talent affects the expansion of the private sector.

JEL Classification: C21, C23, D84, J21, J24, J30, J45

Keywords: subjective expectations, ex ante returns, nonseparable panel, distribution regression, job search, public sector

Corresponding author:
Romuald Méango
Department of Economics
University of Oxford
Manor Road Building
Manor Road
Oxford, OX1 3UQ
United Kingdom

E-mail: romuald.meango@economics.ox.ac.uk

* Romuald acknowledges the support from the Higher Studies Fund - Travel Fund for the Economics of Developing Countries of the University of Oxford, the Centre for Research on Structural Transformation and Growth (CROSTAG) in the Department of Economics at the University of Oxford, which in turn receives funding from the Centre for Study of African Economies (CSAE) at Oxford and the research programme on Structural Transformation and Economic Growth (STEG). Romuald acknowledges support from the AfOX Catalyst Grant, grant number AfOx-233. We thank Richard Moussa for supervising the data collection, Stéphane Bonhomme and three anonymous referees, Adeline Delavande, Marc Henry, Ismaël Mourifié, François Poinas, Sami Stouli, and Martin Weidner and participants to various seminars for useful comments. The Institutional Review Board at the Department of Economics of the University of Oxford approved this study. All remaining errors are ours.

1. Introduction

Suppose that, in a context of skill shortage and competition for talent, an analyst seeks to understand the effect of the labour demand from one sector, say the public sector, on the demand from another sector, say the private sector. One measure of interest is the cost for the private sector to attract additional workers who would otherwise choose to enter the public sector. If, all things being equal, an increase of its labour force by one percent would increase the private-sector wage bill by a similar proportion, then the private sector does not suffer a significant competition from the public sector. Conversely, if this cost elasticity is significantly greater than one, then the competition from the public sector generates high labour costs for the private sector and potentially limits its expansion.

Measuring the aforementioned elasticity requires an understanding of job-seekers' *ex ante returns* on choosing a public job offer rather than a private job offer. Ex ante returns can be defined as the minimum private-sector wage increase that induces the worker to choose the private sector offer. Those returns, say S, encompass both pecuniary and non-pecuniary gains from choosing the public sector offer.[1] Assume that F_S is the distribution of those ex ante returns in the population of job-seekers, and any job-seeker chooses the public sector if they perceive positive ex ante returns. The marginal individual, i.e. who is indifferent between the private and public-sector job, is at the $F_S(0) - th$ percentile. To attract x additional job-seekers, the private sector must compensate up to the job-seeker at $F_S(0) + x - th$ percentile by transferring $F_S^{-1}\Big[F_S(0) + x\Big]$ to this individual. If the private sector is unable to observe workers' private returns and discriminate among them, then it must transfer the same amount even to infra-marginal job-seekers. Therefore, the cost of an expansion of the private sector is the amount of each transfer times the proportion of workers who choose to work in the private sector, as illustrated by the shaded area in Figure 1.1. The key for conducting this analysis is understanding the distribution $F_S(s), s \in \mathbb{R}$.

Estimating a population distribution of ex ante returns F_S in the context of binary investment decisions is the main goal of this paper. The distribution of returns can be used to describe agents' preferences or calculate ex ante policy parameters that require knowledge of different margins of indifference (see another example in Méango and Poinas, 2023). *Ex ante returns*, the agent's perceived returns before they take an investment decision, are seen as the fundamental drivers of individuals' decisions (Heckman et al., 2006). They can be construed as the Willingness-to-Pay (WTP) to take the investment. Most contributions that attempt to understand (the distribution of) ex ante returns use a revealed preference approach (Carneiro et al., 2003; Cunha et al., 2005; Heckman et al.,

[1]The definition of returns adopted in this paper differs from the use of the term in the literature surveyed by Cunha and Heckman (2007) that refers to returns as to ex ante *earnings* returns only. See an extensive discussion in Appendix A.

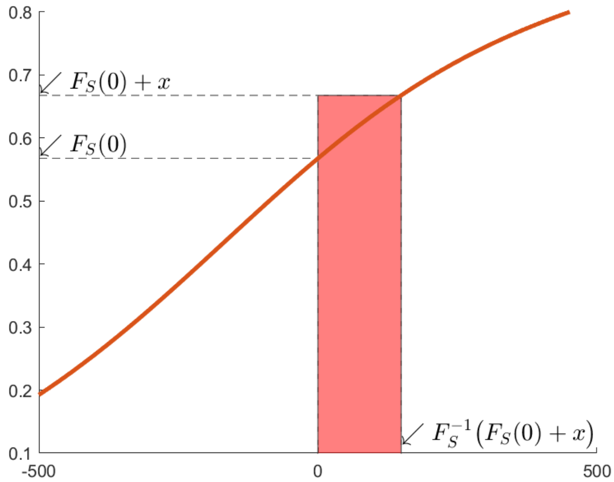


FIGURE 1.1. Illustrative example of the transfers to attract additional workers in the private sectors.

Note: $F_S(0)$ represents the prortion of individuals who would have negative ex ante returns. The marginal individual, i.e. who is indifferent between the private and public-sector job, is at the $F_S(0) - th$ percentile. To attract x additional job-seekers, the private sector must compensate up to the job-seeker at $F_S(0) + x - th$ percentile by transferring $F_S^{-1}\left[F_S(0) + x\right]$ to this individual. The cost of an expansion of the private sector is the shaded area: $F_S^{-1}\left[F_S(0) + x\right] \times (F_S(0) + x)$.

2006; Cunha and Heckman, 2007; Heckman and Navarro, 2007; Stange, 2012; Trachter, 2015; Lee et al., 2015; Eisenhauer et al., 2015; Bhuller et al., 2022).[2]

This paper specialises to the case where the analyst uses a stated choice experiment (also known as stated preference analysis) and asks a sample of the population of interest what they would do or choose in hypothetical situations. Stated choice experiments are increasingly used to describe individual preferences over choice attributes that would be otherwise endogenous, hard to measure in observational data, or hard to vary in an randomised control trial.[3] The public-private-job choice example is a case where job offers would be endogenous, difficult to randomise, but amenable to a stated choice experiment as is shown in the empirical application (see similar exercises on occupational choices in Mas and Pallais, 2017; Maestas et al., 2023; Wiswall and Zafar, 2015, 2018).

A key difficulty of an ex ante evaluation using stated choices resides in the time lag between elicitation and implementation. The decision about which offer to accept happens *tomorrow* but information about the ex ante perception of S must be collected in the

[2]Appendix A discusses in appropriate length the differences between the approach taken in this paper and the revealed-preference approach of the seminal contributions in this literature.

[3]Recent applications of the stated preference approach span many areas of economics, including education choices (Arcidiacono et al., 2020; Delavande and Zafar, 2019), mobility decisions (Gong et al., 2022; Koşar et al., 2022), health and long-term care investments (Kesternich et al., 2013; Ameriks et al., 2020a; Boyer et al., 2020), parental investments (Almås et al., 2023), marriage preference (Adams-Prassl and Andrew, 2019), retirement decisions (Ameriks et al., 2020b; Giustinelli and Shapiro, 2023) and irregular migration (Méango and Poinas, 2023). For recent reviews, see Koşar and O'Dea (2023) and Giustinelli (2023).

population *today*. This distinction is important because it is usually impossible to define all relevant choice attributes in a choice scenario. This leaves room for uncertainty about the unspecified job-specific amenities. For example, in the empirical application below that presents two job offers, one from the public, one from the private sector, the work-life balance, pension benefits, or commuting distance are left unspecified. Some of this uncertainty will be resolved at the time when the respondent faces the actual job offers (tomorrow), but some of it will only be resolved after the individual takes a decision, that is the *day after tomorrow*. Given this environment of sequential resolution, individuals asked today will be uncertain about their perception of S tomorrow; it will depend on the amount of uncertainty that they expect to be resolved between the time of elicitation (today) and the time of decision (tomorrow), the *resolvable uncertainty* (as coined by Blass et al., 2010).

The above discussion highlights the need for an approach to stated-preference analyses that accounts for agents' uncertainty about their tomorrow-returns. More specifically, the uncertainty about possible shocks in perception between the time of elicitation and the time of decision implies that each agent i's perception of returns is best described by a probability distribution over different levels of returns, say $F_{S,i}(s), s \in \mathbb{R}$. If direct elicitation of $F_{S,i}$ were feasible, the analyst would face little difficulty: provided individuals have a correct perception of the resolvable uncertainty, the population distribution of S that will prevail tomorrow can be calculated by aggregating the individual-specific distributions such that $F_S(s) = \sum_i \omega_i F_{S,i}(s)$, where ω_i are individual specific weights. Unfortunately, direct elicitation of individual returns or equivalently WTP can be challenging in many cases. Furthermore, the task at hand is to elicit an entire distribution rather than a single point estimate, which might be cognitively too demanding and prone to measurement error.

Instead of eliciting directly WTP parameters, traditional stated preference analyses routinely retrieve them from individuals intended choices, assuming that S is known to the agent (see a recent example in Maestas et al., 2023). Recent contributions that acknowledge the existence of a resolvable uncertainty only propose methodologies to recover the mean returns for each agent, that is $\mu_i := \int s \, dF_{S,i}(s)$, but carry no other information on $F_{S,i}$ (see, for example, Wiswall and Zafar, 2018; Koşar et al., 2022; Aucejo et al., 2023). However, in the running example, the distribution of means is irrelevant to the analyst, since tomorrow, when the resolvable uncertainty is realised, agent i will act on S_i, not μ_i.

The paper proposes a methodology to retrieve key information about $F_{S,i}$ by using stated choice experiments that elicit the *probability* of choosing a specific option. This type of data is now very common, and the proposed methodology fits well standard elicitation practice. In the empirical application, each respondent i is asked the percent chance to choose to the public/private sector offer in a given scenario t, say P_{it}. The scenario specifies, for example, wages, probability of being laid off, or typical working hours. Juster (1966), Manski (1999) and Blass et al. (2010) already noted that the use of *probabilistic* stated choices, that is, on a scale from 0 to 100 rather than a binary answer, allows respondents to express uncertainty about their future choice. However, they do not exploit this additional information to understand further individuals' uncertainty and focus only on mean preferences expressed as mean WTP. The paper proposes results to study the whole distribution of WTP.

Section 2 formalises the econometric framework. Within it, Theorem 1 shows that information about elicited choice probabilities P_{it} can be translated into information about the distribution of quantiles of $F_{S,i}$ in the population. The key step (Lemma 1) is to relate the *stated demand function*, the probabilistic choice at exogenous values of choice attributes, to quantiles of $F_{S,i}$. An estimator for F_S can then be derived from aggregating the distributions of quantiles of $F_{S,i}$. Theorem 2 in Appendix B shows that the stated demand function can also be used to identify quantile and mean effects of choice attributes (such as the chance of being laid off) on the distribution of $F_{S,i}$. Together, these two results show how to meaningfully extend traditional calculations of WTP parameters from stated choices to case where agents face a degree of (resolvable) uncertainty. Theorem 2 also characterises the distribution of resolvable uncertainty in the population. To the best of our knowledge, these are the first nonparametric characterisations of these objects in the stated-preference literature.

Section 3 is concerned with the nonparametric identification and estimation of the objects of interest. It makes two advances compared to the existing literature: the first advance is to show that given the exogeneity of choice attributes, the objects of interest are identified using the *cross-section*, that is with $T = 1$, if there is no measurement error in the stated choice, P_{it}. Repeated elicitation mainly serves the purpose of controlling for measurement error (or increasing power). In order to identify and estimate a nonparametric *population distribution* of mean WTP parameters, the existing literature has relied on *repeated elicitation* of the stated preference. For example, Wiswall and Zafar (2018); Koşar et al. (2022); Aucejo et al. (2023) leverage on the pseudo-panel structure and estimate separate demand functions for each individual. The population distribution of WTP is then obtained by aggregating the individual estimates. This strategy requires that the number of elicitation points exceeds the number of choice attributes considered.[4] In this paper, Theorem 1 and 2 characterise directly the population distribution of ex ante returns and WTP parameters from the nonparametric stated demand function. This is instrumental for our estimation strategy: instead of estimating and aggregating separate demand functions for each individual, the goal is to identify the quantiles of the nonparametric stated demand function. Given that choice attributes are assigned exogenously, identifying those quantiles does not require panel data, unless there is measurement error in the stated preference. Even when one considers measurement error, the number of elicitation

[4]In a recent paper, Kettlewell et al. (2024) discusses parametric models of preference heterogeneity to estimate a population distribution for WTP parameters. The approach in this paper achieves identification without their restrictive framework.

points for each survey participant can remain moderately low, irrespective of the number of attributes considered.

The second advance is to propose a novel nonparametric/semiparametric estimation strategy for the stated-preference literature. This literature almost exclusively on parametric assumptions to recover estimates of the indirect utility and WTP, namely that the stated demand function is linear (after a transformation) and the resolvable uncertainty is extreme value type I. The proposed strategy takes advantage of the characterisation results and the exogeneity of the choice attributes to relax these assumptions. Congruent with the nonparametric analysis of Section 2, (i) the econometric specification of the stated demand function can remain very flexible, and (ii) the distribution of the resolvable uncertainty is not parameterised. Section 3.6 compares the results from the proposed procedure to the results obtained when using the traditional parametric procedure on simulated data. The proposed procedure reduces significantly the bias associated to the competing estimation strategy at small T.

Section 4 discusses the empirical application. Murphy et al. (1991) argued that the choice of talented individuals to sort into rent-seeking occupations instead of productive (entrepreneurial) activities hurts economic growth. Against this background, a burgeoning literature cautions against the misallocation effects that a generous public sector generates in different contexts (see, for example, Algan et al., 2002; Albrecht et al., 2019; Burdett, 2012; Cavalcanti and Santos, 2020; Duflo et al., 2021; Girsberger and Meango, 2022; Mangal, 2022). Section 4 revisits this question by analysing the preference for public sector jobs of a sample of high-ability students from two highly selective universities in Côte d'Ivoire. The survey reveals differences in beliefs about the characteristics of the public and private sector. On the one hand, the private sector is perceived as a dynamic sector with more abundant, better paid jobs, and more opportunity for progression. However, these benefits are tied with stressful work conditions. On the other hand, the public sector appears as a secure, stable alternative, even if less well paid. It is important to note that beyond these general trends, there is substantial heterogeneity in individual beliefs. Moreover, even for identical observable job attributes, preferences for specific sectors vary significantly across students. This translates into heterogeneous and sometimes large (in absolute value) ex ante returns. The top decile of median returns (50-th percentile of $F_{S,i}$) is as large as 60 percent of the average perceived wage in the public sector, whereas the bottom decile represents a loss of about 69 percent of the average perceived wage in the public sector.

Section 4.5 estimates the cost elasticity associated with a private-sector expansion for the market of top skilled job-seekers. The proposed estimator of the ex ante returns implies that, with their average offers, each sector attracts about half of the job-seekers. Inducing one percent more job-seekers to choose the private sector offer would increase the private-sector wage bill by 2.3 percent, an economically large cost. The finding is consistent with Christiaensen and Premand (2017) who identify high labour cost due to a high skill premium as one possible constraint for firms to grow in Côte d'Ivoire.

2. A Characterisation of the Population Distribution of Ex Ante Returns

This section provides characterisation results for the population distribution of ex ante returns and WTP parameters that, unlike the existing literature, do not rely on parametric assumptions, neither on the individual's utility, nor on the resolvable uncertainty. Section 2.1 describes the econometric framework. Section 2.2 follows with a motivating example. Section 2.3 details the key assumptions for the characterisation result. Section 2.4 delivers a quick intuition, before Section 2.5 provides the main result: a characterisation of the population distribution of ex ante returns as a functional of a stated demand function. Then, Section 2.6 shows how this result can be used to perform a counterfactual analysis. Finally, Section 2.7 compares the framework with the existing literature.

2.1. Econometric framework.

Consider an economic agent i, a binary choice alternative 0 or 1, and two consecutive periods: a time of preference elicitation (today) and a time of decision (tomorrow). At the time of decision, i chooses between option 0 and option 1 based on a threshold-crossing rule:

$$D_i(x) \quad = \quad I\Big\{S\left(x, \eta_i^*\right) \geq 0\Big\} \tag{1}$$

where $I\{A\}$ takes value 1 if A and 0 otherwise. The vector (x, η^*) represents the choice attributes and individual preferences discussed in more details below. $D_i(x)$ is the choice of individual i. The notation borrows from the potential outcome framework, as $D_i(x)$ represents i's choice, when the first set of choice attributes are exogenously set to x.

The vector $x \in \mathcal{X}$ represents choice characteristics that can be manipulated within a hypothetical choice experiment. As is customary in stated preference analyses that translate preferences into pecuniary values, x also contains the vector of expected income (y_0, y_1) in option 0 and option 1 respectively. Thus, $x := (y_0, y_1, z)$, where z summarises the remaining manipulable choice characteristics.

The variable $\eta_i^* \in \mathcal{H}$ subsumes possibly unobserved characteristics of individual i and unobserved choice attributes that matter in the decision but are neither specified nor altered by the choice experiment. This could include a private taste for one of the two options, or some private information about the returns to choosing one option over the other. Finally, $S\left(x, \eta_i^*\right)$ is i's returns to choosing option 1 over option 0 at the time of decision, when the choice attribute are described by (x, η_i^*). The practical characterisation of S is as the pecuniary transfer that would make i indifferent between option 0 and 1 at the time of decision, that is, for all $x = (y_0, y_1, z) \in \mathcal{X}$ and $\eta_i^* \in \mathcal{H}$:

$$S(y_0 + S(x, \eta_i^*), y_1, z, \eta_i^*) = 0 \tag{2}$$

Prior to their decision and at the time of elicitation, i is asked to state their preference over the binary choice alternatives, 0 or 1, in hypothetical scenarios indexed by $t \in \mathbb{N}$. The scenarios are characterised by the different values of observable, manipulable characteristics $\{X_{it} := (Y_{0,it}, Y_{1,it}, Z_{it})\}_{t\in\mathbb{N}}$, where $Y_{d,it}$ is the expected income for option d in scenario t, and Z_{it} subsumes the remaining manipulable characteristics.

The model considers an environment of sequential resolution of uncertainty in which the value of η_i^* is only revealed at the time of decision. Because of the time gap between the time of elicitation and the time of decision, and given that η_i^* is unspecified by the scenario, i does not know the realisation of η_i^*. The model assumes that i entertains a probabilistic distribution over their returns based on their available information, characterised by a vector η_i. Define:

$$F_{S,i}(s; x) := \Pr(S(x, \eta_i^*) \leq s|\eta_i). \tag{3}$$

The perceived distribution of returns $F_{S,i}$ will be the main object of analysis. In traditional stated preference analyses, S is assumed to be known and the distribution is thus degenerated. In this case, $S(\cdot)$ is recovered as a WTP parameter from the stated choices, under the assumption that it is known by the agent. However, when uncertainty is resolved sequentially, agent i does not know $S(\cdot)$. Instead, they entertain a distribution $F_{S,i}$. As the agent reveals their probabilistic choice, they also reveal information about $F_{S,i}(s; x)$.

More formally, during a survey experiment, i is presented with scenario, $\{X_{it}\}_{t\in\mathbb{N}}$. They are asked to state their chance of choosing option 1 over option 0, say P_{it}. Ideally, presented with scenario $X_{it} = x$, respondent i should state:

$$m(x, \eta_i) := \Pr(S(x, \eta_i^*) \geq 0|\eta_i) = 1 - F_{S,i}(0; x) \tag{4}$$

The mapping $x \mapsto m(x, \eta)$ defines the *stated demand function* for an individual with characteristic η. In line with the literature, the stated choice experiment is construed as a *ceteris paribus* experiment. Within it, the respondent is asked to report their stated choice as if X_{it} was determined exogenously.

Instead of the ideal report, it is customary to consider that individuals make mistakes, possibly due to inattention, misunderstanding the survey instrument, and/or lack of effort. The measurement error in elicited choices is such that $P_{it} \neq m(X_{it}, \eta_i)$.

