Chapter 3: Observed and Unobserved Variables in Educational Treatment Effects

With Sergio Urzúa

3.1 Introduction

Decades of research have provided a detailed understanding of how unobservable factors shape heterogeneity in the returns to higher education (Belzil and Hansen, 2002; Belzil, Hansen and Liu, 2022; Carneiro, Heckman and Vytlacil, 2011; Carneiro, Hansen and Heckman, 2003; Cassagneau-Francis, 2022; Chen, 2008; Heckman, Humphries and Veramendi, 2018; Heckman, Stixrud and Urzua, 2006; Prada and Urzúa, 2017). Of course, observed characteristics also play a crucial role in elucidating heterogeneous responses, as they determine schooling choices and potential outcomes. Moreover, they have recently received increased attention in the treatment effects literature, as the functional specification of the outcome equations can be crucial to securing a causal interpretation (Blandhol et al., 2022). However, dissecting the interaction between unobservables and observable dimensions, which are often overlooked in existing literature, is necessary to untangle the factors determining the magnitudes of treatment effects.

This chapter explores the interplay between observed and unobserved variables and proposes a decomposition analysis to evaluate their relative importance in determining the treatment effects of postsecondary schooling decisions. We demonstrate that observable characteristics primarily influence both the level and slope—where the latter represents the diverse compensations—of marginal treatment effects for certain groups and margins of choice. Our focus on the marginal treatment effect is driven by its crucial role in shaping other treatment effects and effective policy interventions (Heckman and Vytlacil, 2007, 2001). By identifying individuals at the margin of indifference, we pinpoint those most likely to respond to incentives introduced by policies, as these individuals require minimal subsidies or incentives to alter their decisions.

Our framework employs a static Roy model with latent cognitive, socioemotional, and mechanical abilities. Previous literature has shown that all these are essential in determining college returns and decisions (Belzil, Hansen and Liu, 2022; Prada and Urzúa, 2017). In this model, individuals can choose between i) staying with a high school diploma, ii) obtaining a two-year degree, or iii) obtaining a four-year degree. We focus on the impact of these decisions among those indifferent between schooling alternatives, decomposing the contribution of unobserved and observed dimensions. Most of the literature studying postsecondary returns and considering heterogeneity in these latent abilities have focused on analyzing only men (e.g., Carneiro, Hansen and Heckman, 2003; Eisenhauer, Heckman and Mosso, 2015; Heckman, Humphries and Veramendi, 2018; Prada and Urzúa, 2017). We focus on highlighting the differences between women and men as we separately estimate our model for each group (see some exceptions: Belzil, Hansen and Liu, 2022; Heckman, Stixrud and Urzua, 2006).

Our identification strategy follows literature standards for Roy models. We rely on a measurement system that imposes structure and interpretation on the latent abilities and exclusion restrictions that affect choice equations but not earnings. Our primary source of information is the National Longitudinal Survey of Youth 1997 (NLSY97), which contains detailed information on earnings, postsecondary schooling trajectories, test scores, and background characteristics, all essential to estimate our model. This data source has been widely employed to study schooling decisions or returns to abilities (Arcidiacono et al., 2023; Belzil, Hansen and Liu, 2022; Deming, 2017; Luo, Hellerstein and Urzúa, 2023). The NLSY97 follows cohorts born between 1980 and 1984 that are slightly older than those analyzed by Mountjoy (2022), who also estimate marginal treatment effects of two-year and four-year college education, allowing us to have a benchmark to compare the magnitudes that we obtain. Limiting the number of schooling choices to three and imposing a static framework enable us to focus on analyzing all possible pairwise comparisons between schooling choices and decomposing marginal treatment effects (MTE) while having well-populated evaluation points of indifference. See the first chapter of this dissertation to focus instead on a dynamic version of this model, which has option values and examines more schooling decisions, including dropping out.