2.2. **Motivating example: a job-choice model.** Consider a job-choice model where individuals have preferences over jobs characterised by a bundle (y, a) of income y and sector-specific amenity a. The individual utility is described by a CES utility function:

$$U_i(y, a) = \left(\alpha_i y^{\beta_i} + (1 - \alpha_i) a^{\beta_i}\right)^{\frac{1}{\beta_i}}.$$

Let option 0 be a job in the public sector, and option 1 a job in the private sector. When deciding between a job offer in the public, (y_0, a_0), and the private sector, (y_1, a_1), the individual chooses the one that maximises their utility. During a stated choice experiment, job-seeker i is presented with T pairs of wages associated with a job in the public sector (y_{0it}) and a job in the private sector (y_{1it}). They are asked to state their probability of choosing option 1 over option 0. Importantly, the stated choice experiment does *not* specify the associated amenities. Thus, the respondent does not know the vector of amenities (a_0, a_1), but holds beliefs about its distribution, $F_{a_0,a_1}(.;\rho_i)$, that depends on a parameter ρ_i. The job-seekers knows $\eta_i := (\alpha_i, \beta_i, \rho_i)$, but these parameters are not observable for the analyst. Thus, $\eta_i^* = (\eta_i, a_{i0}, a_{i1})$, and the vector of amenities represents the resolvable uncertainty. In this simple example, it is possible to derive analytically the ex ante returns:

$$S(y_0, y_1, \eta_i^*) = \Big[y_1^{\beta_i} + \frac{(1-\alpha_i)}{\alpha_i}(a_{1i}^{\beta_i} - a_{0i}^{\beta_i})\Big]^{\frac{1}{\beta_i}} - y_0.$$

The stated preference P_{it} are reported with an error ϵ_{it}, such that $P_{it} = h(Y_{0,it}, Y_{1,it}, \eta_i, \epsilon_{it})$. The next sections present assumptions on h and on the elicitation procedure that permit to characterise the entire distribution $F_{S,i}$ from the stated demand function $m(y_0, y_1, \eta)$ as defined in equation (4).

2.3. **Assumptions on elicited choices.** This section details the type of data observed by the researcher, as well as the main assumptions maintained in the paper. First, introduce few notations: for any variables X, Y, we denote by $F_X(x)$ and $F_Y(y)$ the marginal cumulative distribution functions, $F_{Y|X}(y|x)$, the conditional distribution function of Y given X, and $Q_{Y|X}(y|x)$ the conditional quantile function of Y given X, which is the generalised inverse of $F_{Y|X}$. For any variable X, denote by X' an i.i.d. copy of X.

Assumption 1 (Data)**.** *Let (X, η, ϵ) be a data generating process. The sample consists of copies $(X_{it}, \eta_i, \epsilon_{it})_{i\in\mathbb{N}, t\in\mathbb{N}}$, independent across individuals i. Moreover $X_{it} \perp\!\!\!\perp (\eta, \{X_{it'}\}_{t'\neq t})$.*

The analyst observes $(P_{it}, X_{it})_{i\in\mathbb{N}, t\in\mathbb{N}}$ where there is a real function h such that $P_{it} = h(X_{it}, \eta_i, \epsilon_{it})$ and $Q_{P|X,\eta}(0.5|X_{it}, \eta_i) = m(X_{it}, \eta_i)$ for all i, t.

The analyst's interest is in $m(X_{it}, \eta_i)$, but except at the median of ϵ_{it}, $h(X_{it}, \eta_i, \epsilon_{it}) \neq m(X_{it}, \eta_i)$.

Assume that h is strictly increasing in the third argument ϵ. Hence, without loss of generality, normalise $\epsilon_{it}|X_{it}, \eta_i \sim \mathcal{U}[0,1]$. Thus $h(X_{it}, \eta_i, \epsilon_{it}) = Q_{P|X,\eta}(\epsilon_{it}|X_{it}, \eta_i)$, and $h(X_{it}, \eta_i, 0.5) = m(X_{it}, \eta_i)$.

Assumption 1 describes the pseudo-panel data obtained from a stated choice experiment. It allows for measurement errors that are deviations around the stated demand function. The restriction on the distribution of ϵ is a normalisation in the sense that one can define a random variable ε, such that $P = g(X, \eta, \varepsilon)$, with ε possibly associated with (X, η), and $\epsilon = F_{\varepsilon|X,\eta}(\varepsilon|X, \eta)$. The normalisation is standard in a nonparametric identification framework.

Individuals characterised by an unobserved heterogeneity η and presented with the same scenario x_t will sometimes give answers above or below the stated demand function $m(x_t, \eta)$, however, the median answer will correctly identify the stated demand function. An alternative assumption is to consider that the reports are correct on average. The restriction on the median is the one entertained in the seminal paper of Blass et al. (2010) and the subsequent probabilistic stated choice literature.

Assumption 2 (Monotonicity)**.** *The map $y_0 \mapsto S(y_0, y_1, z, \eta^*)$ is strictly decreasing for any (y_1, z, η^*).*

Assumption 2 states that the returns on choosing option 1 over option 0 decrease with the income in option 0. This is easily satisfied for utility functions that are strictly increasing with income.

2.4. **A quick intuition of the characterisation.** This section delivers a quick intuition of the characterisation result. The analyst is interested in the distribution: $\Pr(S(x, \eta^*) \leq s|\eta)$, the distribution of ex ante returns. The respondent provides information about: $\Pr(S(X_{it}, \eta_i^*) \geq 0|\eta_i) := m(X_{it}, \eta_i)$, the *stated demand*. Because choices attributes X_{it} are varying exogenously, the analyst can use the variations in X_{it} to learn about the distribution of interest. Useful for this is the restriction on S contained in equation (2), $S(y_0 + S(x, \eta_i^*), y_1, z, \eta_i^*) = 0$. This restriction follows from the definition of S as the pecuniary transfer that makes individuals indifferent between the two investment options. It further implies that $\Big\{\eta^* : S(y_0, y_1, z, \eta^*) \leq s\Big\} = \Big\{\eta^* : S(y_0 + s, y_1, z, \eta^*) \leq 0\Big\}$ (see Lemma 1). In other words, those in the population whose returns is lower than some value s are those whose returns would still be lower than 0 even if one would top up their income from y_0 to $y_0 + s$. Thus:

$$\Pr(S(y_0, y_1, z, \eta_i^*) \leq s|\eta_i) = \Pr(S(y_0 + s, y_1, z, \eta_i^*) \leq 0|\eta_i) = 1 - m(y_0 + s, y_1, z, \eta).$$

The LHS is the quantity of interest. The RHS uses the stated demand function m. Theorems 1 and 2 show how to use the stated demand function to obtain the objects of interest. Those two theorems are our *characterisation results*.

2.5. **From the stated demand function to the distribution of returns.** Theorem 1 below shows how to recover the full distribution of $F_{S,i}$ from the stated demand function m. The stated demand fucntion will in turn be derived from the elicited choices P_{it}. Theorem 1 builds on the following Lemma.

Lemma 1. *In the model described by equations (1)-(3) and under Assumption 2, the following holds:*

$$F_{S,i}(s; x) = 1 - m(y_0 + s, y_1, z, \eta_i), \text{ for any } s \in \mathbb{R}. \tag{5}$$

Lemma 1 articulates the link between the private distribution of returns (LHS), our object of interest, and the stated demand function (RHS). It states that, at any given η_i, the chance of having returns below some value s is the same as the chance that the returns are lower than 0 even if the income will change from y_0 to $y_0 + s$.

Unfortunately, the vector η_i is not observed by the analyst. So the individual distribution of returns $F_{S,i}$ is not available by using Equation (5). Nevertheless, by averaging η_i out, one can learn about the distribution of quantiles of $F_{S,i}$ in the population. Ensuring that this operation is well defined requires the following assumption.

Assumption 3 (Support Condition)**.** *Let $\mathcal{X}$ be the support of x, $\mathcal{H}$, the support of η, and $\mathcal{H}|x$ the support of η conditional on $X = x$. For all $x \in \mathcal{X}$, $\mathcal{H}|x = \mathcal{H}$.*

Assumption 3 requires that the support of the unobserved heterogeneity does not change with X. This is akin to the large support condition of Imbens and Newey (2009). It will be satisfied by design since the scenarios are generated independently from the unobserved heterogeneity.

Theorem 1 is one of the main results. It shows how information about the structural function m can be harnessed to learn about the distribution of quantiles of $F_{S,i}(s;x)$. Note that the vector x contains choice attributes that the policy maker can manipulate (e.g. the chance of being laid off, the chance of being promoted). Theorem 1 shows that one can learn about the distribution of quantile $F_{S,i}$, for any given distribution of a random variable of choice attribute, say $F_{\tilde{X}}$.

Theorem 1. *Let $F_{\tilde{X}}$ be the cumulative distribution function of the variable $\tilde{X}$, which is of interest for the policy maker. Define $Q_{S,i}(\tau;\tilde{X}) := \inf\{s : F_{S,i}(s;\tilde{X}) \geq \tau\}$ and $F_Q(s;\tau,F_{\tilde{X}}) := \Pr\left(Q_{S,i}(\tau;\tilde{X}) \leq s\right)$. Under the conditions of Lemma 1 and Assumption 3, the following holds: For any real value s such that $(y_0+s, y_1, z) \in \mathcal{X}$ and $\tau \in [0,1]$,*

$$\begin{aligned} F_Q(s;\tau,F_{\tilde{X}}) &= \Pr\left[F_{S,i}(s;\tilde{X}) \geq \tau\right], \\ &= \int_{\mathcal{X}}\int_{\mathcal{H}} I\left\{m(y_0+s,y_1,z,n) \leq 1-\tau\right\} dF_{\eta|X}(n|y_0,y_1,z)dF_{\tilde{X}}(y_0,y_1,z) \quad (6) \end{aligned}$$

The proof is immediate by Lemma 1. Theorem 1 recovers the distribution of the cdf $F_{S,i}(s;\tilde{X})$ in the population. Hence, it describes individuals' perceived returns, accounting for their perceived uncertainty. Alternatively, it can be represented as a set of distributions of quantiles $F_Q(s;\tau,F_{\tilde{X}}), \tau \in (0,1)$. This has an intuitive interpretation: $F_Q(s;\tau,F_{\tilde{X}})$ is the proportion of the population for whom the τ-quantile of returns is lower than some given value s.

2.6. **From the distribution of returns to the counterfactual distribution of realised returns.** The distributions of quantiles allow to construct counterfactual distributions of realised returns tomorrow, the main object of interest for ex ante policy evaluation. Indeed, if shocks are i.i.d. and the policy maker considers that the distribution $F_{S,i}$ is correct in the sense that when uncertainty is resolved, $S(x,\eta_i^*)$ is a draw for the distribution $F_{S,i}$, the best predictor for the realised distribution of returns is a mixture of the distributions of quantiles with equal weights; that is:

$$F_S(s,F_{\tilde{X}}) := \int_0^1 F_Q(s;\tau,F_{\tilde{X}})d\tau. \quad (7)$$

The policy maker can explore other assumptions about the correctness of respondents' perception, which result in different weighting schemes for $F_Q(s;\tau,F_{\tilde{X}})$.

The stated demand function m serves to characterise further objects of interests for the policy maker. For example, it might be of interest to understand the effect of a given choice attribute x_k on the perceived returns. Theorem 2 in Appendix B presents a characterisation of the distribution of quantile effects and mean effects of choice attributes on ex ante returns in the population. It also characterises the distribution of dispersion of $F_{S,i}$ using the inter-quantile range (IQR). The latter describes the amount of uncertainty that the agents expect to be resolved at the time of decision.

2.7. **Comparison with the existing literature.** The model and restrictions from equations (1)-(3) and Assumptions 1 and 2 on the data generating process, the decision mechanism, and the resolvable uncertainty are significantly milder than the restrictions commonly imposed in the literature. For example, a common representation following Blass et al. (2010) and Wiswall and Zafar (2018) is to impose: $S(x, \eta_i^*) = (y_1 - y_0) + \gamma_i + (z_1 - z_0)'\delta_i + \nu_i$ where ν_i is the (additively separable) resolvable uncertainty, and follows a logistic distribution with variance σ_i. If the parametric model would map one-to-one with the true model, one should have $\eta_i = (\gamma_i, \delta_i, \sigma_i)$ and $\eta_i^* = (\eta_i, \nu_i)$. In addition, they assume that the log-odds $\log(P_{it}/(1 - P_{it}))$ are reported with an additively separable measurement error, ϵ_{it}, which has a zero median, that is $Q_{\epsilon|X,\eta}(0.5|X_{it}, \eta_{it}) = 0$.

The linearity restriction of the above representation facilitates identification and estimation, because the WTP parameter is identified by using the ratio between two coefficients of a linear regression (see, for example, Wiswall and Zafar, 2018, equation (8), p.488).[5] The parametric restriction on the resolvable uncertainty is made for technical convenience and has no compelling motivation, as acknowledged by Blass et al. (2010).

The framework in this paper proposes several important relaxations: it permits nonlinearity and nonseparability of the returns with respect to choice attributes, so that choice attributes can freely interact between them and with other unobservable characteristics. For example, the returns to option 1 depend not only on the net pecuniary returns $(y_1 - y_0)$ as for risk neutral individuals, but also on the income level in option 0, y_0. The resolvable uncertainty is not restricted to additive separability from the ex ante returns. Hence, beliefs about an event can depend on the (dis)utility of the event (contained in η_i), as in models of motivated beliefs (see Bénabou and Tirole, 2016). Furthermore, the

[5]The theoretical motivation for the linear form is an additive random utility model (ARUM). One can see a parallel between the ARUM model used for stated choice experiments and the subjective expected utility model (SEU) of structural models of decision using subjective expectations (see, for example, Manski, 2004). For instance, some of the choice attribute x pertain to uncertain outcomes as in the case of the probability of being laid off in the future. These uncertain outcomes enter the utility as an additive term, which consists of the product between the probability of the event and the (dis)utility associated to the event. Giustinelli (2022) notes two limitations of the linear-SEU model: (1)"An implication of the linear SEU specification is that decision makers are assumed to be risk neutral with respect to continuous outcomes, notably expected earnings", and (2) "Although a standard feature of canonical SEU [...], multiplicative separability rules out the possibility that a person's subjective probability of an event depends on his/her (dis)utility for the event, as in models of utility-based or motivated beliefs." By remaining silent on the underlying utility model and allowing for unrestricted decision mechanisms, the present framework avoids these limitations.

resolvable uncertainty is not parameterised, allowing for unrestricted forms of individuals beliefs about the uncertainty that will be resolved at the time of decision. Finally, the measurement error is nonseparable so that one does not need to resort to restrictions on the log-odd transformation of the stated choices.

The model retains one common assumption about stated preference experiments: they are *ceteris paribus* experiments. The respondents is invited to assume that only X_{it} changes across experiments and that these changes are exogenous. Therefore, η_i remains stable across scenarios. The above treatment of the unobserved heterogeneity differs from the one in Wiswall and Zafar (2018). They entertain the possibility that individuals presented with various manipulated choice attributes infer different unobserved choice attributes. Thus, η_i (or more precisely, the part of η_i related to unobserved choice attributes) is not stable across experiments. To achieve identification, they need to assume that, within each scenario, the unobserved choice attributes are exactly the same across the choice options. To ensure this, they instruct the respondents to consider that, within a scenario, the unobserved choice attributes are exactly the same across the choice options. Crucially, they also need to assume separable returns/utilities. Under these two assumptions the effects of these unobserved attributes 'cancel out'.

From the theory side, nothing precludes adopting the same strategy: one could augment equations (1)-(3) to add separable unobserved choice characteristics that change across experiments, but are identical within pairs of choices. However, whereas Wiswall and Zafar (2018) consider 'anonymous jobs', the empirical application in the paper considers public-sector jobs against private-sector jobs that respondents perceive to be very different. It is not plausible to consider that unobserved choice attributes can be perceived as exactly the same across sectors. To ensure that respondents perceived choice attributes as stable, the choice experiment explicitly instruct, that across scenarios, the only choice attributes that change are the one defined in the experiment. The remaining ones are exactly the same. Hudomiet et al. (2018, 2021) provide evidence that the concern of changing unobserved heterogeneity is minor in some contexts and can be mitigated by carefully designed elicitation procedures.

3. Identification, estimation and inference

From Theorem 1, the stated demand function $m(x, \eta)$ stands out as the main object to estimate. Assumption 1 is the basis for this estimation from hypothetical choice experiments $\{P_{it}, X_{it}\}_{i,t}$.

An approach consistent with the stated preference literature is to leverage on the pseudo-panel structure and estimate separate demand functions for each individual (Wiswall and Zafar, 2018). One challenge with stated preference data is that (pseudo-)panels are relatively short. For example, Blass et al. (2010) elicit up to 10 stated choices, Wiswall and Zafar (2018) up to 16, Koşar et al. (2022) up to 22, and Aucejo et al. (2023) up to 42 per individual. The empirical application in this paper has $T = 5$. A solution adopted in the literature is to impose a parametric form of the stated demand function and of the resolvable uncertainty that leaves enough degrees of freedom to estimate individual-specific parameters. One common example sets:

$$\log\left(\frac{P_{it}}{1-P_{it}}\right) = X_{it}'\eta_i + \epsilon_{it}, \text{ with } Q_{\epsilon_{it}|X_{it},\eta_i}(0.5|X_{it},\eta_i) = 0, \tag{8}$$

and estimate $\hat{\eta}_i$ by Least Absolute Deviation, $\hat{m}(X_{it},\hat{\eta}_i) = \exp(X_{it}'\hat{\eta}_i)/(1+\exp(X_{it}'\hat{\eta}_i))$, for each individual in the sample. The population distribution is obtained by aggregating over each individual. It is important to note that (i) this estimation procedure does not exploit the cross-sectional variation, except at the aggregation stage, and (ii) the panel dimension must always exceed the number of attributes considered.

To address the case of limited panel length and remain consistent with the nonparametric characterisation of Section 2, this section develops a novel nonparametric/semiparametric estimation procedure. Identification results for several panel models with nonseparable unobserved heterogeneity and fixed T have emerged in the recent literature, e.g. Evdokimov (2010); Freyberger (2018); Sasaki (2015), and could be adapted in the present context. However, they would not fully take advantage of the context of stated choice experiments. The approach taken here is to show a constructive identification result for T unrestricted, and demonstrate in simulations that estimation following this constructive result performs much better with small-T samples than the traditional procedure used in the stated preference literature.

Section 3.1 explains that given the exogeneity of choice attributes, the objects of interest could be identified using the *cross-section*, that is with $T = 1$, if there would be no measurement error. Repeated observations only serve the purpose of controlling for measurement error in the stated preferences. Given this result, it seems counter-intuitive not to use the cross-section of variation.

Section 3.2 shows that both the cross-sectional and the panel variations can be used to produce an estimate for measurement error. Once controlled for, estimation can proceed as in the case of no measurement error. By taking advantage of the cross-sectional variation, this procedure does not require that the number of scenarios exceeds the number of attributes to exploit the panel variation.

Section 3.3 and 3.4 details the estimation and inference procedure following the constructive identification argument. Sections 3.5 briefly discusses asymptotic theory by drawing a parallel to previous work from Chernozhukov et al. (2020). Section 3.6 confirms that the proposed estimation methodology performs much better with small-T samples ($T \leq 20$) than the traditional procedure.

The nonparametric estimation strategy developed here for a nonseparable panel model can be of independent interest. We think that a more systematic study of its properties is beyond the scope of the present paper, but should be a fruitful avenue for future research.

Before we proceed, note that because the unobserved heterogeneity, η, is ultimately averaged out in Theorem 1, it is appropriate to work with the quantile treatment response

(QTR) function, as defined by Chernozhukov and Hansen (2005), rather than with the structural function $m(.)$. For any $x \in \mathcal{X}$, define $q(x, \alpha) := Q_{m(x,\eta)}(\alpha)$, the α-quantile of the random object $m(x, \eta)$.

3.1. **Identification in the absence of measurement error.** To convey the intuition of identification, suppose that at $t = 1$, there is no measurement error, that is $\epsilon_{i1} = 0.5$ and $P_{i1} = q(X_{i1}, \alpha_{i1})$ for all i. In this case, the QTR is identified as:

$$q(x, a) = Q_{P_1|X_1}(a|x) \text{ for all } x \text{ in the support of } X_1 \text{ and } a \in (0, 1). \tag{9}$$

Hence, in the absence of measurement error, the distributions of quantiles as characterised by Theorem 1 and the additional parameters of interest in Theorem 2 are all identified from eliciting probabilistic stated choices in *one* scenario. More specifically:

$$\begin{aligned} F_Q(s; \tau, F_{\tilde{X}}) &= \int_{\mathcal{X}} \int_0^1 I\left\{Q_{P_1|X_1}(a|y_0 + s, y_1, z) \leq 1 - \tau\right\} da\, dF_{\tilde{X}}(y_0, y_1, z) \\ &= \int_{\mathcal{X}} \int_0^1 I\left\{a \leq F_{P_1|X_1}(1 - \tau|y_0 + s, y_1, z)\right\} da\, dF_{\tilde{X}}(y_0, y_1, z) \\ &= \int_{\mathcal{X}} F_{P_1|X_1}(1 - \tau|y_0 + s, y_1, z) dF_{\tilde{X}}(y_0, y_1, z) \end{aligned} \tag{10}$$

The RHS is identified from the observed joint distribution of (P_1, X_1) and any given distribution $F_{\tilde{X}}$. Hence the distribution of quantiles (LHS) is identified. This results makes clear that repeated elicitation serves mainly the purpose of correcting for measurement error in the elicitation procedure (and possibly increasing the precision of estimation). A pseudo-panel is not required for identifying the *population distribution* of ex ante returns or WTP for choice attributes.

3.2. **Identification in the presence of measurement error.** This section deals with a non-classical measurement error (potentially correlated with the choices attributes and the unobserved heterogeneity). This results in a stated choice $P_{it} = h(X_{it}, \eta_i, \epsilon_{it}) \neq m(X_{it}, \eta_i)$. The restriction of Assumption 1 is that, conditional on (X_{it}, η_i), the median stated choice is unbiased. That is, under the normalisation of Assumption 1, $h(X_{it}, \eta_i, 0.5) = m(X_{it}, \eta_i)$.

Proposition 1. *Under Assumptions 1, $q(x, a) = Q_{P|X,\epsilon}(a|x, 0.5)$, for any $a \in (0, 1)$ and any $x \in \mathcal{X}$.*

Proposition 1 establishes the link between the QTR and the conditional quantile function of P. Comparing to the case without measurement error, the main difference is the presence of ϵ in the conditioning set. The next proposition shows that the data are rich enough to control for the effect of ϵ on this conditional quantile.

It is worth noticing that from Assumption 1, $h(X_{it}, \eta_i, \epsilon_{it}) = Q_{P|X,\eta}(\epsilon_{it}|X_{it}, \eta_i)$, hence $\epsilon_{it} = F_{P|X,\eta}(P_{it}|X_{it}, \eta_i)$. To rephrase this result, with T unrestricted, if the analyst could identify the conditional distribution of P_{it} given (X_{it}, η_i), they would identify ϵ_{it}, so that the QTR is identified. Now η_i is unobserved, yet, the conditional distribution is identified nonparametrically by considering separately the conditional distribution of $F_{P|X}$ for each individual, effectively controlling for η_i. In practice, nonparametric estimation following this result would require a long panel to be able to 'match' within individual on the random vector X_{it} and span its support for each individual. This would not improve on the procedure suggested by Wiswall and Zafar (2018) that estimates separate demand functions, would not use the fact that $X \perp\!\!\!\perp \eta$, and would ignore the cross-sectional variation.

Instead, the paper proposes a procedure that relies on the fact that $X \perp\!\!\!\perp \eta$ to exploit the cross-sectional variation through the following iterative procedure.

Step 1: Consider $V_0 = F_P(P)$. Construct $V_1 = F_{V_0|X}(V_0|X)$. This step uses the cross-sectional variation.
Step 2: Consider V_1 as defined in step 1. Construct $V_2 = F_{V_1|\eta}(V_1|\eta)$. Although η is unobserved, this is possible by considering separately each individual, and ranking the values V_1. This step uses the panel variation.
...
Step $2k+1$**:** Consider V_{2k}. Construct $V_{2k+1} = F_{V_{2k}|X}(V_{2k}|X)$.
Step $2(k+1)$**:** Consider V_{2k+1}. Construct $V_{2(k+1)} = F_{V_{2k+1}|\eta}(V_{2k+1}|\eta)$.

One way to understand the procedure is that, at each step, it 'partials out' the effect of either X or η. Indeed, $V_{2k} \perp\!\!\!\perp \eta$, for all $k \geq 1$, and $V_{2k+1} \perp\!\!\!\perp X$, for all $k \geq 1$ (see, for example, Matzkin, 2008). Step 1 partials out the effect of X on V_0 (or equivalently P). However, V_1 still depends on η (and ϵ). Step 2 partials out the effect of η on V_1. Unfortunately, because V_1 is *not* jointly independent from X and η, this may re-introduce a dependence with respect to X. One can then iterate step 1, and partial out the effect of X on V_2. However, because V_2 is *not* jointly independent from X and η, this may re-introduce a dependence with respect to η. One can then iterate Step 2. And so on, and so forth. The key result is that, because $X \perp\!\!\!\perp \eta$, this construction will ultimately lead to a fixed point where the effect of X and η are jointly purged, leaving the variation in ϵ only.

Proposition 2. *Assume that P is continuous and the mapping $e \mapsto h(x, n, e)$ is strictly monotone for every $(x, n) \in \mathcal{X} \times \mathcal{H}$. Under Assumption 1, the sequence $\{V_k\}_{k=1}^{\infty}$ converges to $V_\infty = \epsilon$.*

Proposition 2 suggests that an estimate of the measurement error can be obtained through an iterative procedure. It represents an alternative to estimating the conditional distribution $F_{P|X,\eta}$.

The above results are the rationale for the following estimation procedure:

(1) Let $V_0 = P$. Across the whole population, estimate $\hat{V}_{1,it} = \widehat{F}_{P|X}(P_{it}|X_{it})$, for example using kernel estimation or using a semi-parametric estimator.
(2) Consider each individual i separately and rank $\hat{V}_{1,it}$ across scenarios t. Let $\hat{V}_{2,it}$ be this rank.

(3) Iterate step 1 and 2, by replacing $V_{k,it}$ by $V_{k+2,it}$ until it satisfies a convergence criterion or after a pre-specified number of iterations. Call $\hat{V}_{\infty,it}$ the estimated fixed-point.

(4) Estimate $\widehat{q}(x,a) = \widehat{Q}_{P|X,\hat{V}_\infty}(a|x,0.5)$, for example using a kernel estimator or a semi-parametric estimator. Alternatively, to use the equivalent of equation (14), estimate $\hat{F}_{P|X,\hat{V}_\infty}(a|x,0.5)$.

The identification result relies on T being unrestricted and consistent estimation relies on large T. Still, we expect that in finite sample, this procedure outperforms the alternative procedure that estimates separate demand functions for each individual because it takes advantage of the cross-sectional variation. Section 3.6 shows that it is the case for our simulated data.

One important question is the number of steps required for the sequence $\{V_k\}_{k=1}^{\infty}$ to reach its fixed-point. A general result is difficult to obtain, however, we offer three comments. First, one can show that for the following class of function $\{h : \mathcal{X} \times \mathcal{H} \times [0,1] \to [0,1]$ such that there exists two real functions f, g, and $h(X,\eta,\epsilon) = f(X, g(\eta,\epsilon))\}$, a fixed-point is attained at V_2. This class of function reduces the unobserved heterogeneity to a scalar, and imposes some version of rank similarity (Chernozhukov and Hansen, 2005). Second, in our simulation, the fixed-point seems to be approached very rapidly. The correlation between $\hat{V}_7$ and $\hat{V}_8$ exceeds 0.98, and concomitantly, the correlation between $\hat{V}_k$ and X or η approaches zero after only few iterations. Third, the empirical application displays a similar speed of convergence to a fixed point, so that correlation between $\hat{V}_5$ and $\hat{V}_6$ exceeds 0.97. In light of this, the remaining discussion considers only the case of a pre-specified number of iterations, which is the simplest. Inference and asymptotic properties for the case of a convergence criterion is left for future research.

3.3. **Estimation for the distribution of quantiles.** The previous section suggests a multistage estimation procedure. This section provides a detailed description of each step based on a Distribution Regression (DR) estimation for the conditional distribution $F_{V_k|X}, k \geq 1$, and $F_{P|X,\epsilon}$. This semiparametric estimator provides a feasible alternative to the Kernel estimator when the dimension of X is large, as in our empirical application. The multistage procedure is closely related to the estimation and inference Chernozhukov et al. (2020) (henceforth, CFNSV), which notations are adopted here. Inference is conducted using a weighted bootstrap procedure.

First stage: Estimation of the conditional distribution $\hat{F}_{V_k|X}, k \geq 0$, $\hat{V}_k$, $k \geq 1$, *and* $\hat{\epsilon}$. The conditional distributions $F_{V_k|X}(p|x), k \geq 0$, $p \in (0,1)$ and $x \in \mathcal{X}$ are estimated by distribution regression.

Let $V_0 = P$. Consider the estimator for V_1. Let Λ be the logistic CDF. The DR estimator for each $t \leq T$ is:

$$\hat{F}^e_{V_0|X}(p|X_{it};t) = \Lambda\left(R'_{it}\hat{\pi}^e_t(p)\right), \text{ where } R_{it} = r(X_{it}),\ p \in (0,1], \text{ and}$$
$$\hat{\pi}^e_t(p) \in \arg\min_{\pi \in \mathbb{R}^{dim(R)}} -\textstyle\sum_i e_i \left[1\{V_{0,it} \leq p\}\log\left(\Lambda(R'_{it}\pi)\right) + 1\{V_{0,it} > p\}\log\left(1 - \Lambda(R'_{it}\pi)\right)\right] \tag{11}$$

Note that the DR regression is done separately at each t. When $e_i = 1$ for all $i = 1, \ldots, N$, equation (11) defines the estimator $\hat{F}_{V_0|\boldsymbol{X}}$. For $V_{0,it}, X_{it}$ in $[0,1] \times \mathcal{X}$, the estimator of the weighted bootstrap version of the conditional cdf are $\hat{F}^e_{V_0|\boldsymbol{X}}(p|X_{it};t) = \Lambda\left(R'_{it}\hat{\pi}^e_t(p)\right)$.

The estimator of $V_{1,it}$ is given by $\hat{V}_{1,it} = \hat{F}_{V_0|\boldsymbol{X}}(V_{0,it}|X_{it};t)$ and the weighted bootstrap version is $\hat{V}^e_{1,it} = \hat{F}^e_{V_0|\boldsymbol{X}}(V_{0,it}|X_{it};t)$.

Consider V_2. To estimate $\hat{V}_{2,it}$, for each individual i, rank $V_{1,it}$ using:

$$\hat{V}^e_{2,it} = \frac{1}{T}\sum_{t'=1}^{T} I\{\hat{V}^e_{1,it'} \leq \hat{V}^e_{1,it}\} \tag{12}$$

Iterate the construction by replacing $\hat{V}_{2(k-1)}$ by $\hat{V}_{2k}$, for $k \geq 1$. Repeat K times. Alternatively, one can repeat until the correlation between $\hat{V}_{2k-1}$ and $\hat{V}_{2k+1}$ exceeds a pre-specified threshold.

Second stage: Estimation of the conditional distribution $F_{P|X,\epsilon}$. The conditional distributions $F_{P|X,\epsilon}(p|x,v)$, $p \in (0,1)$, $x \in \mathcal{X}$, and $v \in (0,1)$ are estimated by distribution regression. The DR estimator is defined similarly as above:

$$\hat{F}^e_{P|X}(p|X_{it},\hat{\epsilon}_{it}) = \frac{1}{T}\textstyle\sum_{t'=1}^{T} \Lambda\left(W'_{it}\hat{\beta}^e_{t'}(p)\right), \text{ where } W_{it} = w(X_{it},\hat{\epsilon}_{it}),\ p \in (0,1], \text{ and}$$
$$\hat{\beta}^e_t(p) \in \arg\min_{\beta \in \mathbb{R}^{dim(W)}} -\textstyle\sum_i e_i \left[1\{P_{it} \leq p\}\log\left(\Lambda(W'_{it}\beta)\right) + 1\{P_{it} > p\}\log\left(1 - \Lambda(R'_{it}\beta)\right)\right] \tag{13}$$

Third stage: Estimation of the distribution of quantiles F_Q. Let $F_{\tilde{X}}$ be a distribution of interest for the policy maker. Given the estimator $\hat{F}_{P|X,\epsilon}$ and their bootstrap draws $\hat{F}^e_{P|X,\epsilon}$, we can form estimators of distribution of quantiles F_Q as functionals of these building blocks.

$$\hat{F}^e_Q(s;\tau,F_{\tilde{X}}) = \int_{\mathcal{X}} \hat{F}^e_{P|X,\epsilon}(1-\tau|y_0+s,y_1,z,0.5)dF_{\tilde{X}}(y_0,y_1,z), \text{ for any } \tau \in (0,1).$$

3.4. **Inference.** This section considers inference over regions of values $s \in \tilde{\mathcal{S}} \subset \mathbb{R}$. The weighted bootstrap versions of the distributions of interest are obtained by rerunning the estimation procedure in Section 3.3 with sampling weights that satisfy Assumption 3 in CFNSV (p.518). There are used to perform uniform inference. A $(1-\alpha)$-confidence region for $F_Q(s;\tau,F_{\tilde{X}})$ over the region $\tilde{\mathcal{S}}$ is given by

$$\left[\hat{F}_Q\left(s;\tau,F_{\tilde{X}}\right) \pm \hat{k}_\tau\left(1-\alpha\right)\hat{\sigma}_\tau, s \in \tilde{\mathcal{S}}\right], \tag{14}$$

where $\hat{\sigma}_\tau = IQR[\hat{F}^e_Q(s;\tau, F_{\tilde{X}})]/1.349$ and $\hat{k}_\tau(1-\alpha)$ denotes the $(1-\alpha)$-quantile of bootstrap draws for the maximal t-stat

$$\|t^e_\tau(s)\|_{\tilde{S}} = \sup_{s\in\tilde{S}} \left| \frac{\hat{F}^e_Q(s;\tau, F_{\tilde{X}}) - \hat{F}_Q(s;\tau, F_{\tilde{X}})}{\hat{\sigma}^e_\tau} \right|$$

3.5. **Asymptotic Theory.** Asymptotic results for the estimators builds on existing results in Chernozhukov et al. (2010, 2013); Melly and Santangelo (2015) and CFNSV pertaining to the DR estimator and the operators involved at each step of the estimation. Starting from the FCLT for the DR estimator (Chernozhukov et al., 2013), it suffices to show that each operator map involved at each subsequent stage of estimation is Hadamard differentiable. The proof for each of these operators can be found, for example, in Chernozhukov et al. (2010, 2013); Melly and Santangelo (2015) and CFNSV. A functional central limit theorem (FCLT) and a bootstrap FCLT for the estimators of the distributions of interest follow by applying the chain rule of Hadamard differentialbility and the functional Delta method. The exposition here omits a full derivation of the FCLT, which is tedious and mainly requires careful bookkeeping about the cascading stochastic processes. It provides little additional insight into the main point of the paper, beyond providing a theoretical background for conducting inference by using the exchangeable bootstrap.

3.6. **Illustrative simulations.** To illustrate the performance of the proposed estimation strategy, this section presents the results of a small-scale simulation study based on the job-choice example in section 2.2.

In the following simulation exercise, the dgp is such that $\log(a_0, a_1)$ follows a normal distribution with mean $(0.5, 0)$ and variance $\begin{pmatrix} 0.5 & -0.1 \\ & 0.3 \end{pmatrix}$. This mimics the case where one sector has amenities that are higher on average than the other sector, but also less dispersed. α has a uniform distribution on the interval $[0.25, 0.75]$. In the main specification, β has a discrete support $\{1, 2, \ldots, 10\}$ with equal probability for each point of the support. An alternative specification, $\beta_i = 1$ for all i, is of particular interest. In this case, the traditional procedure (hereafter, WZ2018) is correctly specified: the returns are linear and separable in income and amenity, and the resolvable uncertainty follows a Gaussian distribution (thus, very close to the assumed logistic distribution).

The choice experiment is assumed to elicit stated preferences for values (y_{0it}, y_{1it}), which are quantiles of a standard log-normal distribution. The stated preference P_{it} are reported with an error e_{it}, such that $P_{it} = m_{it} + e_{it}$, where $m_{it} = m(y_{0t}, y_{1t}, \eta_i)$ and e_{it} has a uniform distribution on the interval $[\pm w \times m_{it}(1 - m_{it})]$. This specification allows considering four cases:

(1) the case of negligible measurement error $w = 0.1$,
(2) the case of moderate measurement error, $w = 0.5$, where the stated preferences of half of the population are off by 4 pp, and for 20 percent of the population by 7pp,

(3) the case of severe measurement error, $w = 1$, where the stated preferences of half of the population is off by 8 pp, and for 20 percent of the population by 15pp.
(4) and the case of increasing variance of measurement error, $w = 0.1t$.

The simulation exercise also investigates the effect of rounding to the nearest 5 or 10 percent, which is common in stated preference experiments (Manski, 2004; Manski and Molinari, 2010).

Using both the traditional (hereafter, WZ2018) and the proposed procedure (hereafter, 2S-KR/DR for two-step Kernel regression or Distribution regression), this section compares the estimated function $F_Q(s; \tau, F_{\tilde{X}})$ for $\tau = 0.20, 0.25, \dots, 0.80$ respectively, and for the counterfactual values $\tilde{X} = (\tilde{y}_0, \tilde{y}_1) = (0.5, 0.7)$. It considers $T = 5, 10$, and 20 scenarios, $N = 500, 1,000$, and replicates the estimation for $N_{sim} = 50$ samples in each case. In the baseline scenario, the measurement error is assumed to be the intermediate case ($w = 0.5$), and rounding to the nearest 5 percent. The 2S-KR/DR simulation considers $\hat{V}_\infty = \hat{V}_{10}$, for which the correlation between successive iterated values exceeds 0.98 in the overwhelming majority of the simulated samples.

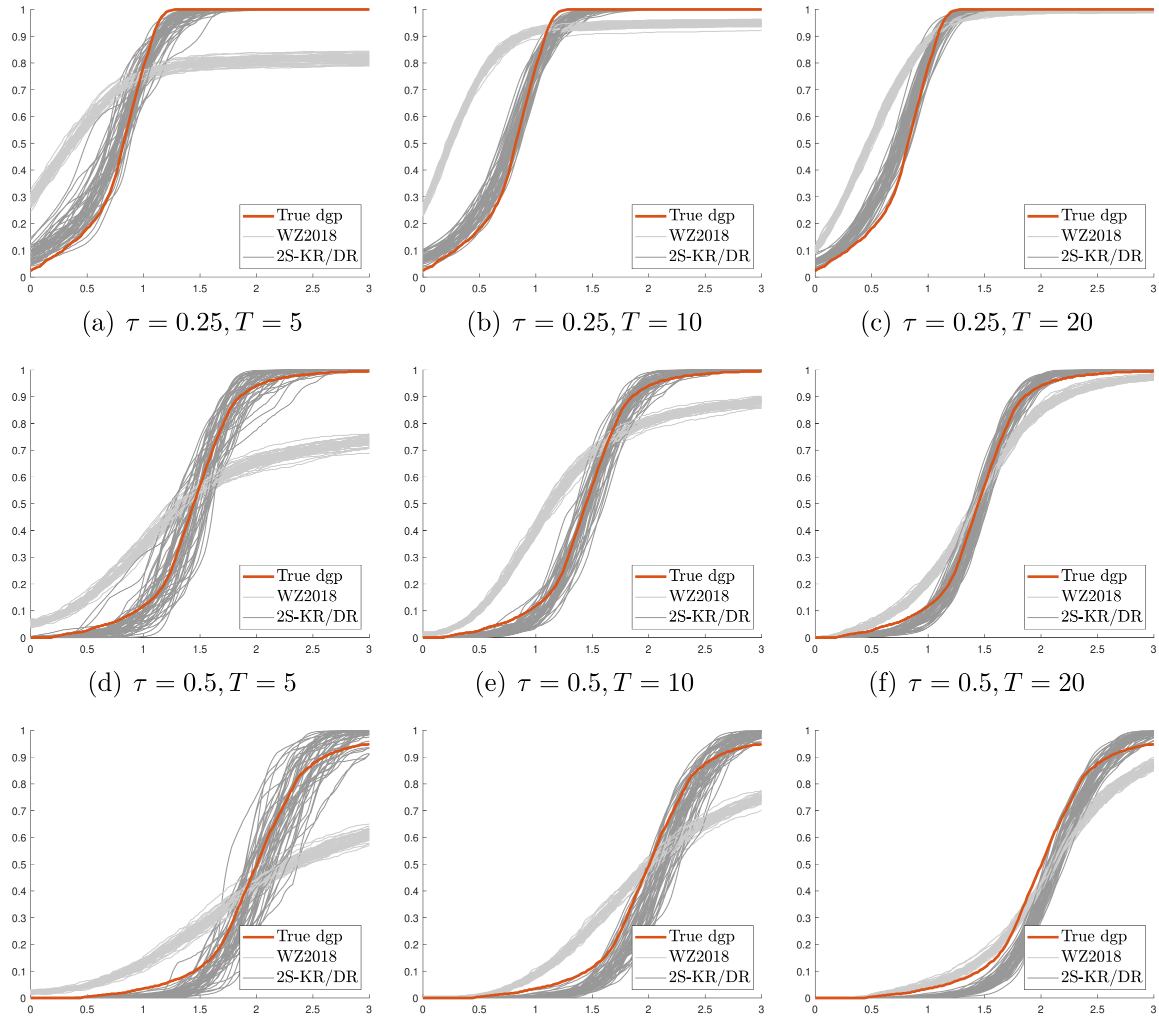
The performance of each procedures is measured through the root-integrated-square-bias (RISB), which averages the squared bias between the true function and its estimate, at N_S points on the support, over all simulations,

$$RISB = \left[\frac{1}{N_{sim}} \sum_{j=1}^{N_{sim}} \frac{1}{N_S} \sum_{s=1}^{N_S} \Big(\hat{F}_{Q,j}(s; \tau, \tilde{X}) - F_Q(s; \tau, \tilde{X}) \Big)^2 \right]^{1/2}.$$

More details are presented in Appendix D, which also collects results from other specifications, for example; (i) a severe measurement error to show the robustness of the procedure, (ii) a coarser rounding (to the nearest 10 percent), (iii) the case $\beta_i = 1$ for all i, where the traditional procedure is correctly specified, or (iv) an increasing variance of the measurement error. To save space, results are only reported for $\tau = 0.25, 0.50, 0.75$

Table 3.1 and Figure 3.2 illustrate two facts: first, the traditional procedure (light grey) may be significantly biased and fail to capture the distribution of quantiles, especially at small $T(\leq 10)$. The bias is reduced, though not eliminated with $T = 20$. A cause of this poor performance is the misspecification of the model. Table D.1 in Appendix D shows that, in the case where the parametric model is correctly specified, the bias remains substantial for $T = 5$ but decreases quickly with longer panels. Second, although it is slightly noisier, 2S-KR/DR outperforms the traditional procedure and exhibits smaller bias for all T considered. The gains are large with the RISB reduced as low as one sixth of its value using WZ2018. The performance is similar whether using KR or DR in the second stage, with a small advantage of KR.[6] Even in the case where the parametric model is correctly specified, the 2S-KR/DR can outperform the traditional procedure (see Table D.2). 2S-KR/DR is robust to measurement error and rounding. Although it performs less

[6]When used in the first stage, DR produces less noisy estimates (not reported).