We contribute to three strands of the literature. First, we contribute to the literature that analyzes the different components of an MTE. We follow previous work focusing on decomposing the MTE in multinomial settings. Where more than two alternatives are available, the MTE can be expressed as a weighted average of the MTEs obtained from all the pair-wise comparisons between choices. Thus, we examine the MTE for individuals who are at different margins of indifference depending on the individual's next-best alternative (e.g., Heckman, Urzua and Vytlacil, 2006; Mountjoy, 2022). In addition, we propose a definition of the MTE that integrates over the distribution of observed characteristics for the in-sample compliers at each point of indifference.[1] This distinction allows us to decompose the MTE into the observables and each ability's contributions. The practice of decomposing parameters into observable and unobservable components

[1]Marginal treatment effects are usually defined as conditional functions of characteristics or abilities instead of integrating over them.

is a long-standing tradition in Labor Economics, dating back to Oaxaca-Blinder decompositions (Angrist, Pathak and Walters, 2013; Blinder, 1973; Oaxaca, 1973). These methods have been extended to treatment effect parameters, with several papers proposing methodologies or estimates for decomposing these parameters into systematic and idiosyncratic variations. (Ding, Feller and Miratrix, 2019; Djebbari and Smith, 2008; Heckman, Smith and Clements, 1997). Moreover, Blandhol et al. (2022) develop a decomposition to diagnose if the two-stage least squares estimate is level-dependent on outcomes and, hence, not rich in covariates. In contrast to previous work, we focus on analyzing MTEs and further decomposing unobservable components considering the role of latent abilities, an essential driver of heterogeneous schooling returns.

Second, we contribute to the articles that analyze the effect of higher education when more than two choices are available (e.g, Ferreyra, Galindo and Urzúa, 2022; Mountjoy, 2022). In particular, we can contribute to the literature that characterizes and analyzes the next-best alternatives. For instance, Kirkeboen, Leuven and Mogstad (2016) and Heinesen et al. (2022) rely on administrative data from centralized admission systems to identify the next-best alternative for students applying to different fields and postsecondary institutions. The structure of the model allows us to fully characterize the ranking of preferences without requiring detailed data on individual preferences.

Finally, we also contribute to the literature that analyzes the role of latent abilities on schooling choices by documenting the empirical consequences of modeling the ordered nature of socioemotional tests. Most of the available socioemotional tests use a Likert scale. Therefore, the psychological tests usually employed to derive noncognitive ability distributions ask respondents to declare which statement, each associated with an integer, best describes their personality or behavior. Consequently, the different values reported for these tests only contain information about the order of preferences as they are not cardinal measures. Previous work using measurement systems usually models these tests as if they were cardinal test scores (see Attanasio et al., 2020; Hansen, Heckman and Mullen, 2004, for some exceptions). In this chapter, we show that the estimated distribution of the socioemotional latent ability is different if the ordered nature is considered. Assuming cardinality causes the estimated distribution to depend heavily on one test, the one informing the most about the latent ability. This dependency will increase with the number of mixtures determining the degree of flexibility of the ability's estimated distribution. Hence, the higher the flexibility of the distribution of the latent variable, the higher the over-dependence on this test. However, a common limitation in studies utilizing the NLSY97 to analyze socioemotional abilities is the timing of data collection, which often occurs either after individuals have made significant postsecondary schooling decisions or after completing them. Consequently, the interpretation of findings related to this ability dimension must be approached cautiously, as observed tests may be influenced by the educational experiences individuals experience during college. As in the second chapter of this dissertation, the implicit assumption in interpreting the lack of sorting observed for socioemotional abilities is that they remain stable from high school graduation into the mid- or late-twenties—when these measures are collected in the NLSY97.

As a preview of our results, we find different ability distributions and returns to ability for women and men. We observe that latent abilities assume crucial and distinctive roles depending on both the schooling decision and the group. Notably, the correlation between cognitive and mechanical abilities among men stands at 0.46, while it nearly doubles for women, reaching 0.86. Moreover, the socioemotional ability determines women's earnings associated with a high school diploma but not men's. However, our decomposition analysis shows that unobserved abilities are not always the primary factor that explains the level and slope of the MTE. Cognitive ability predominantly shapes the observed heterogeneity in men's marginal treatment effects. Conversely, observed factors are crucial in determining the level and slope of women's marginal treatment effects, with mechanical and cognitive abilities playing a comparatively minor role. In addition, we examine each observed variable's effect on MTEs. This further decomposition shows heterogeneous effects in terms of variables such as the number of children, urban residency, and region of residency, revealing changes in the composition of those at the margin of indifference.