(a) $\tau = 0.25, T = 5$ (b) $\tau = 0.25, T = 10$ (c) $\tau = 0.25, T = 20$

(d) $\tau = 0.5, T = 5$ (e) $\tau = 0.5, T = 10$ (f) $\tau = 0.5, T = 20$

(g) $\tau = 0.75, T = 5$ (h) $\tau = 0.75, T = 10$ (i) $\tau = 0.75, T = 20$

FIGURE 3.1. Simulated $F_Q(s; \tau, F_{\tilde{X}})$: WZ2018 compared to 2S-KR/DR with DR (first stage) + KR (second stage).

Specification: $\alpha_i \sim \mathcal{U}[0.25, 0.75]$, $\beta_i \in \{1, 2, \ldots, 5\}$, $\tau = 0.25, 0.5, 0.75$, $(\tilde{y}_0, \tilde{y}_1) = (0.5, 0.7)$, $T = 5, 10$, and 20 scenarios, $N = 1,000$, $N_{\text{sim}} = 50$.

well in the lower quartile when the measurement error is severe, it still outperforms the traditional procedure (see Table D.2).

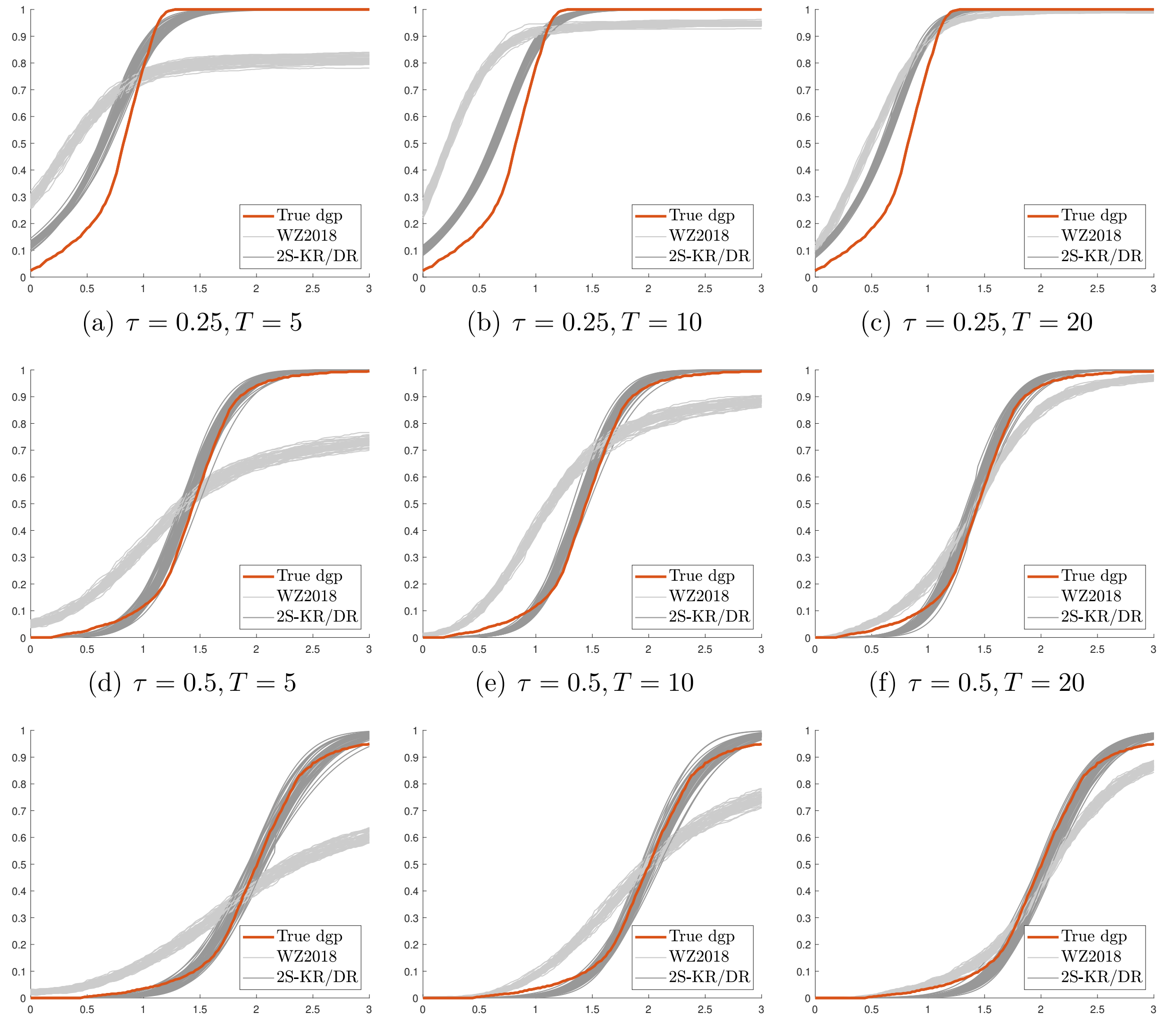
4. Preference for public sector opportunities among young Ivorians

In line with the running example, this section estimates high-ability students' valuation of a job in the public sector. It is worth noting that the methodology of Wiswall and Zafar (2018) is not implementable with the data at hand, because the pseudo-panel is too short (T=5) compared to the number of attributes considered (ten attributes).[7]

4.1. Data.

The survey was conducted in the first week of February 2024 at two elite higher education institutions in Ivory Coast (*Institut National Polytechnique Houphouët-Boigny*,

[7]In light of the high familiarity of the sample, we view the empirical application as a case where measurement is negligible. The results remain very stable even if not considering a measurement error.



(a) $\tau = 0.25, T = 5$ (b) $\tau = 0.25, T = 10$ (c) $\tau = 0.25, T = 20$

(d) $\tau = 0.5, T = 5$ (e) $\tau = 0.5, T = 10$ (f) $\tau = 0.5, T = 20$

(g) $\tau = 0.75, T = 5$ (h) $\tau = 0.75, T = 10$ (i) $\tau = 0.75, T = 20$

FIGURE 3.2. Simulated $F_Q(s; \tau, F_{\tilde{X}})$: WZ2018 compared to 2S-KR/DR with DR (first stage) + DR (second stage).

Specification: $\alpha_i \sim \mathcal{U}[0.25, 0.75]$, $\beta_i \in \{1, 2, \ldots, 5\}$, $\tau = 0.4, 0.5, 0.6$, $(\tilde{y}_0, \tilde{y}_1) = (0.5, 0.7)$, $T = 5, 10$, and 20 scenarios, $N = 1,000$.

INP-HB and *Ecole Nationale de Statistiques et d'Economie Appliquée*, ENSEA), that train students in STEM (Science, Technology, Engineering and Mathematics) degrees, and Statistics, Business and Economics, respectively. Entry in each of these institutions is through a selective exam. Together, they bring about 900 to 1,000 new graduates on the labour market each year. A number of well-known alumni of these two universities serve in high-ranked positions both in the Ivorian government and in private corporations.[8] Hence, the sample represents some of the best students in the country.

The context is one of skill shortage where firms struggle to satisfy their demand for labour. Quotes from two human resources directors illustrate vividly the skill shortage

[8] Accurate proportions of those who are civil servants and those working in the private sector are not available yet, as these information have only started to be systematically collected recently. Among tertiary educated, the public sector offers about half of formal wage employment in Côte d'Ivoire (Girsberger and Meango, 2022).

Baseline DR (first stage) + KR (second stage)

		$N = 500$						$N = 1,000$					
		RISB			Std. deviation			RISB			Std. deviation		
	τ	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
WZ2018	$T = 5$	0.240	0.213	0.190	0.355	0.570	0.501	0.241	0.214	0.190	0.355	0.570	0.494
(1)	$T = 10$	0.260	0.162	0.128	0.420	0.787	0.682	0.260	0.161	0.129	0.426	0.794	0.686
	$T = 20$	0.157	0.064	0.074	0.654	0.915	0.749	0.157	0.063	0.075	0.650	0.918	0.752
2S-KR/DR	$T = 5$	0.067	0.062	0.080	0.813	1.055	0.924	0.060	0.057	0.072	0.844	1.062	0.902
(2)	$T = 10$	0.052	0.040	0.065	0.788	1.050	0.885	0.044	0.037	0.071	0.852	1.063	0.930
	$T = 20$	0.044	0.031	0.072	0.850	1.076	0.886	0.048	0.031	0.067	0.824	1.059	0.895
Ratio	$T = 5$	3.588	3.453	2.382	0.436	0.540	0.542	3.989	3.758	2.635	0.421	0.536	0.547
(1)/(2)	$T = 10$	4.984	4.004	1.956	0.533	0.750	0.770	5.874	4.315	1.812	0.500	0.746	0.737
	$T = 20$	3.565	2.098	1.028	0.770	0.851	0.845	3.274	2.063	1.112	0.789	0.867	0.840

DR (first stage) + DR (second stage)

		$N = 500$						$N = 1,000$					
		RISB			Std. deviation			RISB			Std. deviation		
	τ	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
WZ2018	$T = 5$	0.241	0.215	0.189	0.380	0.592	0.510	0.240	0.215	0.190	0.353	0.581	0.493
(1)	$T = 10$	0.262	0.163	0.127	0.426	0.783	0.670	0.261	0.162	0.129	0.410	0.774	0.663
	$T = 20$	0.157	0.065	0.075	0.645	0.920	0.751	0.157	0.064	0.075	0.649	0.918	0.756
2S-KR/DR	$T = 5$	0.097	0.041	0.035	0.726	1.031	0.915	0.097	0.037	0.029	0.749	1.048	0.906
(2)	$T = 10$	0.095	0.038	0.026	0.750	1.061	0.916	0.099	0.034	0.027	0.740	1.054	0.894
	$T = 20$	0.107	0.036	0.032	0.748	1.067	0.931	0.109	0.033	0.033	0.743	1.071	0.930
Ratio	$T = 5$	2.480	5.195	5.360	0.523	0.575	0.558	2.468	5.780	6.478	0.472	0.555	0.544
(1)/(2)	$T = 10$	2.748	4.310	4.825	0.568	0.738	0.732	2.638	4.768	4.804	0.555	0.734	0.741
	$T = 20$	1.469	1.810	2.344	0.863	0.863	0.806	1.440	1.935	2.272	0.873	0.857	0.812

TABLE 3.1. Root-integrated-square-bias and Standard deviation in baseline scenario using DR +KR and DR + DR

Note: The table summarises the results of the simulations for two cases: DR + KR: Distribution regression is employed in the first stage, and Kernel regression in the second stage. DR + DR: Distribution regression is employed in the first stage, and in the second stage. It compares these results to the generalisation of Wiswall and Zafar (2018) (WZ2018), as detailed in Appendix D. The last block of lines in each panel represents the ratio between the RISB and the standard deviation.

on one side: 'Industrial engineers are rare, only the INP-HB trains good profiles but not enough of them are being trained to meet the needs of all the companies in Côte d'Ivoire.', and the comparative advantage of students from these institutions on the other side: 'There are other schools that train industrial engineers but we will not entrust our factories to these young people trained in schools other than the INP-HB.'[9]

The target population was students in their last year, after either of a 3-year, 4-year or 5-year degree. The survey was implemented as a Computer Assisted Personal Interview, with five enumerators meeting students in or around their dorm-rooms. The usable sample

[9]Both cited from IOM (2023).

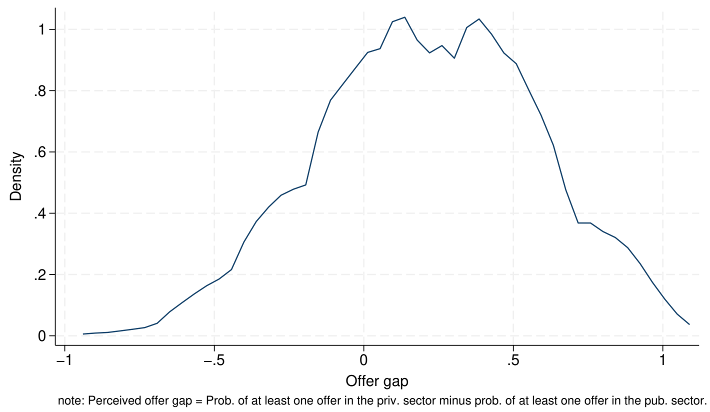


FIGURE 4.1. Perceived offer gap distribution

Note: The figure represents the probability distribution of the average perceived offer gap between two sectors. The offer gap is defined as the perceived probability of receiving a private-sector offer minus the same perceived probability for a public sector offer.

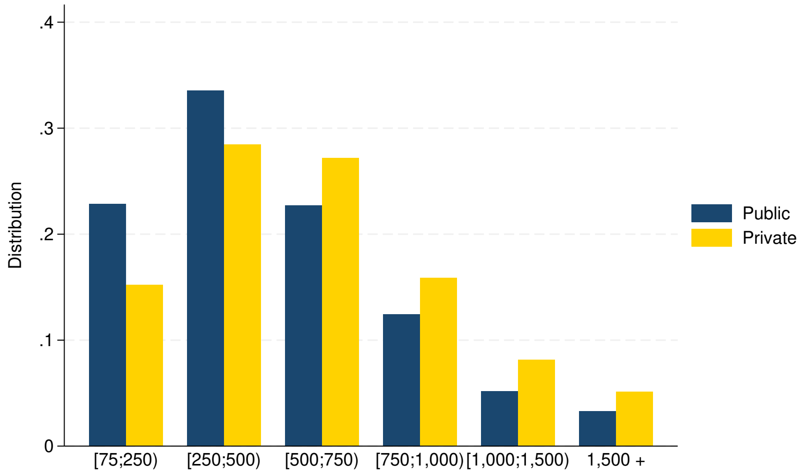
covers students 587 interviews. The vast majority of respondents are between 21 and 25 year old, with a proportion of 30 percent women, which is representative of this population.

The survey is unique in its scope as it provides unprecedented insights in the perception and preferences of some of the best and brightest young Africans about their labour market prospects. On the one hand, it collects novel information on young Ivorians' perception about job characteristics in the public and private sector. The perceptions garnered consist of beliefs about offer arrival rates, wage distribution, job destruction rate, and the likelihood of moving up in the hierarchy or obtaining a wage rise. On the other hand, the survey elicits preferences for jobs in different sectors. First, it asks students to rank different employers by their attractiveness. Second, it conducts a stated choice experiment where respondents are invited to state their preferences over two job offers, one in the public, and one in the private sector. Jobs differ by their wage and non-wage characteristics (employer, hours worked, likelihood of losing the job and likelihood obtaining a wage rise).

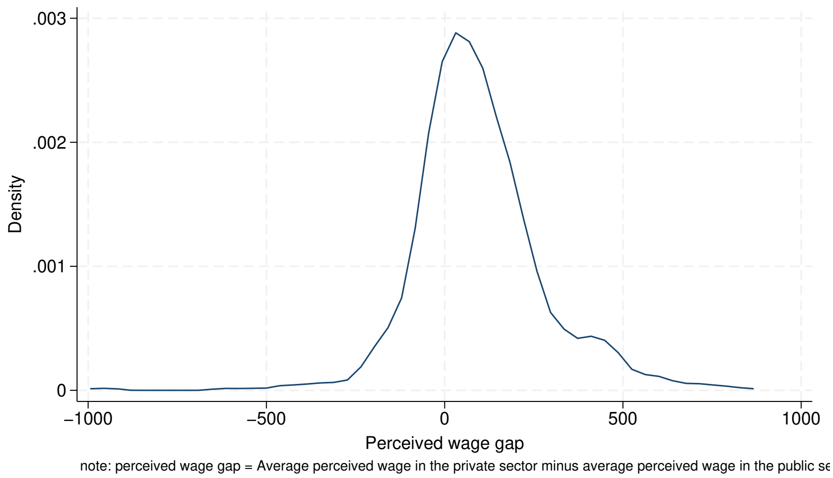
4.2. **Perception of the labour market.** Perceptions in this section are elicited by asking the students to think about 20 students with similar characteristics as theirs (same age, gender, education, and family background).[10] To elicit the perceived sector-specific wage distribution, the students are instructed to think instead of one wage offer for each in a given sector.[11] Perceptions of students in either track are broadly similar, aside from a level difference in wage. Therefore, they are pooled in the following exposition.

[10]For example, for the likelihood of receiving a job offer from the public sector: *Imagine 20 young Ivorians your age, i.e. [age variable]. Imagine they have an education similar to yours, i.e. [degree and field of study], all [male / female] like you, with similar family backgrounds. Think of these young people as they enter the job market. Out of these 20 young people, how many do you think will receive at least one job offer in the public sector?*

[11]More precisely, the question states: *Imagine that these twenty young people each receive a job offer in the public sector. These job offers may vary according to the administration or public company offering them. We'll show you categories representing salary ranges, and ask you to allocate the twenty job offers to each range.*



(a) Perceived wage distribution by sector



(b) Average perceived wage gap

FIGURE 4.2. Perceived wage by sector and average perceived wage gap

Note: In Panel (a), each bar represents the reported a probability of a wage offer following within a given bin by sector averaged on the sample. The intervals are in 1,000 CFA Francs. Panel (b) represents the probability distribution of the average perceived wage gap between two sectors. For each individual, for each sector, an average perceived wage is calculated by taking the median of the interval and the mass associated with the interval. The wage gap is defined as the average perceived wage in the private sector minus the average perceived wage in the public sector.

First, three quarters of students perceive that jobs are equally or more abundant in the private sector than in the public sector. The average likelihood of receiving a job offer stands at 0.50 in the public sector, and 0.71 in the private sector. Figure 4.1 shows the distribution of perceived offer rate gap between the public and the private sector (offer rate in private minus public), which is tilted toward positive values. The median individual perceived a 25-percentage-point higher chance to receive a job offer from the private sector. Yet, a quarter of the population perceives better opportunities in the public sector.