We show, aligned with previous results by Mountjoy (2022), that the MTE of a two-year college education is only positive for those whose next-best alternative is staying with a high school diploma. In other words, incentivizing students to switch from four- to two-year colleges generates, on average, reduced earnings at the margin of indifference. At the same time, we would observe positive effects if these incentives were provided to those who would stay as high school graduates in the absence of a policy. On the other hand, the MTE of a four-year college education is positive for all groups. However, it is usually higher for those who would switch from being a high school graduate to attaining a four-year college degree. These results underscore the importance of understanding the ranking of preferences, as the next-best alternative can determine the level and sign of the marginal effect. We show that for those who decide to stay with a high school diploma, the most common second-best alternative is obtaining a four-year college degree for both men and women. We also show that 10.9% of women and 8.3% of men ranked getting a four-year college as their first best and attaining a two-year degree as their second best. Within these fractions, those closer to the margin of indifference would exhibit a negative effect if they were incentivized to get a two-year college degree.

The following section introduces our model and the MTE in a simplified setting, considering only two sectors or alternatives. Section 3.3 explains the decomposition into observed and unobserved components and discusses the practical implementation of the MTE and its decomposition. Section 3.4 extends the discussed methodology to a framework with three sectors and details the further decomposition considering all pairwise comparisons of different margins of indifference. Section 3.5 unifies all these elements in a static Roy model with latent abilities. Then, Section 3.6 discusses the estimation method and describes our data and sample in detail. This section also discusses the goodness of fit of our model and the empirical consequences of incorporating the ordered nature of socioemotional measures in the measurement system's assumptions. Section 3.7 presents our main results, and Section 3.8 discusses our main conclusions.

3.2 Basic Model: Two-sector case

First, we examine a Roy model with only two sectors, where a specific sector is denoted with i ($i \in \{0, 1\}$). This framework has been used extensively to bridge the treatment effect and structural econometrics literature. In this model, potential outcomes (Y_0, Y_1) can be decomposed in terms of their mean, conditional on observed characteristics X, and individual deviations from the mean or unobserved components (U_0, U_1), as follows:

$$Y_1 = \mu_1 (X) + U_1, \tag{3.1}$$

$$Y_0 = \mu_0 (X) + U_0. \tag{3.2}$$

Y_i denotes the outcome that agent i receives upon selecting sector i. Please note the omission of individuals' subscripts for notation simplicity.

A choice model dictates how individuals choose one of the available sectors. A latent variable D^* captures the utility derived from choosing sector 1 while normalizing the utility of sector 0 to zero. Consequently, we define the following binary indicator that takes the value corresponding to the chosen sector:

$$D = \begin{cases} 1 \text{ if } D^* = \mu_c(Z) - V \geqslant 0 \\ 0 \text{ otherwise.} \end{cases} \tag{3.3}$$

This expression means that individuals choose the sector that generates the highest utility, which is determined as a mean function of observed characteristics Z and an unobserved component V. We assume that:

$$U_1 \not\!\perp\!\!\!\perp U_0 \not\!\perp\!\!\!\perp V,$$

where $\not\!\perp\!\!\!\perp$ denotes no independence; the unobserved components are permitted to be correlated. Given the definition of D, the binary indicator, we can conveniently express the observed outcome Y as follows:

$$Y = Y_1 D + Y_0 (1 - D),$$

This expression encapsulates the central challenge of causal inference, which is fundamentally a missing data problem. For each individual unit, only one of either Y_0 or Y_1 is observable, with the observed outcome contingent upon the individual's choice. Notice that:

$$\begin{aligned} Y &= Y_1 D + Y_0 (1 - D) \\ &= \mu_0(X) + \beta D + U_0, \end{aligned}$$

where $\beta = Y_1 - Y_0$. This implies a distribution of gains, meaning that the effect of the treatment is heterogeneous, with β varies across units even after accounting for X. Consequently, the response to treatment, or the choice of D, is partly shaped by idiosyncratic unobserved heterogeneity, a factor that influences even individuals who appear observationally identical (see Heckman, Urzua and Vytlacil (2006) for a more detailed discussion).