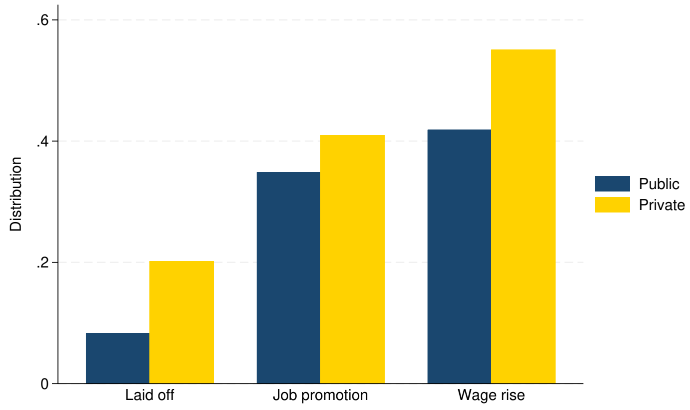


FIGURE 4.3. Other perceived characteristics by sector

Note: Each bar represents the reported a probability by sector averaged on the sample. 'Laid off' represents the probability of being laid off within two years, 'job promotion', the probability of being promoted to a more senior position, 'wage rise', the probability of receiving a 20 percent wage rise within two years.

Students also believe that the private sector provides jobs with higher pay. Figure 4.2 shows in Panel (a) the distribution of wages by sector averaged across all respondents. The average distribution of wages in the private sector is clearly shifted to right, with less offers in the bottom categories and more offers in the middle and top categories. Panel (b) constructs an average perceived wage for each individual and each sector, and computes the distribution of the average wage gap in the population. Four out of five respondents perceive that jobs are, on average, equally or better paid in the private sector than in the public sector. The perceived average gain amounts to about 96,600 CFA Franc, that is 16.7 percent of the perceived average wage in the public sector. Note that about two third of the population perceives the wage distribution in the public sector to be more concentrated than the one in the private sector.

Furthermore, students perceive private-jobs as less secure, but offering more possibility for career and wage progression (Figure 4.3). The perceived job destruction rate after two years in the private sector is, on average, 20 percent, more than double the job destruction rate in the public sector (8 percent). Concomitantly, opportunities for moving up the job ladder are seen as better in the private sector, with 6 percentage points (pp) gap, on average in the chance of obtaining a career advancement and a 13 pp gap in the chance of obtaining a 20 percent wage rise within two years.[12]

Finally, respondents are asked to describe in three words their opportunities in each sector. The top-three words describing public-sector opportunities are 'guarantee', 'stable', and 'flexible'.[13] By contrast, the top three words describing the private-sector opportunities are 'stressful', 'lucrative', and 'competitive.'

[12]All differences are significant at standard levels.

[13]Within the Top-10 words, one finds: 'security', 'reliable', 'insurance'. The word 'corruption' also appears in the Top-10 list, while it is never mentioned in relationship with the private sector.

The general insight from these results is that public and private sector jobs are perceived very differently in the population. One the one hand, the private sector appears as a dynamic sector with more abundant, better paid jobs, and more opportunity for progression. However, these benefits are tied with stressful work conditions. On the other hand, the public sector appears as a secure, stable alternative, even if less well paid. It is important to note that beyond these general trends, there is substantial heterogeneity in individual perceptions. This explains the heterogeneity in preferences which we turn to in the next section.

It is difficult to compare the stated beliefs to actual labour market statistics. Our efforts to construct a comparable sample of workers from existing (survey) data was unsuccessful. For example, the latest publicly available labour force survey ERI/ESI 2017 contains only 69 workers aged between 25 and 34 with university education and available wage information, and may not represent well students from these elite universities. Comparing with other data sources is still instructive. The average wage perception of those high-ability student securely places them in the top quartile of the wage distribution in Côte d'Ivoire (compare with Christiaensen and Premand, 2017, p.124), which reflects well their ability. Looking at the distribution of their wage perception, students seem aware that even for them, offers below the mean wage in STEM occupations can be frequent (23 percent for public-sector jobs, 15 percent for private-sector jobs).[14] The perceived wage penalty for public sector jobs may seem counter-intuitive given the received knowledge of a wage premium in the public sector. Girsberger and Meango (2022) using a regional survey data conducted in 2003 for francophone countries in West Africa, including Côte d'Ivoire, find that a public-sector wage premium exists only for low education groups, not for tertiary educated. However, destruction rates in the public sector are significantly lower than in the private sector. Gindling et al. (2020), comparing 68 countries including several low-income countries, also conclude that high skilled public-sector employees in several of those countries pay a wage penalty for working in the public sector. Hence, the evidence gathered suggest that students' perception about the labour market are not too far off. Following their career development and comparing those to their initial beliefs is a fruitful research avenue that we hope to pursue in future.

4.3. **Preferences over sectors.** Respondents are asked first to rank five type of employers according to their attractiveness: public administration, public-sector firms, small and medium-sized (private) enterprises (SMEs), large-sized enterprises, and international institutions.[15] Figure 4.4 presents the result of this exercise. International institutions are at the top of the ranking, being ranked as the most attractive employers by more than half of the sample. They are closely followed by large private firms, which are ranked first

[14]According to the ILO database, the 2019 average wage in STEM occupations (for all education levels) was 227,876 CFA Franc.

[15]This includes regional institutions, as the WAEMU or ECOWAS and their agencies, panafrican institutions, for example, the African Union and its agencies, or intercontinental institution, for example, the United Nations and its agencies.

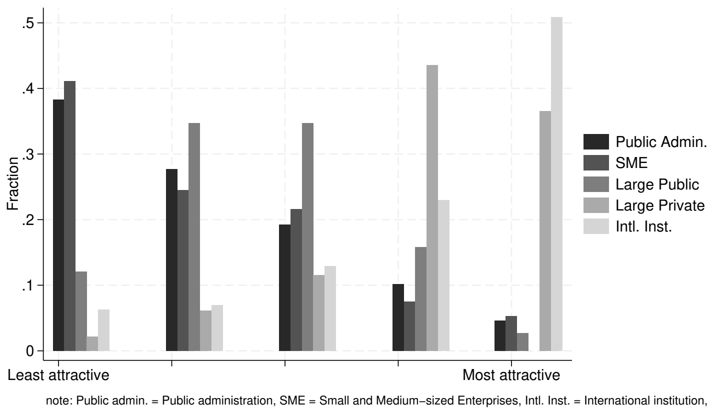


FIGURE 4.4. Attractiveness of sectors

Note: The figure represents the average ranking of employers. Ranks go from 1 (Most attractive) to 5 (least attractive). International institutions and large private firms are, in general, the most attractive, SMEs and public administration are the least attractive.

by close to four out of ten respondents. Large public firms are generally in the middle position. Public administration and SMEs share the bottom place with seventy percent of the population ranking them as one of the two least attractive employers. Thus, it appears that the divide between public and private is not enough to explain preferences and the ranking is influenced also by the size of the firm.

To obtain a deeper insight in individual preferences, survey participants are presented with a choice experiment with five pairs of hypothetical job offers, one offer from the public, and one from the private sector. Each pair specifies:

(i) the type of employer, public administration or public-sector firm for the public-sector job, SME or large-size firm for the private,
(ii) the typical number of weekly hours associated with the job, to capture the flexibility of the job,
(iii) the percentage of employees in this firm/administration who lose their job after two years,
(iv) the percentage of employees who are promoted after two years,
(v) the monthly wage attached to the job.[16]

For each pair/scenario, they are asked to state the probability of choosing either sector. The first scenario always refers to identical offers, i.e, the scenario equates all items (ii) - (v) above for an offer from the public administration against an offer from an SME.[17] For the remaining pairs, the attributes are randomly drawn from a support displayed in Table E.1.

Figure 4.7 displays the distribution of the probability to accept the public sector offer for the first scenario, where offers are identical. The histogram shows evidence of rounding at

[16]In Côte d'Ivoire, wages for salaried work are typically expressed in monthly unit.

[17]The employers are 'public administration' and 'SME' respectively, the weekly hours worked 40, the chance of losing job is 5 percent, the chance of job job promotion, 10 percent, and the starting salary 750,000 CFA Francs (25 percent above the average perceived wage).

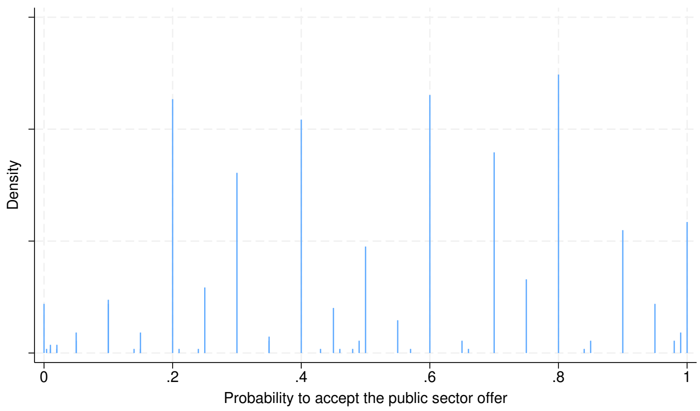


FIGURE 4.5. Probability of choosing the public sector for two identical offers

Note: The figure represents the reported probability to choose a public-sector offer, when offer are otherwise identical. The employers are 'public administration' and 'SME' respectively, the weekly hours worked 40, the chance of losing job is 5 percent, the chance of job job promotion, 10 percent, and the starting salary 750,000 CFA Francs (25 percent above the average perceived wage).

the nearest 10 percent. Otherwise, there is no conspicuous use or extreme values or of the value 50 percent that is sometimes found in stated preference data. The relative quality of this data can be partly attributed to the respondents' experience with probability, which is higher than in the average population. A striking insight is that the answers are balanced on both side of the 50-percent mark. The median answer is 55. About 35 percent of the sample states a probability between 0.15 and 0.45, whereas about 36 percent gives an answer between 0.55 and 0.85. Thus, it appears that for identical offers, there is no sweeping preference for one sector over the other. On the contrary, preferences are very heterogeneous in the population. Besides, even when respondents express a preference for one sector over the other, the great majority does not exclude the possibility of accepting the competing offer (91 percent gives an interior solution). This hints at a high prevalence of (resolvable) uncertainty. The stated choice probabilities are the main inputs to infer the distribution of perceived ex ante returns for a public-sector job.

4.4. **Distribution of perceived ex ante returns for a public-sector job.** Following the development of Section 2, this section defines the ex ante returns to a public-sector job S as the pecuniary transfer (or the wage change in the private sector offer) that would make a job-seeker indifferent between a public-sector job (option 1) and a private sector job (option 0). Because uncertainty is resolved sequentially, job-seekers do not know their returns S. Instead, they entertain an individual-specific distribution $F_{S,i}$.

In the empirical application, $r(X_{it})$ in equation (11) is linear in each choice attribute. An interaction term is added between the two random variables measuring monthly wage in each sector. Similarly, $w(X_{it}, \hat{\epsilon}_{it})$ in equation (13) is linear in each choice attribute

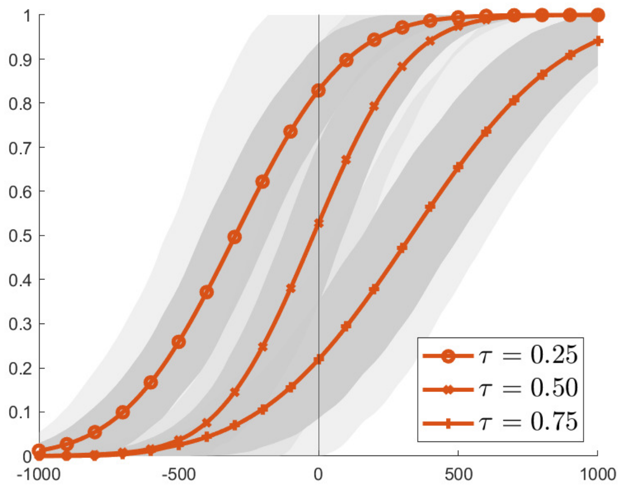


FIGURE 4.6. Distribution of quantiles $\hat{F}_Q\left(.; \tau, F_{\tilde{X}}\right), \tau = 0.25, 0.50, 0.75$ for identical offers.

Note: Offers mimic the average perception about a public-sector job offer. Working hours are set to 40 hours. The dark-grey area shows the 90 percent pointwise confidence interval. The light-grey area shows the 90 percent uniform confidence interval.

and the measurement error enters as a separable variable. An interaction term is added between the two random variables measuring monthly wage in each sector.

Figure 4.6 represents the distributions of quartiles $\hat{F}_Q\left(.; \tau, F_{\tilde{X}}\right), \tau = 0.25, 0.50, 0.75$. It shows these distributions for identical offers that both mimic the typical public-sector offer. The construction of offers uses the respondents' perceptions described in section 4.3 to compute average values of the public-sector-job attributes in the population. $F_{\tilde{X}}$ is set to take these average values. For example, the wages are set at $525,000$ CFA Francs, the probability to lose a job both in the public and the private sector is set to 0.08, and the chance to obtain a job promotion to 0.30. The distribution of median returns (middle line) suggests that for identical offers, the population of job-seekers is almost equally divided between those who perceive positive (43.2 percent) and negative median returns from choosing the public-sector offer. Yet, returns are very heterogeneous and can be large in absolute values. At the top decile, the returns are as large as 60 percent of the average perceived wage in the public sector, whereas at the bottom decile they represent a loss of about 69 percent of the average perceived wage in the public sector.

The horizontal distance between the first and third-quartile distribution gives an indication of the individual-specific IQR. The distributions are well-separated, which hints at a high prevalence of uncertainty about ex ante returns. More formally, the distribution of individual-specific IQR can also be calculated, thanks to equation (21) in Theorem 2. For the majority of the population (contained between the 10-*th* and the 90-*th* percentile) the estimated values range between 0.8 and 1.6 times the average perceived wage in the public-sector. These results confirm the high prevalence of uncertainty in the population. It implies that the job-seekers have a high value of waiting and collecting more information

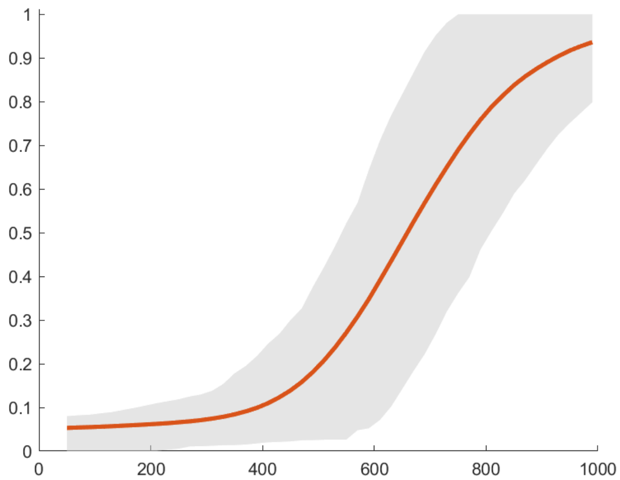


FIGURE 4.7. Distribution of individual-specific IQR

Note: Offers mimic the average perception about a public-sector job offer. Working hours are set to 40 hours. The grey area shows the 90 percent pointwise confidence interval.

on the sectors instead of committing ex ante to one sector (on the option-value arising from colecting information, see, for example, Gong et al., 2020; Méango and Poinas, 2023).

Figure F.2 in Appendix F shows the distributions for perceived public-sector amenities: wages are set to be equal in both sectors so that there is no pecuniary gain from being in the private sector. Only perceived amenities vary.[18] The distributions are somewhat flatter, reflecting the existence of dispersed beliefs. At the median returns, about 6 out of 10 students perceive positive amenities from being in the public sector. Close to 15 percent of the population perceives amenities larger than the average wage in the public sector.[19]

To investigate further the value of amenities, one can compute the distribution of quantile treatment effect (QTE) for the observed choice attributes using equation (19) in Theorem 2.[20] At any quantile τ, they answer the question: 'how does the treatment (choice attribute) changes the τ-quantile of returns for any given individual?' Figure 4.8 displays the distribution of QTEs on the distribution of median returns for treatments

[18]The wage is set at the individual-specific average perceived wage in the private sector. The survey missed a question about the number of weekly working hours expected in the public or the private sector. The number of working hours in the public sector is set to 41.4 and the working hours for the private sectors are set to 45.8 according to Christiaensen and Premand (2017). The average working time for salaried workers as calculated from the publicly available Ivorian labour force survey (ERI-ESI 2017) is 48 hours. Comparable numbers for other countries of francophone west Africa range between 45 and 51 hours.

[19]Mangal (2024) uses a sample of 147 candidates preparing for competitive exams for government jobs to infer a lower bound on the total value of a government job, including amenities. He finds that the amenity value of a government job in Pune, India, comprises at least two-thirds of total compensation. The above estimates of median returns are lesser in magnitude and provide a more nuanced picture.

[20]Unlike traditional QTE estimates, these are individual-specific QTEs. For the traditional QTEs to produce individual-specific treatment effect, one also needs to assume rank invariance.

that equate the individual's perceptions.[21] For example, panel (a) changes the individual's perception about the perceived probability of being laid off in the public-sector to equate it to the probability of being laid off in the private sector. Similarly, panel (b) equates the perception about the probability of obtaining a job promotion and Panel (c) equates the working hours. As the gap in the probability of being laid off is closed off, the median returns to a public-sector offer decrease by up to one fifth of average wages in the public-sector, for the majority of the population contained within the two extreme deciles. This reflects the importance of job stability in those returns. The effect of equating the perceived probability of a job promotion depends strongly on whether this probability is believed to be larger in the private or the public sector. The effect are strong in either case, with the top quartile perceiving gains larger than one sixth of the average wage in the public sector. Finally, in our counterfactual exercise where the number of hours in the public sector is increase from 41.4 to 45.8, the returns to public-sector offer decrease by about one fourth to one third of average perceived wages in the public-sector.

Overall, the above results are consistent with the verbal description of the students who value the public sector for its stability and flexibility, and identify the private sector foremost with stressful but pecuniary rewarding working conditions. Beyond the choice attributes controlled for, the perceived returns are very heterogeneous, implying that a significant proportion of the population views both sectors as very different. Some people have a high value for the public-sector offer, whereas other have an equally high value for the private sector offer. Nevertheless, there is considerable uncertainty about those returns, which leaves room for substantial changes between the time of elicitation and the time of decision.

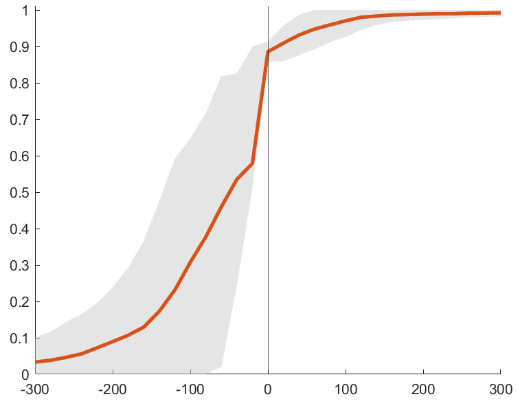
4.5. **Policy implication.** Given the context of shortage in high-skilled workers, this section investigates the externalities imposed by the public sector on the private sector in the labour market for top-skilled workers. One metric for these externalities is the cost for the private sector to attract an additional fraction of top-skilled workers. As discussed in the introduction, a predictor of the realised distribution of returns, F_S, is necessary to conduct this counterfactual analysis.

From an ex ante perspective, the resolution of uncertainty will generate a mixture of shocks $\eta^*|\eta$ in the population. For some individuals, the realised uncertainty $\eta^*|\eta$ will correspond to lower quantiles τ, whereas for others, it will correspond to higher quantiles τ. Thus, if preferences are stable, and for a given distribution $F_{\tilde{X}}$, the realised distribution of returns will be a mixture of the ex ante distribution of quantiles $F_Q(s;\tau,F_{\tilde{X}})$.