In this study, we will focus on two treatment parameters:

$$\Delta^{MTE}(x, v) = E\left(Y_1 - Y_0 \mid X = x, V = v, D^* = 0\right), \tag{3.4}$$

$$\Delta^{TT}(x) = E\left(Y_1 - Y_0 \mid X = x, D = 1\right). \tag{3.5}$$

The first equation corresponds to the marginal treatment effect (MTE) parameter, which measures the mean gain in outcomes for persons or units who would be indifferent between treatment or not when $V = v$ and $X = x$. The second equation corresponds to the treatment on the treated (TT) parameter, which measures the mean gain for individuals who receive treatment (or choose sector 1), given $X = x$.

Define $P(Z)$ as the probability of receiving treatment given Z, or in the lenses of a discrete choice model, of choosing sector 1 given Z. Then $\Pr(Z) \equiv \Pr(D = 1 \mid Z) = F_V(\mu_c(Z))$, where $F_V(\cdot)$ denotes the distribution function of V. If we assume that $(X, Z) \perp\!\!\!\perp (U_1, U_0, V)$, the MTE can be estimated over the support of P and is additive separable in X and V.

3.3 Practical Implementation

The MTE can be empirically identified using the Local Instrumental Variables (LIV) estimator. Let $U_D = F_V(V)$, which follows a uniform distribution on the $[0, 1]$ interval. Then, the

LIV estimator is:

$$\frac{\partial}{\partial P(z)} E\left(Y \mid X = x, P(Z) = P(z)\right)\Big|_{P(z)=u_D} = E(Y_1 - Y_0 \mid X = x, U_D = u_D). \tag{3.6}$$

The left-hand side of this equation provides an expression for the MTE based on only observable variables (Y, X, Z). The LIV parameter can be estimated for any $P(z)$ within the support of $P(Z)$ and defines a treatment effect for those with a given unobserved propensity to participate (u_D) that is equal to $P(z)$. Note that the expression on the right-hand side is equivalent to equation 3.4, using u_D instead of v for determining the evaluation point of unobservables and having $P(z) = u_D$ as the point of indifference to satisfy the equality. Observe that a high level of $P(z)$ identifies the MTE for a high level of $U_D = u_D$, meaning that it identifies the effect for individuals whose unobservables make them less likely to choose sector 1. Recall that $P(z)$ is a monotonic transformation of $\mu_c(z)$, thus, high levels of $\mu_c(z)$ would be required to compensate for high levels of v under indifference ($\mu_c(z) - v = 0$).

Following Carneiro, Heckman and Vytlacil (2010), consider the introduction of a metric function $q(\cdot, \cdot)$ to capture the distance between $P(Z)$ and U_D. This function empirically defines indifference between sectors. Therefore, the MTE can be re-expressed as follows:

$$\Delta^{MTE}(X, u_D; \epsilon) = E(Y_1 - Y_0 \mid X = x, q(P(z), u_D) < \epsilon),$$

where ϵ denotes a user-defined tolerance threshold. As discussed by Carneiro, Heckman and Vytlacil (2010), different metrics will yield different estimates and identify distinct individuals at the indifference margin. However, the metric can be selected in a fashion such as the change in the decision rule has a policy change interpretation.

Note that the MTE considered so far is an expression conditional on X, the observed variables affecting sector-specific outcomes. However, conditioning on X might be empirically infeasible to satisfy if X contains continuous variables. Therefore, we consider instead an MTE averaged over the distribution of X. For a given ϵ:

$$\begin{aligned}
\Delta^{MTE}(u_D;\epsilon) &= \int \Delta^{MTE}\left(x, u_D;\epsilon\right) dF_{X|U_D=u_D, q(P,U_D)<\epsilon}(x) \\
&= E(Y_1 - Y_0 \mid U_D = u_D, q(P,U_D) < \epsilon).
\end{aligned}$$