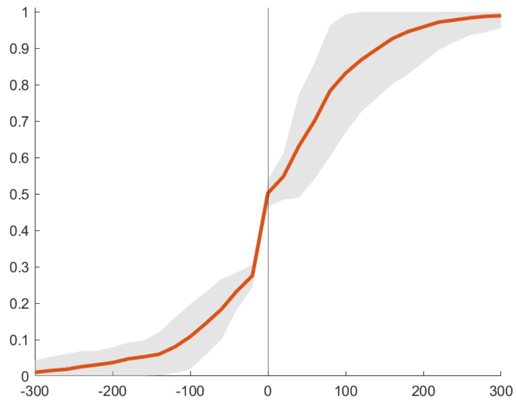
If the analyst considers that all students have equally valid prior on the distribution of $\eta^*|\eta$ and shocks are i.i.d., the best predictor for the realised distribution of returns is a mixture of the distributions of quantiles with equal weights; that is:

$$F_S(s;F_{\tilde{X}}) := \int_0^1 F_Q(s;\tau,F_{\tilde{X}})\omega_\tau d\tau \tag{15}$$

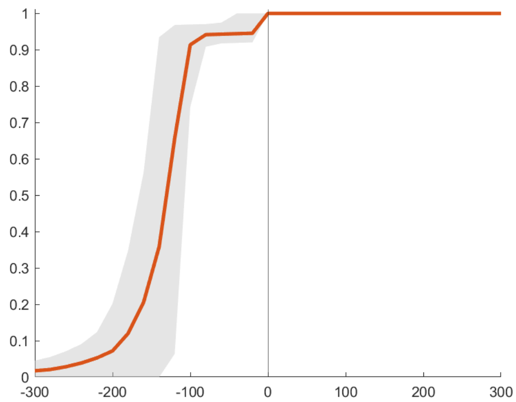
[21]Incidentally, the QTEs are similar at the first and third quartiles. See, for example, figure F.3.



(a) Being laid off



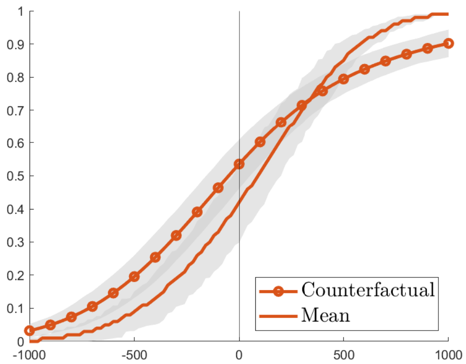
(b) Obtaining a job promotion



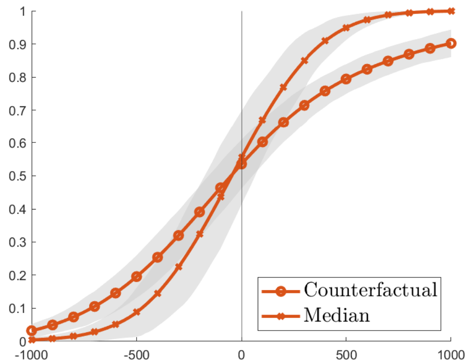
(c) Working hours

FIGURE 4.8. QTEs for choice attributes

Note: The treatment consists in closing the perceived gap between choice attributes one by one. Panel (a) changes the individual's perception about the perceived probability of being laid off in the public-sector to equate it to the probability of being laid off in the private sector. Similarly, panel (b) changes the perception about the probability of obtaining a job promotion in the public sector to equate it to the perceived probability of obtaining a job promotion. Panel (c) changes the working hours from 41.4 in the public-sector, to 45.8 as in the private sector offer.



(a) against the distribution of average returns



(b) against the distribution of median returns

FIGURE 4.9. Comparing the predicted distribution of realised returns to the distribution of average returns and median returns.

Note: The predicted distribution of realised returns ('counterfactual') is calculated by equation (15), using quantiles $\tau = 0.05, 0, 10, \ldots, 0.95$. The distributions of mean and median returns are calculated using Theorem 1 and 2. $F_{\tilde{X}}$ is set for identical offers that mimic the public sector offer.

where $\omega_\tau = 1$ for all τ. Instead, the analyst may entertain a different prior: for example, if individuals are deemed to be too optimistic about the resolvable uncertainty, realised shocks would be overwhelmingly concentrated on the lowest quantiles. Thus, a policy maker may wish to assign low weights to the highest quantiles. If instead, individuals are assumed to place too much weight on extreme quantiles, the policy maker can discount extreme quantiles. A policy maker who bases their analysis solely on the distribution of median returns assigns a weight one to the median and zero to all other quantiles.

Before turning to the result of the counterfactual exercise, it is important to note that the predicted distribution of realised returns may correspond neither to the distribution of a particular quantile nor to the distribution of average returns. Previous literature has emphasised the estimation of these two objects, which are not the most relevant from the perspective of ex ante policy evaluation. Figure 4.9 demonstrates the discrepancy using the empirical application. The counterfactual distribution of returns using equation (15) is flatter than the distributions of average and median returns, and would understate the cost elasticity of an expansion. Intuitively, the latter distributions fail to account for the fact that some members of the population will receive extreme values of their perceived returns rather than average or median values.

Figure 4.10 shows the relationship between the private sector expansion (percentage of additional workers) and the transfers as a proportion of the original wage bill. $F_{\tilde{X}}$ is set for average offers in each sector. For these offers, the population is almost equally split between those with positive returns (43.2 percent) and negative returns to the public-sector offer. This reflects well the equal split on the actual labour market, where the each sector provides about one half on the formal wage employment (see Christiaensen and Premand, 2017, Table A.1, p.43). Achieving a one percent increase of private-sector workers requires

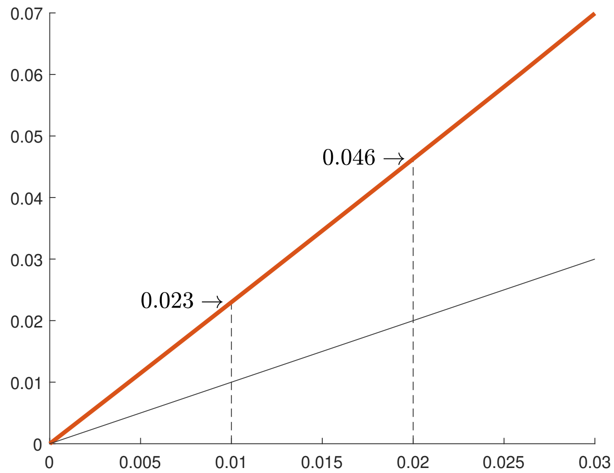


FIGURE 4.10. Estimated transfers to attract additional workers in the private sectors.

Note: The predicted distribution of realised returns is calculated by equation (15), using quantiles $\tau = 0.05, 0, 10, \ldots, 0.95$. $F_{\tilde{X}}$ is set for typical offers in the public and the private sector. The figure shows the relationship between the private sector expansion (percentage of additional workers) and the transfers as a proportion of the original costs/wage bill. A one percent increase of the private sector would entail a 2.3 percent increase of the wage bill.

an increase of 2.3 percent of the wage bill. This appears as an economically significant cost and supports the hypothesis that the competition from a public sector offering more job security might constrain the expansion of the private sector.

5. Conclusion

Stated preference analyses have served mainly two goals: (1) to describe individual preferences over choice attributes, presented in the form of Willingness-To-Pay (WTP) parameters, and (2) to conduct ex ante policy evaluation. These tasks are more complex in an environment where agents are uncertain about their returns and uncertainty is sequentially resolved. This paper provides: (i) an econometric framework to model the elicitation and decision process, (ii) characterisation results for the population distribution of ex ante returns, the related uncertainty, and a policy relevant distribution for the purpose of ex ante policy evaluation. These characterisations results are obtained while accounting for the uncertainty of agents but without appealing to ad-hoc parametric assumptions, for example, that the resolvable uncertainty is extreme value type I. They are new to the stated preference literature.

The paper introduces nonparametric identification results and a novel nonparametric/semiparametric estimation methodology for the objects of interest. The main innovation is to use the cross-sectional variation to identify the distribution of interest. In fact, without measurement error, the distributions of interest are identified with a single cross-section. Repeated observation serves only to deal with measurement error. The proposed procedure shows a significant reduction of the bias for relatively small T.

The paper exploits these new tools to analyse the preference for public sector jobs of high-ability students in Côte d'Ivoire, with data on close to 600 students from two highly selective universities. The data delivers unprecedented insights on the perceptions of Ivorian students about their labour market perspectives: students value the public sector for its stability and flexibility, and identify the private sector foremost with stressful but pecuniary rewarding working conditions. The preferences are very heterogeneous: even when wage offers are identical, some students have a high value for a public-sector job, whereas other have an equally high value for a private sector job. Given these preferences, there is evidence that the presence of competing public-sector offers possibly increases the labour costs for private sector and limits its expansion.

The nonparametric estimation strategy developed for a nonseparable panel model can be of independent interest. Although the systematic study of its property is beyond the scope of the present paper, it should be a fruitful avenue for future research.

References

Adams-Prassl, A. and Andrew, A. (2019). Preferences and beliefs in the marriage market for young brides. CEPR Discussion Paper No. DP13567.

Albrecht, J., Robayo-Abril, M., and Vroman, S. (2019). Public-sector employment in an equilibrium search and matching model. *The Economic Journal*, 129(617):35–61.

Algan, Y., Cahuc, P., and Zylberberg, A. (2002). Public employment and labour market performance. *Economic Policy*, 17(34):7–66.

Almås, I., Attanasio, O., and Jervis, P. (2023). Economics and measurement: New measures to model decision making. Technical report, National Bureau of Economic Research.

Ameriks, J., Briggs, J., Caplin, A., Lee, M., Shapiro, M. D., and Tonetti, C. (2020a). Older americans would work longer if jobs were flexible. *American Economic Journal: Macroeconomics*, 12(1):174–209.

Ameriks, J., Briggs, J., Caplin, A., Shapiro, M. D., and Tonetti, C. (2020b). Long-term-care utility and late-in-life saving. *Journal of Political Economy*, 128(6):2375–2451.

Arcidiacono, P., Hotz, V. J., Maurel, A., and Romano, T. (2020). Ex ante returns and occupational choice. *Journal of Political Economy*, 128(12):4475–4522.

Aucejo, E. M., French, J., and Zafar, B. (2023). Estimating students' valuation for college experiences. *Journal of Public Economics*, 224:104926.

Bénabou, R. and Tirole, J. (2016). Mindful economics: The production, consumption, and value of beliefs. *Journal of Economic Perspectives*, 30(3):141–164.

Bhuller, M., Eisenhauer, P., and Mendel, M. (2022). Sequential choices, option values, and the returns to education. *arXiv preprint arXiv:2205.05444*.

Blass, A. A., Lach, S., and Manski, C. F. (2010). Using elicited choice probabilities to estimate random utility models: Preferences for electricity reliability. *International Economic Review*, 51(2):421–440.

Bowman, A. W. and Azzalini, A. (1997). *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*, volume 18. OUP Oxford.

Boyer, M. M., De Donder, P., Fluet, C., Leroux, M.-L., and Michaud, P.-C. (2020). Long-term care insurance: Information frictions and selection. *American Economic Journal: Economic Policy*, 12(3):134–69.

Burdett, K. (2012). Towards a theory of the labor market with a public sector. *Labour economics*, 19(1):68–75.

Carneiro, P., Hansen, K. T., and Heckman, J. J. (2003). 2001 lawrence r. klein lecture estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review*, 44(2):361–422.

Cavalcanti, T. and Santos, M. (2020). (MIS)Allocation Effects of an Overpaid Public Sector. *Journal of the European Economic Association*, 19(2):953–999.

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.

Chernozhukov, V., Fernández-Val, I., and Melly, B. (2013). Inference on counterfactual distributions. *Econometrica*, 81(6):2205–2268.

Chernozhukov, V., Fernández-Val, I., Newey, W., Stouli, S., and Vella, F. (2020). Semiparametric estimation of structural functions in nonseparable triangular models. *Quantitative Economics*, 11(2):503–533.

Chernozhukov, V. and Hansen, C. (2005). An iv model of quantile treatment effects. *Econometrica*, 73(1):245–261.

Christiaensen, L. and Premand, P. (2017). Côte d'ivoire jobs diagnostic: employment, productivity, and inclusion for poverty reduction.

Cunha, F., Heckman, J., and Navarro, S. (2004). Separating heterogeneity from uncertainty an aiyagari-laitner economy. In *Goldwater Conference on Labor Markets in Arizona*, volume 608.

Cunha, F., Heckman, J., and Navarro, S. (2005). Separating uncertainty from heterogeneity in life cycle earnings. *oxford Economic papers*, 57(2):191–261.

Cunha, F. and Heckman, J. J. (2007). Identifying and estimating the distributions of ex post and ex ante returns to schooling. *Labour Economics*, 14(6):870–893.

Delavande, A. and Zafar, B. (2019). University choice: The role of expected earnings, nonpecuniary outcomes, and financial constraints. *Journal of Political Economy*, 127(5):2343–2393.

Duflo, E., Dupas, P., and Kremer, M. (2021). The impact of free secondary education: Experimental evidence from ghana. National Bureau of Economic Research No. 28937.

Eisenhauer, P., Heckman, J. J., and Vytlacil, E. (2015). The generalized roy model and the cost-benefit analysis of social programs. *Journal of Political Economy*, 123(2):413–443.

Evdokimov, K. (2010). Identification and estimation of a nonparametric panel data model with unobserved heterogeneity. Working Paper, Princeton University.

Freyberger, J. (2018). Non-parametric panel data models with interactive fixed effects. *The Review of Economic Studies*, 85(3):1824–1851.

Gindling, T., Hasnain, Z., Newhouse, D., and Shi, R. (2020). Are public sector workers in developing countries overpaid? evidence from a new global dataset. *World Development*, 126:104737.

Girsberger, E. M. and Meango, R. (2022). The puzzle of educated unemployment in west africa. IZA Discussion Paper No. 15721.

Giustinelli, P. (2022). Expectations in education: Framework, elicitation, and evidence. *Available at SSRN 4318127*.

Giustinelli, P. (2023). Expectations in education. *Handbook of Economic Expectations*, pages 193–224.

Giustinelli, P. and Shapiro, M. D. (2023). Seate: Subjective ex ante treatment effect of health on retirement. Technical report. Unpublished manuscript.

Gong, Y., Stinebrickner, T., and Stinebrickner, R. (2020). Perceived and actual option values of college enrollment. *Journal of Applied Econometrics*, 35(7):940–959.

Gong, Y., Stinebrickner, T., Stinebrickner, R., and Yao, Y. (2022). The role of non-pecuniary considerations: Locations decisions of college graduates from low income backgrounds. Technical report, University of Western Ontario, Centre for Human Capital and Productivity (CHCP).

Heckman, J. J., Lochner, L. J., and Todd, P. E. (2006). Earnings functions, rates of return and treatment effects: The mincer equation and beyond. *Handbook of the Economics of Education*, 1:307–458.

Heckman, J. J. and Navarro, S. (2007). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics*, 136(2):341–396.

Hudomiet, P., Hurd, M. D., Parker, A. M., and Rohwedder, S. (2021). The effects of job characteristics on retirement. *Journal of pension economics & finance*, 20(3):357–373.

Hudomiet, P., Hurd, M. D., and Rohwedder, S. (2018). The causal effects of economic incentives, health and job characteristics on retirement: estimates based on subjective conditional probabilities. In *2018 Working Longer and Retirement Conference*.

Imbens, G. W. and Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512.

IOM (2023). National study of the labour market in côte d'ivoire. International Organization for Migration (IOM), Geneva.

Juster, F. T. (1966). Consumer buying intentions and purchase probability: An experiment in survey design. *Journal of the American Statistical Association*, 61(315):658–696.

Karr, A. F. (1993). *Probability*. Springer New York, New York, NY.

Kesternich, I., Heiss, F., McFadden, D., and Winter, J. (2013). Suit the action to the word, the word to the action: Hypothetical choices and real decisions in medicare part d. *Journal of Health Economics*, 32(6):1313–1324.

Kettlewell, N., Walker, M. J., and Yoo, H. I. (2024). Alternative models of preference heterogeneity for elicited choice probabilities. *Available at SSRN*.

Koşar, G. and O'Dea, C. (2023). Expectations data in structural microeconomic models. In *Handbook of Economic Expectations*, pages 647–675. Elsevier.

Koşar, G., Ransom, T., and Van der Klaauw, W. (2022). Understanding migration aversion using elicited counterfactual choice probabilities. *Journal of Econometrics*, 231(1):123–147.

Lee, S. Y. T., Shin, Y., and Lee, D. (2015). The option value of human capital: Higher education and wage inequality. Technical report, National Bureau of Economic Research.

Maestas, N., Mullen, K. J., Powell, D., Von Wachter, T., and Wenger, J. B. (2023). The value of working conditions in the united states and implications for the structure of wages. *American Economic Review*, 113(7):2007–2047.

Mangal, K. (2022). The long-run costs of highly competitive exams for government jobs. *Unpublished manuscript*.

Mangal, K. (2024). How Much Are Government Jobs in Developing Countries Worth? *The World Bank Economic Review*, page lhad047.

Manski, C. F. (1999). Analysis of choice expectations in incomplete scenarios. *Journal of Risk and Uncertainty*, 19:49–66.

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

Manski, C. F. and Molinari, F. (2010). Rounding probabilistic expectations in surveys. *Journal of Business & Economic Statistics*, 28(2):219–231.

Mas, A. and Pallais, A. (2017). Valuing alternative work arrangements. *American Economic Review*, 107(12):3722–3759.

Matzkin, R. L. (2008). Identification in nonparametric simultaneous equations models. *Econometrica*, 76(5):945–978.

Méango, R. and Poinas, F. (2023). The option value of overstaying. CESifo Working Paper 10536.

Melly, B. and Santangelo, G. (2015). The changes-in-changes model with covariates. *Universität Bern, Bern*.

Murphy, K. M., Shleifer, A., and Vishny, R. W. (1991). The allocation of talent: Implications for growth. *The quarterly journal of economics*, 106(2):503–530.

Sasaki, Y. (2015). Heterogeneity and selection in dynamic panel data. *Journal of Econometrics*, 188(1):236–249.

Stange, K. M. (2012). An empirical investigation of the option value of college enrollment. *American Economic Journal: Applied Economics*, 4(1):49–84.

Trachter, N. (2015). Stepping stone and option value in a model of postsecondary education. *Quantitative Economics*, 6(1):223–256.

Wiswall, M. and Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, 82(2):791–824.

Wiswall, M. and Zafar, B. (2018). Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, 133(1):457–507.

Appendix

Appendix A. A comparison with the revealed preference approach to the identification of ex ante returns

In a series of influential papers, Carneiro et al. (2003); Cunha et al. (2004, 2005) make a distinction between *ex ante* and *ex post* returns, where their focus is mainly on *earnings returns*, the increase in the agent's lifetime earnings from a human capital investment (college education) (see also a review in Cunha and Heckman, 2007). They emphasise the difference between components of earnings variability that are forecastable and acted upon at the time students decide to go to college (heterogeneity) and components that are unforecastable. More specifically, the analyst observes the investment decision at time $t = 0$, say D, and the associated ex post stream of income of each individual, say Y_t at time $t > 0$, but observes neither the potential outcomes $Y_t(0), Y_t(1)$, nor the direct costs to college education C, or the full information set $\mathcal{I}_0$ at the time of decision. The agent is assumed to take the investment if the expected gains from schooling are greater than or equal to the expected costs $\mathbb{E}(Y(1) - Y(0) - C|\mathcal{I}_0)$, where $Y(d)$ is the lifetime earnings with education d. This literature decomposes the variability of the ex post returns $Y(1) - Y(0) - C$ into its component that relate to $\mathbb{E}(Y(1) - Y(0) - C|\mathcal{I}_0)$, the (forecastable) heterogeneity, and $Y(1) - Y(0) - C - \mathbb{E}(Y(1) - Y(0) - C|\mathcal{I}_0)$, the unforecastable component ('luck').

In the revealed preference approach, identifying predictable components that enter the information set $\mathcal{I}_0$ from unpredictable components relies critically on assumptions about the market structure facing agents and their preferences. For example, Carneiro et al. (2003) considers an environment of complete autarky. Cunha et al. (2005) assumes complete markets. The main identification condition is that the choice of education D is not associated with the forecast error $Y(1) - Y(0) - C - \mathbb{E}(Y(1) - Y(0) - C|\mathcal{I}_0)$. Operationalising this condition requires an estimate of the joint distribution of potential outcome $F_{Y(1)-Y(0)}$, which the analyst can obtain by embedding a factor model structure in the generalised Roy model (Carneiro et al., 2003).

Similar to the aforementioned literature, the interest in this paper is in the distribution of ex ante returns, pecuniary and nonpecuniary, $\mathbb{E}(Y(1) - Y(0) - C|\mathcal{I}_0)$ in their notations. Stated preference data are different in nature as they pertain to ex ante perceptions of agents. To be clear, for the objects we identify, it is not necessary for the analyst to observe the respondents perception of potential outcome *at the time of decision*, $F_{Y(1),Y(0)|\mathcal{I}_0}$. In the empirical application, we do observe the perceived distribution of earnings *at the time of elicitation*, which is close to the time of decision. What is instrumental within the choice experiment is that the analyst observes stated counterfactual choices for different, exogenous, combinations of these outcomes, say $D(y_1, y_0)$. Hence, this bypasses the problem of estimating $F_{Y(1)-Y(0)}$, which is controlled for by the analyst. When agents are allowed to express uncertainty about their decision, the analyst observes the probability that agents make the investment, say $\Pr(D(y_1, y_0) = 1)$. The results in the present paper is that the latter contains information rich enough to identify the distribution of ex ante returns for counterfactual realisations of earnings (y_1, y_0), that is, $\mathbb{E}(y_1 - y_0 - C|\mathcal{I}_0, Y(1) = y_1, Y(0) = y_0)$ in Cunha and Heckman (2007)'s notations. Crucially, identification can be achieved without relying on assumptions about the market structure facing agents and their preferences, or on a particular model structure.

If in addition to stated preference data, information about realised outcomes Y_t and actual choices D are available, the analyst can recover the ex post returns $Y(1) - Y(0) - C$ with usual instrumental variables methodologies (e.g. Eisenhauer et al., 2015). This would allow decomposing the variability of returns in the spirit of Cunha et al. (2005). This case is left for future research.

Stated preference data permit an additional layer of complexity. Symmetrically to Cunha and Heckman (2007)'s distinction between the time of decision (where *outcomes are anticipated*) and the time of outcome realisation, the analyst can distinguish the time of elicitation (where the *decision is anticipated*) from the time of decision. At the time of elicitation, say $t = -1$, the agent is not certain of their returns, and holds beliefs about the distribution of $\mathbb{E}(Y(1) - Y(0) - C|\mathcal{I}_0)$ given their information set $\mathcal{I}_{-1}$. This is the resolvable uncertainty, the uncertainty that agents expect to be resolved at the time of decision. The results in this paper is that choice experiments eliciting the probability to make an investment are rich enough to measure the magnitude of the resolvable uncertainty. Quantifying the resolvable uncertainty is essential to understand the option value for waiting to collect more information rather than committing to one choice ex ante (Heckman and Navarro, 2007; Stange, 2012; Gong et al., 2020; Méango and Poinas, 2023).

Appendix B. Additional Results

This section presents additional results of characterisation and identification. More specifically, Section B.1 introduces characterisation results for additional parameters of interest: mean ex ante returns, quantile effects, mean effects, and IQR.

B.1. **Characterisation of parameter of interests.** The identification of ex ante returns and their distribution in the population is of interest in and of itself. The ex ante returns are also fundamental to characterise further parameters that describe individual preferences. Complementing the definition of quantiles, it is straightforward to define parameters routinely encountered in the literature:

(1) Mean returns: $\mu_S(x, \eta) := \mathbb{E}\left[S(x, \eta^*)|\eta\right]$;
(2) Quantile effects: $QE_{j,h}(\tau, x, \eta) := Q_S(\tau; x + h\boldsymbol{e}_j, \eta) - Q_S(\tau; x, \eta)$, where $h \in \mathbb{R}$ and $\boldsymbol{e}_j$ is a vector of zeros except for the j-th component, which equals one.
(3) Mean effects: $ME_{j,h}(x, \eta) := \mu_S(x + h\boldsymbol{e}_j, \eta) - \mu_S(x, \eta)$.

Quantile and mean effects measure the contribution of a specific choice characteristic to agents' utility; in other words, the willingness to pay for these attributes. The quantile effect measures the shift in quantile that is attributable to a change in the j-th component of the vector of choice characteristics. In our empirical application, it would measure, for example, how the τ-quantile of returns changes with an increase in the chance of becoming regularised. The mean effect has the same interpretation.

The spread of the conditional distribution $F_{S,i}$ provides a useful tool to quantify the importance of (resolvable) uncertainty. One measure of this spread is the interquantile range (IQR). For any $0 < \tau_1 < \tau_2 < 1$, the interquantile range for an individual with beliefs and characteristics (X, η) is defined by:

$$IQR(X, \eta; \tau_1, \tau_2) = Q_S(\tau_2; X, \eta) - Q_S(\tau_1; X, \eta) \tag{16}$$

Start with the following remark:

Remark 1. *Given knowledge of the distribution of $F_{S,i}(s; x)$ (Lemma 1), one can derive the first moment, say $\mu_S(x, \eta)$, and conditional quantiles, say $A^\tau(x, a)$, as in Chernozhukov et al. (2020) (cf. also Karr, 1993, pp. 113-114). Let $\mathcal{S}$ denote the support of S, and assume that μ_S is bounded. Denote by $t(s, x) := (y_0 + s, y_1, x)$. Then:*

$$\begin{aligned} \mu_S(x, \eta) &= \int_{\mathcal{S}^+} [m(t(s, x), \eta)]\, ds - \int_{\mathcal{S}^-} [1 - m(t(s, x), \eta)]\, ds, \\ A^\tau(x, \eta) &= \int_{\mathcal{S}} 1\{s \geq 0\} - 1\{[1 - m(t(s, x), \eta)] \geq \tau\}\, ds \end{aligned} \tag{17}$$

Theorem 2 provides a characterisation for the remaining distributions of interest.

Theorem 2. *Recall that $\mathcal{S}$ denotes the support of S. Under the conditions of Theorem 1 and for μ_S bounded, the population distributions of interest are characterised as follows: Define*

$$\begin{aligned} &A^+(x, \eta) = \textstyle\int_{\mathcal{S}^+} [m(t(s, x), \eta)]\, ds, \\ &A^-(x, \eta) = \textstyle\int_{\mathcal{S}^-} [1 - m(t(s, x), \eta)]\, ds, \\ &\mu_S(x, \eta) = A^+(x, \eta) - A^-(x, \eta) \\ &A^\tau(x, \eta) = \textstyle\int_{\mathcal{S}} 1\{s \geq 0\} - 1\{[1 - m(t(s, x), \eta)] \geq \tau\}\, ds, \end{aligned}$$

where $\mathcal{S}^+ = \mathbb{R}^+ \cap \mathcal{S}$ and $\mathcal{S}^- = \mathbb{R}^- \cap \mathcal{S}$. Then, we have:

$$\Pr(\mu_S(X, \eta) \leq y) = \int_{\mathcal{X}} \int_{\mathcal{H}} 1\{\mu_S(x, n) \leq y\}\, dF_{X,\eta}(x, n), \tag{18}$$

$$\Pr(QE_{j,h}(\tau; X, \eta) \leq y) = \int_{\mathcal{X}} \int_{\mathcal{H}} 1\{A^\tau(x + h\boldsymbol{e}_j, n) - A^\tau(x, n) \leq y\}\, dF_{X,\eta}(x, n), \tag{19}$$

$$\Pr(ME_{j,h}(X, \eta) \leq y) = \int_{\mathcal{X}} \int_{\mathcal{H}} 1\{\mu_S(x + h\boldsymbol{e}_j, n) - \mu_S(x, n) \leq y\}\, dF_{X,\eta}(x, n), \tag{20}$$

$$\Pr(IQR(X, \eta; \tau_1, \tau_2) \leq y) = \int_{\mathcal{X}} \int_{\mathcal{H}} 1\{A^{\tau_2}(x, n) - A^{\tau_1}(x, n) \leq y\}\, dF_{X,\eta}(x, n). \tag{21}$$

As with Theorem 1, Theorem 2 implies that identification of the function $m(x, \eta)$ and the conditional distribution of η is sufficient for identification of the distributions of interest.

Appendix C. Proofs

This section collects the proofs from the main text. Few notations: for any variables X, Y, we denote by $F_X(x)$ and $F_Y(y)$ the marginal cumulative distribution functions, $F_{Y|X}(y|x)$, the conditional distribution function of Y given X, and $Q_{Y|X}(y|x)$ the conditional quantile function of Y given X, which is the generalised inverse of $F_{Y|X}$. For any variable X, denote by X' an i.i.d. copy of X.

C.1. **Proof of Lemma 1.** By equation (2) and Assumption 2, for any pair $(x, \eta^*) \in \mathcal{X} \times \mathcal{H}$ and $s \in \mathbb{R}$: $S(y_0 + S(x, \eta^*), y_1, z, \eta^*) = 0$ if and only if $S(x, \eta^*) = S^{-1}(0, y_1, z, \eta^*) - y_0$, where S^{-1} is the inverse of S with respect to its first argument. It follows that: $\Big\{\eta^* : S(x, \eta^*) \leq s\Big\} = \Big\{\eta^* : S^{-1}(0, y_1, z, \eta^*) - y_0 \leq s\Big\} = \Big\{\eta^* : S(y_0 + s, y_1, z, \eta^*) \geq 0\Big\}$. Taking the conditional expectation, it follows that:

$$\begin{aligned} F_{S,i}(s; x) &= \Pr(S(x, \eta_i^*) \leq s|\eta_i), \text{ by definition.} \\ &= \Pr(S(y_0 + s, y_1, z, \eta_i^*) \geq 0|\eta_i) \\ &= 1 - m(y_0 + s, y_1, z, \eta_i), \text{ by definition of } m(.). \end{aligned}$$

C.2. **Proof of Proposition 1.** For any $a \in (0, 1)$ and any $x \in \mathcal{X}$

$$\begin{aligned} q(x, a) &= Q_{m(x,\eta)}(a) \text{ by definition of the QTR,} \\ &= Q_{m(X,\eta)|X}(a|x) \text{ by independence as from Assumption 1,} \\ &= Q_{h(X,\eta,0.5)|X}(a|x) \text{ by Assumption 1,} \\ &= Q_{h(X,\eta,\epsilon)|X,\epsilon}(a|x, 0.5) \text{ since, by the normalisation, } \epsilon \perp\!\!\!\perp (X, \eta) \\ &= Q_{P|X,\epsilon}(a|x, 0.5) \end{aligned}$$

C.3. **Proof of Proposition 2.** This section is concerned with showing identification for both N and T being infinite. The identification is constructive and suggests an estimation procedure that performs reasonably well even with small T.

Claim 1: The sequence $\{V_k\}_{k=1}^{\infty}$ converges (a.s.) to a fixed point V_∞ such that $V_\infty \perp\!\!\!\perp X$ and $V_\infty \perp\!\!\!\perp \eta$. Another way of understanding the construction of the sequence $\{V_k\}_{k=1}^{\infty}$ is that it alternates orthogonal projections. Indeed, let V_k' be an i.i.d. copy of V_k. V_{2k+1} is the orthogonal projection of $I\{V_{2k}' \leq V_{2k}\}$ on the space of functions of X. $V_{2(k+1)}$ is the orthogonal projection of $I\{V_{2k+1}' \leq V_{2k+1}\}$ on the space of functions of η. One useful of the property of the conditional expectation is that it is the solution to the problem:

$$\min_g \mathbb{E}\left((I\{V_k' \leq p\} - g(X))^2\right), \text{ for each } p \text{ and each } kge0.$$

Thus, we have:

$$\mathbb{E}\left((I\{V_0' \leq p\} - F_{V_0}(p))^2\right) \geq \mathbb{E}\left(\left(I\{V_0' \leq p\} - F_{V_0|X}(V_0|x)\right)^2\right), \text{ for all } p, x,$$

with equality if and only if $V_0 \perp\!\!\!\perp X$. Since the above inequality holds for any $(p, x) \in [0,1] \times \mathcal{X}$, it follows that:

$$\mathbb{E}\left((I\{V_0' \leq V_0\} - F_{V_0}(V_0))^2\right) \geq \mathbb{E}\left(\left(I\{V_0' \leq V_0\} - F_{V_0|X}(V_0|X)\right)^2\right) = \mathbb{E}\left((I\{V_0' \leq V_0\} - V_1)^2\right) \tag{22}$$

Note now that, since $p \mapsto F_{P|X}(p|x)$ is strictly increasing for all x, $I\{V_0' \leq V_0\} = I\{F_P(P') \leq F_P(P)\} = I\{P' \leq P\}$ a.s.. Furthermore, $I\{V_1' \leq V_1\} = I\{F_{V_0|X}(V_0'|X) \leq F_{V_0|X}(V_0|X)\} = I\{V_0' \leq V_0\} = I\{P' \leq P\}$ a.s.. Similarly, one can show that, for any $k > 1$, $I\{V_k' \leq V_k\} = I\{P' \leq P\}$, a.s..

Note also that $F_{V_0}(V_0) = V_0$, because $V_0 = F_P(P) \sim \mathcal{U}[0,1]$. Hence, equation (22) simplifies to:

$$\mathbb{E}\left((I\{P' \leq P\} - V_0)^2\right) \geq \mathbb{E}\left((I\{P' \leq P\} - V_1)^2\right) \tag{23}$$

with equality if and only if $V_0 \perp\!\!\!\perp X$.

With a similar argument, we can show, *mutatis mutandis*, that since $F_{V_1|\eta}$ minimises the problem:

$$\min_g \mathbb{E}\left((I\{V_1' \leq p\} - g(\eta))^2\right), \text{ for all } p.$$

we have

$$\mathbb{E}\left((I\{P' \leq P\} - V_1)^2\right) \geq \mathbb{E}\left((I\{P' \leq P\} - V_2)^2\right) \tag{24}$$

with equality if and only if $V_1 \perp\!\!\!\perp \eta$. Note that by construction, $V_1 \perp\!\!\!\perp X$.

If V_1 is independent of η, we have reached a fixed point. Indeed, $V_2 = F_{V_1|\eta}(V_1|\eta) = F_{V_1}(V_1) = V_1$, a.s., where the first equality follows by independence and the second follows because V_1 has a uniform distribution. However, since V_2 is by construction independent of X, $V_3 = F_{V_2|X}(V_2|X) = F_{V_2}(V_2) = V_2 = V_1$, a.s.. Similarly $V_k = V_{k-1} = \ldots = V_1$ a.s. for any $k > 1$.

Reproducing the above argument shows that the sequence $\{V_k\}_{k=1}^{\infty}$ is such that, for any k:

$$\mathbb{E}\left((I\{P' \leq P\} - V_k)^2\right) \geq \mathbb{E}\left((I\{P' \leq P\} - V_{k+1})^2\right) \tag{25}$$

with equality if and only if $V_{2k+1} \perp\!\!\!\perp \eta$ or $V_{2k} \perp\!\!\!\perp X$, for all $k > 1$. If either V_{2k+1} is independent of η or V_{2k} is independent of X, we have reached a fixed point.

Otherwise, the implied sequence $C_n = \mathbb{E}\left((I\{P' \leq P\} - V_k)^2\right)$ is a bounded decreasing sequence on a compact ($[0,1]$). Thus, it must converge to a fixed-point within the sequence, say C_∞, which is the minimum of the sequence. The constructed variable V_∞ that realises C_∞ must be a fixed point of the sequence $\{V_k\}_{k=1}^{\infty}$ and

$$V_\infty = F_{V_\infty|X}(V_\infty|X) = F_{V_\infty|\eta}(V_\infty|\eta), a.s..$$

Otherwise, suppose that $V_\infty \neq F_{V_\infty|X}(V_\infty|X)$, then:

$$\mathbb{E}\left(\left(I\{P' \leq P\} - F_{V_\infty|X}(V_\infty|X)\right)^2\right) < \mathbb{E}\left((I\{P' \leq P\} - V_\infty)^2\right) = C_\infty,$$

which is a contradiction. The same holds if $V_\infty \neq F_{V_\infty|\eta}(V_\infty|\eta)$.

Claim 2: $V_\infty = \epsilon$. To show this result, it is useful to consider V_k as a function of (X, η, ϵ). To see this, first, note that, since $V_0 = F_P(P)$, $V_1 = F_{V_0|X}(V_0|X) = F_{P|X}(P|X)$, by monotonicity of F_P. Now, consider $F_{P|X}(p|x)$, the conditional distribution of P given X. Note that $X \perp\!\!\!\perp \eta$, and $\epsilon \perp\!\!\!\perp (X, \eta)$ both imply that $X \perp\!\!\!\perp (\epsilon, \eta)$.

$$
\begin{aligned}
F_{P|X}(p|x) &= \Pr\left(P \le p|X = x\right) \\
&= \Pr\left(h(X, \eta, \epsilon) \le p|X = x\right) \\
&= \Pr\left(h(x, \eta, \epsilon) \le p\right), \text{ since } X \perp\!\!\!\perp (\epsilon, \eta), \\
&= \int I\left\{h(x, n, e) \le p\right\} dF_{\eta,\epsilon}(n, e).
\end{aligned}
$$

Applying the conditional distribution to the pair (P, X) gives the following expression for V_1:

$$
\begin{aligned}
V_1 &= \int I\left\{h(X, n, e) \le P\right\} dF_{\eta,\epsilon}(n, e) \\
&= \int I\left\{h(X, n, e) \le h(X, \eta, \epsilon)\right\} dF_{\eta,\epsilon}(n, e) \\
&:= \varphi_1(X, \eta, \epsilon).
\end{aligned}
$$

It is important to note that the mapping $e \mapsto \varphi_1(x, n, \epsilon)$ inherits a monotonicity property from the monotonicity of the mapping $e \mapsto h(x, n, \epsilon)$, for all (x, n).

A similar argument leads to:

$$
\begin{aligned}
V_{2k} &= \int I\left\{\varphi_{2k-1}(x, \eta, e) \le \varphi_{2k-1}(X, \eta, \epsilon)\right\} dF_{X,\epsilon}(x, e) := \varphi_{2k}(X, \eta, \epsilon), \text{ and} \\
V_{2k+1} &= \int I\left\{\varphi_{2k}(X, n, e) \le \varphi_{2k}(X, \eta, \epsilon)\right\} dF_{\eta,\epsilon}(n, e) := \varphi_{2k+1}(X, \eta, \epsilon), \text{ for all } k \ge 1,
\end{aligned}
$$

where the mapping $e \mapsto \varphi_{2k+1}(x, n, e)$ is strictly monotone.

Let $\varphi_\infty(X, \eta, \epsilon) = V_\infty$ be the limit of the sequence $\{\varphi_k(X, \eta, \epsilon)\}_{k=1}^{\infty}$. We have established that:

(i) $V_\infty \perp\!\!\!\perp \eta$, i.e, $F_{V_\infty|\eta}(p|\eta) = F_{V_\infty}(p) = p$; and
(ii) $V_\infty \perp\!\!\!\perp X$, i.e, $F_{V_\infty|X}(p|X) = F_{V_\infty}(p) = p$.

Hence, by (i), for any $n \in \mathcal{H}$,

$$
\begin{aligned}
\Pr\left(\varphi_\infty\left(X, \eta, \epsilon\right) \le p|\eta = n\right) &=_{(1)} \Pr\left(\varphi_\infty\left(X, n, \epsilon\right) \le p\right) \\
&=_{(2)} \Pr\left(\epsilon \le \varphi_\infty^{-1}\left(X, n, p\right)\right) \\
&=_{(3)} \int_{\mathcal{X}} \varphi_\infty^{-1}\left(X, n, p\right) dF_X(x) = p. \qquad (26)
\end{aligned}
$$

where equality (1) follows by independence of η and (X, ϵ). In (2), $\varphi_\infty^{-1}(X, n, p)$ denotes the inverse of the mapping $e \mapsto \varphi_\infty(x, n, e)$ and is the result of the strict monotonicity of this mapping for all x, n. Finally, equality (3) follows from the fact that ϵ is uniformly distributed and independent of X. For equation (26) to be satisfied for any $n \in \mathcal{H}$, it must be that there exists function $\overline{\varphi_\infty^{-1}}(X, t)$, such that $\varphi_\infty^{-1}(X, n, p) = \overline{\varphi_\infty^{-1}}(X, p) > 0$, a.s..

However, by (ii), it must be that applying the same argument, for any x in $\mathcal{X}$:

$$\int_{\mathcal{X}} \varphi_\infty^{-1}(x,n,p)\, dF_\eta(n) = \int_{\mathcal{H}} \overline{\varphi_\infty^{-1}}(x,p) dF_\eta(n) = \overline{\varphi_\infty^{-1}}(x,p) = p$$

Hence, for any (x,n), $\varphi_\infty^{-1}(x,n,p) = p$. We conclude that: $V_\infty = \varphi_\infty(X,\eta,\epsilon) = \epsilon$.

Appendix D. Additional Details of Simulations

This section presents additional simulation details and results.

We generalise the estimation procedure from Wiswall and Zafar (2018) to estimate distribution of quantiles. The LAD model is the workhorse model in the stated preference literature:

$$\log\left(\frac{P_{it}}{1-P_{it}}\right) = r(Y_{0it}, Y_{1it}) + \epsilon_{it}, \text{ with } Q_{\epsilon_{it}|X_{it},\eta_i}(0.5|X_{it},\eta_i) = 0. \tag{27}$$

where $r(Y_{0it}, Y_{1it})$ is a polynomial that includes levels and interactions of choice attributes. The usual assumption is that: $r(Y_{0it}, Y_{1it}) = c_i + b_i(Y_{0it} - Y_{1it})$. Assuming that the resolvable uncertainty has a logistic distribution, the cdf $F_{S,i}$ is given by:

$$\Pr\left(S_{it} \le s | Y_{0it}, Y_{i1t}, \eta_i\right) = \left[1 + \exp\left(r(Y_{0it} + s, Y_{1it})\right)\right]^{-1}. \tag{28}$$

The estimation of r assumes a quadratic form: $r(Y_{0it}, Y_{1it}) = c_i + b_{0i}Y_{0it} + b_{1i}Y_{1it} + b_{01i}Y_{0it} \times Y_{1it} + b_{3i}Y_{0it}^2 + b_{4i}Y_{1it}^2$. The quadratic terms are considered only for $T = 20$. Estimation of $(\hat{c}_i, \hat{b}_{0i}, \hat{b}_{1i}, \hat{b}_{01i}, \hat{b}_{3i}, \hat{b}_{4i})$ and thus $\hat{F}_{S,i}$ is conducted separately for each individual using a Least-Absolute Deviation estimator on the (winsorised) log-odds of P_{it}. An estimator for $F_Q(s;\tau)$ is obtained by calculation the proportion of the sample such that: $\hat{F}_{S,i} \le 1 - \tau$.

One alternative that generalises Blass et al. (2010) (BLM2010) is to assume a random effect model. Let $r(Y_{0it}, Y_{1it}) = c_i + b_i(Y_{0it} - Y_{1it})$ and $c_i \sim N(0, \sigma_1)$ and $b_i \sim N(0, \sigma_2)$. This performs uniformly worse than the other strategies, as evidenced by Figure F.1.

For the first stage of the 2S-KR/DR, the DR uses logit regressions for $p \in \mathcal{P}_{\text{grid}} = \{0, 0.05, 0.1, \ldots, 0.95\}$ to estimate $\hat{F}_{P|X}(p|x)$ for each cross section and predict $\hat{F}_{V_k|X}(P_{it}|X_{it})$. The estimation is conducted with the routine 'fitglm' in Matlab. When V_{kit} takes values outside the estimation set $\mathcal{P}_{\text{grid}}$, the estimate is obtained by interpolation. The function $r(.)$ in equation (11) is defined as the saturated interaction between the squared-values of (Y_{0it}, Y_{1it}).

For the second stage of the 2S-KR/DR, the estimated $\hat{F}_{P|X}(p|X_{it}, \hat{\epsilon}_{it}).p = 0.20, 0.25, \ldots, 0.80$, is performed for each cross-section and the estimates are averaged out across cross-sections. The KR estimation uses the matlab routine 'ksrmv.m', authored by Yi Cao. The code implements the Nadaraya-Watson kernel regression using a Gaussian kernel. The bandwidth is the optimal bandwidth suggested by (Bowman and Azzalini, 1997, p. 31). The DR estimation uses the routine 'fitglm' in Matlab. The function $r(.)$ in equation (13) is defined as the saturated interactions between the values of (Y_{0it}, Y_{1it}) and the estimated value $\hat{\epsilon}_{it}$ added as a separable variable. The tables report the result for $p = 0.25, 0.50, 0.75$. The results for the remaining quantiles are very similar, and available upon request.

$\beta_i = 1$ for all i													
		N=500						N=1,000					
		RISB			Std. deviation			RISB			Std. deviation		
	τ	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
WZ2018	$T = 5$	0.105	0.069	0.040	0.323	0.592	0.528	0.103	0.066	0.036	0.306	0.583	0.538
(1)	$T = 10$	0.032	0.027	0.035	0.368	0.714	0.617	0.031	0.021	0.031	0.366	0.712	0.599
	$T = 20$	0.032	0.031	0.027	0.379	0.690	0.591	0.031	0.026	0.020	0.365	0.692	0.613
2S-KR/DR	$T = 5$	0.035	0.081	0.119	0.432	0.817	0.802	0.030	0.080	0.114	0.448	0.854	0.799
(2)	$T = 10$	0.029	0.076	0.115	0.444	0.790	0.766	0.027	0.077	0.114	0.481	0.870	0.828
	$T = 20$	0.022	0.077	0.115	0.439	0.849	0.787	0.023	0.076	0.115	0.437	0.852	0.771
Ratio	$T = 5$	2.969	0.848	0.335	0.746	0.724	0.659	3.394	0.831	0.310	0.682	0.683	0.673
(1)/(2)	$T = 10$	1.096	0.353	0.304	0.830	0.905	0.806	1.143	0.278	0.272	0.760	0.819	0.724
	$T = 20$	1.440	0.401	0.236	0.864	0.813	0.751	1.315	0.337	0.174	0.836	0.813	0.794

Low measurement error ($w = 0.1$), no rounding													
		N=500						N=1,000					
		RISB			Std. deviation			RISB			Std. deviation		
	τ	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
WZ2018	$T = 5$	0.223	0.213	0.202	0.422	0.612	0.488	0.222	0.213	0.200	0.428	0.614	0.493
(1)	$T = 10$	0.258	0.153	0.133	0.446	0.781	0.635	0.257	0.150	0.132	0.454	0.793	0.663
	$T = 20$	0.148	0.067	0.098	0.707	0.929	0.718	0.144	0.068	0.102	0.700	0.938	0.739
2S-KR/DR	$T = 5$	0.078	0.079	0.098	0.897	1.098	0.891	0.080	0.072	0.074	0.851	1.072	0.924
(2)	$T = 10$	0.068	0.053	0.073	0.894	1.048	0.907	0.079	0.051	0.068	0.907	1.078	0.964
	$T = 20$	0.074	0.049	0.073	0.899	1.067	0.953	0.083	0.050	0.068	0.868	1.073	0.932
Ratio	$T = 5$	2.865	2.705	2.058	0.470	0.557	0.548	2.788	2.967	2.710	0.503	0.573	0.534
(1)/(2)	$T = 10$	3.800	2.886	1.812	0.500	0.746	0.700	3.243	2.947	1.928	0.501	0.736	0.688
	$T = 20$	1.992	1.349	1.341	0.787	0.870	0.754	1.736	1.355	1.506	0.806	0.874	0.793

Moderate measurement error ($w = 0.5$), no rounding													
		N=500						N=1,000					
		RISB			Std. deviation			RISB			Std. deviation		
	τ	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
WZ2018	$T = 5$	0.234	0.206	0.185	0.394	0.609	0.510	0.234	0.208	0.188	0.378	0.601	0.502
(1)	$T = 10$	0.255	0.149	0.123	0.450	0.813	0.670	0.256	0.150	0.121	0.444	0.797	0.657
	$T = 20$	0.164	0.060	0.082	0.653	0.932	0.749	0.165	0.059	0.081	0.658	0.935	0.764
2S-KR/DR	$T = 5$	0.085	0.067	0.081	0.848	1.055	0.905	0.079	0.069	0.080	0.836	1.085	1.014
(2)	$T = 10$	0.075	0.050	0.064	0.856	1.080	0.978	0.082	0.055	0.060	0.856	1.083	0.980
	$T = 20$	0.079	0.049	0.056	0.826	1.067	0.984	0.086	0.051	0.058	0.870	1.095	0.991
Ratio	$T = 5$	2.765	3.083	2.291	0.464	0.577	0.564	2.975	3.033	2.358	0.453	0.554	0.495
(1)/(2)	$T = 10$	3.389	2.998	1.932	0.526	0.753	0.685	3.116	2.734	2.003	0.518	0.735	0.670
	$T = 20$	2.088	1.222	1.470	0.791	0.873	0.761	1.912	1.175	1.408	0.756	0.854	0.771

Table D.1. Additional specifications (1)

Note: The table summarises the results of the simulations for three cases described in the first line of each panel and: DR + KR: Distribution regression is employed in the first stage, and Kernel regression in the second stage. It compares these results to the generalisation of Wiswall and Zafar (2018) (WZ2018), as detailed in Appendix D. The last block of lines in each panel represents the ratio between the RISB and the standard deviation.

Large measurement error ($w = 1$), rounding to the nearest 5%

		N=500						N=1,000					
		RISB			Std. deviation			RISB			Std. deviation		
	τ	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
WZ2018	$T = 5$	0.257	0.231	0.192	0.320	0.543	0.505	0.255	0.228	0.190	0.317	0.548	0.501
(1)	$T = 10$	0.246	0.170	0.132	0.407	0.760	0.679	0.246	0.172	0.131	0.410	0.759	0.671
	$T = 20$	0.170	0.086	0.073	0.587	0.885	0.757	0.169	0.085	0.073	0.589	0.892	0.763
2S-KR/DR	$T = 5$	0.100	0.061	0.069	0.759	1.017	0.886	0.095	0.050	0.054	0.732	1.025	0.878
(2)	$T = 10$	0.087	0.040	0.050	0.739	1.047	0.930	0.076	0.035	0.048	0.751	1.049	0.947
	$T = 20$	0.071	0.031	0.056	0.781	1.041	0.899	0.065	0.029	0.049	0.773	1.054	0.933
Ratio	$T = 5$	2.567	3.785	2.780	0.421	0.534	0.569	2.676	4.526	3.507	0.433	0.535	0.571
(1)/(2)	$T = 10$	2.833	4.233	2.640	0.551	0.726	0.730	3.243	4.903	2.726	0.546	0.723	0.708
	$T = 20$	2.394	2.798	1.307	0.752	0.850	0.842	2.604	2.890	1.487	0.761	0.846	0.817

Moderate measurement error ($w = 0.5$), rounding to the nearest 10%

		N=500						N=1,000					
		RISB			Std. deviation			RISB			Std. deviation		
	τ	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
WZ2018	$T = 5$	0.245	0.221	0.194	0.348	0.562	0.487	0.244	0.220	0.191	0.340	0.555	0.491
(1)	$T = 10$	0.255	0.171	0.136	0.423	0.764	0.674	0.256	0.170	0.137	0.419	0.756	0.658
	$T = 20$	0.156	0.074	0.071	0.651	0.907	0.767	0.159	0.074	0.072	0.633	0.908	0.764
2S-KR/DR	$T = 5$	0.072	0.059	0.068	0.891	1.013	0.883	0.064	0.059	0.080	0.936	1.035	0.982
(2)	$T = 10$	0.058	0.041	0.058	0.966	1.031	0.936	0.057	0.040	0.064	0.915	1.018	0.898
	$T = 20$	0.038	0.035	0.050	0.947	1.015	0.947	0.038	0.034	0.056	0.968	1.013	0.915
Ratio	$T = 5$	3.402	3.733	2.841	0.391	0.555	0.551	3.796	3.744	2.385	0.364	0.536	0.500
(1)/(2)	$T = 10$	4.371	4.165	2.332	0.438	0.741	0.720	4.528	4.285	2.131	0.458	0.743	0.733
	$T = 20$	4.125	2.111	1.427	0.687	0.893	0.810	4.169	2.180	1.283	0.654	0.896	0.835

Increasing variance of the measurement error ($w = 0.1t$)

		N=500						N=1,000					
		RISB			Std. deviation			RISB			Std. deviation		
	τ	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75	0.25	0.5	0.75
WZ2018	$T = 5$	0.237	0.215	0.194	0.372	0.589	0.501	0.237	0.214	0.194	0.362	0.586	0.487
(1)	$T = 10$	0.258	0.163	0.128	0.415	0.787	0.684	0.259	0.164	0.129	0.421	0.779	0.680
	$T = 20$	0.165	0.088	0.077	0.588	0.880	0.759	0.165	0.087	0.075	0.605	0.884	0.753
2S-KR/DR	$T = 5$	0.065	0.062	0.092	0.907	1.057	0.911	0.049	0.062	0.076	0.890	1.047	0.932
(2)	$T = 10$	0.052	0.043	0.062	0.845	1.056	0.914	0.045	0.036	0.064	0.837	1.044	0.920
	$T = 20$	0.071	0.032	0.058	0.769	1.004	0.840	0.066	0.027	0.055	0.757	1.043	0.906
Ratio	$T = 5$	3.673	3.481	2.098	0.410	0.558	0.550	4.866	3.464	2.558	0.406	0.559	0.522
(1)/(2)	$T = 10$	4.932	3.781	2.075	0.492	0.746	0.748	5.782	4.610	2.023	0.503	0.747	0.739
	$T = 20$	2.340	2.725	1.317	0.764	0.877	0.904	2.481	3.181	1.372	0.799	0.847	0.831

Table D.2. Additional specifications (2)

Note: The table summarises the results of the simulations for three cases described in the first line of each panel and: DR + KR: Distribution regression is employed in the first stage, and Kernel regression in the second stage. It compares these results to the generalisation of Wiswall and Zafar (2018) (WZ2018), as detailed in Appendix D. The last block of lines in each panel represents the ratio between the RISB and the standard deviation.

Appendix E. Further details about the survey

This section presents the instructions for the choice experiment. They were presented using a short video.

> **Choice experiment instructions.** Now we're going to show you a series of hypothetical scenarios. In these scenarios, you receive two job offers for a position that corresponds to your degree. One offer comes from the public sector, the other from the private sector. In these scenarios, these two offers differ in the starting salary they offer, the number of hours of work expected, the risk of losing the job, and the progression in position.
>
> For example, let's compare these two offers:
>
> **Offer 1:**
> Your employer: Public administration
> Starting salary 500,000
> Number of hours per week 40
> Risk of losing job within two years 2/100
> Number of people who become managers with salary +20%: 20/100
>
> **Offer 2:**
> Your employer: SME
> Starting salary 800,000
> Number of hours per week 60
> Risk of losing job within two years 10/100
> Number of people who become managers with salary +20%: 1/10
>
> The starting salary in the public sector offer is 500,000 FCFA per month for a job in public administration. It is normal to work 40 hours a week. 2 out of 100 people lose their job within two years. Unemployment lasts 6 months. 20 people out of 100 become managers in their administration or company after two years.
>
> The starting salary in the private sector is 800,000 FCFA per month for a job in an SME. It's normal to work 60 hours a week. 20 people out of 100 lose their job within two years. Unemployment lasts 6 months. 10 people out of 100 become managers in their administration or company after two years.
>
> We are going to ask you to consider several such scenarios. Considering all the factors that might affect your decision to enter the job market (e.g. family background, your parents' wishes, macroeconomic conditions), we'd like to know which offer you'd be most likely to choose. We'll ask

you how likely you would be to choose the [randomize= private/public] sector. Between different scenarios, the only variations are indicated by the scenario. For example, the level of responsibility or the number of days off in the public sector job remain the same. Only the stated characteristics change. The same goes for the private sector offer. You can consider that these offers are your only options, and that you would not receive another offer within two years if you turned down both offers.

Q1 - Q5. Consider the following two offers, what is the chance that you will choose the offer from the [randomize= private/public] sector.

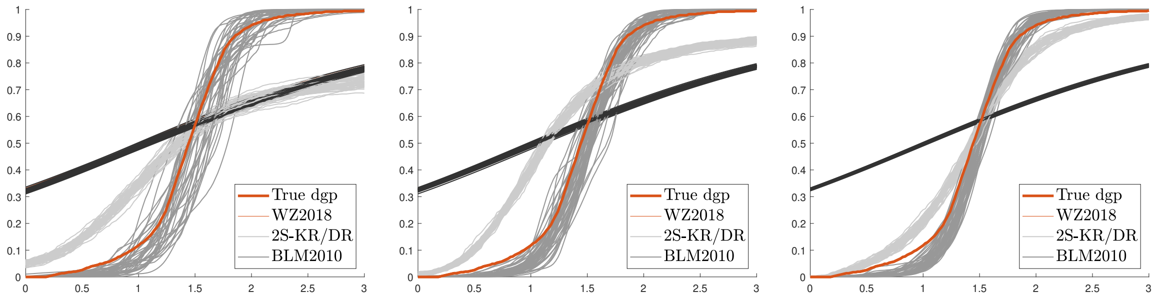
Table E.1 presents the support of choice attributes in the choice experiment. The values are chosen randomly for the second to the fifth scenario. In the first scenario, the employer are public administration and SME, the weekly hours worked 40, the chance of losing job is 5 percent, the chance of job job promotion, 10 percent, and the starting salary 750,000 CFAF (25 percent above the average perceived wage).

Public [Randomised between the values]	
Employer	Public administration, Public-sector company
Number of hours per week	35, 40
Chance of losing job within two years	2/100, 5/100, 10/100
Chance of job promotion within two years	5/100, 10/100, 20/100
Starting salary	from 300K to 1,000K with steps of 50K
Private [Randomised between the values]	
Employer	SME, Large-sized enterprise
Number of hours per week	40, 50, 60
Chance of losing job within two years	10/100, 20/100, 30/100
Chance of job promotion within two years	5/100, 10/100, 20/100
Starting salary	from 300K to 1,000K with steps of 50K

Table E.1. Support of choice attributes in the choice experiment

Note: Each row reports the possible values that a choice attribute can assume. Values are randomly assigned across individuals.

Appendix F. Additional figures



(a) $\tau = 0.5, T = 5$ (b) $\tau = 0.5, T = 10$ (c) $\tau = 0.5, T = 20$

FIGURE F.1. Simulated $F_Q(s; \tau, F_{\tilde{X}})$: BLM 2010 and WZ2018 compared to 2S-KR/DR with DR (first stage) + KR (second stage). BLM2010 performs poorly.

Specification: $\alpha_i \sim \mathcal{U}[0.25, 0.75]$, $\beta_i \in \{1, 2, \ldots, 5\}$, $\tau = 0.5$, $(\tilde{y}_0, \tilde{y}_1) = (0.5, 0.7)$, $T = 5, 10$, and 20 scenarios, $N = 1,000$, $N_{\text{sim}} = 50$.

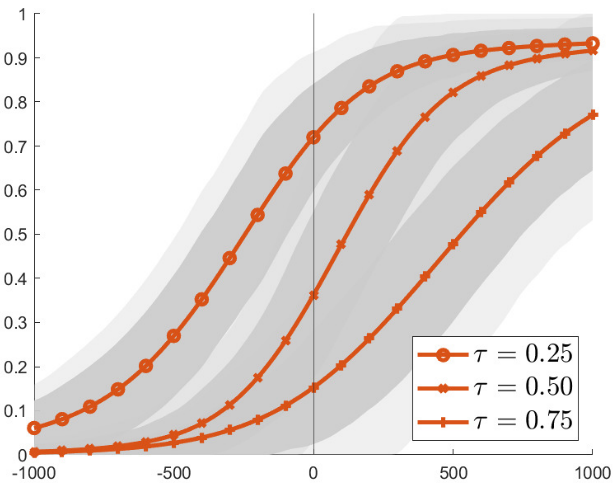


FIGURE F.2. Distribution of quantiles $\hat{F}_Q\left(.; \tau, F_{\tilde{X}}\right), \tau = 0.25, 0.50, 0.75$ for individual-specific perceptions

Note: The dark-grey area shows the 90 percent pointwise confidence interval. The light-grey area shows the 90 percent uniform confidence interval.

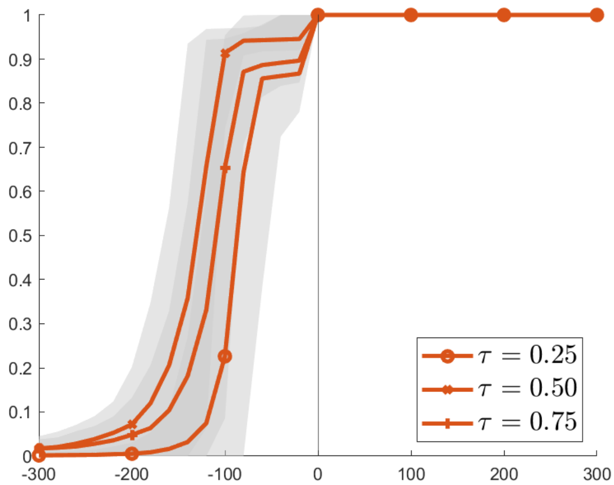


FIGURE F.3. Distribution of QTEs for $\tau = 0.25, 0.50, 0.75$ where the treatment is 'closing the working hours gap'

Note: The dark-grey area shows the 90 percent pointwise confidence interval.