Then, every point of indifference involves integrating over a distinct relevant distribution ($F_{X|U_D=u_D,q(P,U_D)<\epsilon}$). Note that no assumption required for estimating the MTE precludes the possibility that Z and X are correlated over the set defined by $q(P,u_D) < \epsilon$. Therefore, the contribution of X to the MTE might change as one evaluates different values of $(u_D, P(z))$. We propose the following decomposition to examine the extent to which this correlation with observed variables or with unobserved components (U_1, U_0) explains the MTE:

$$\begin{aligned}
\Delta^{MTE}(u_D;\epsilon) &= \int \left[\mu_1(x) - \mu_0(x) + U_1 - U_0\right]\ dF_{X|U_D=u_D,q(P,U_D)<\epsilon}(x) \\
&= \int \left[\mu_1(x) - \mu_0(x)\right]\ dF_{X|U_D=u_D,q(P,U_D)<\epsilon}(x) \\
&\quad + \int \left[U_1 - U_0\right]\ dF_{X|U_D=u_D,q(P,U_D)<\epsilon}(x) \\
&= \beta_X(U_D = u_D, q(P(Z),U_D) < \epsilon) \\
&\quad + \beta_{u_D}(U_D = u_D, q(P(Z),U_D) < \epsilon).
\end{aligned} \tag{3.7}$$

The first term captures how much of the MTE is explained by observed factors ($\mu_1(x) -$

$\mu_0(x)$), and the second term captures how much of the MTE is explained by unobserved factors ($U_1 - U_0$), that in our specific application will refer to the role of abilities. It is crucial to recognize that the integration over the distribution of X, conditional on each indifference point, means the sample for each u_D may vary. Consequently, this decomposition also reveals how the characteristics influencing sector selection or treatment (Z, V) relate to the outcome components.

We examine this decomposition in a multinomial setting, introduced in the following section.

3.4 Unordered model: HS, 2YC, and 4YC,

This section introduces a generalization of the previously described model, expanding the choice framework to allow individuals to select from a broader range of alternatives beyond the binary options considered initially. Let $\mathcal{J}$ be the set of all sectors under consideration, with no predetermined order of preference between elements. Consequently, this model acknowledges that individual preferences will lead to varied rankings of these options, with no inherent hierarchy among them. Additionally, let C_j determine the utility derived from selecting sector $j \in \mathcal{J}$:

$$C_j = \varphi_j(Z_j) - V_j, \tag{3.8}$$

and, similarly, define Y_j as the outcome associated with selecting sector j:

$$Y_j = \mu_j(X_j) - U_j, \tag{3.9}$$

Then, we can define the discrete choice model governing sector selection as follows:

$$D_{\mathcal{J},j} = \begin{cases} 1 & \text{if} \quad C_j \geqslant C_k \quad \forall k \in \mathcal{J} \\ \\ 0 & \text{otherwise.} \end{cases}$$

where $D_{\mathcal{J}_j}$ is an indicator variable defined for each sector j, taking the value of one exclusively when the utility from sector j, denoted by C_j, surpasses that of all other considered sectors.

Define $C_{\mathcal{J}}$ as the highest level of utility achievable within the choice set $\mathcal{J}$, represented as:

$$C_{\mathcal{J}} = \max_{j \in \mathcal{J}} \{C_j\} = \sum_{i \in \mathcal{J}} D_{\mathcal{J},j} C_j. \tag{3.10}$$

Given that individuals select only the sector yielding the highest utility, the econometrician is limited to observing:

$$Y_{\mathcal{J}} \equiv \sum_{j \in \mathcal{J}} D_{\mathcal{J},j} Y_j, \tag{3.11}$$

which is the outcome associated with the utility-maximizing sector.

Following Heckman, Urzua and Vytlacil (2006), we can use these elements to define the MTE in a multinomial setup. The main challenge when more than two sectors or alternatives are available is that we do not have a clear counterfactual to calculate the relative change in outcomes, while in the binary setup, there is only one choice left for comparisons. To address this, we consider two distinct choice sets defined by the next-best alternative. First, we define $\mathcal{K} = \{k\}$, a singleton set with only alternative k. Then, we define $\mathcal{L} = \mathcal{J} \backslash \{k\}$, which is the choice set with all available sectors but k. These two sets allow us to define the following MTE:

$$\Delta^{MTE}_{\mathcal{K},\mathcal{L}}(x, u_{D_{\mathcal{K}}}) = E\left[Y_{\mathcal{K}} - Y_{\mathcal{L}} \mid X = x, C_{\mathcal{K}} = C_{\mathcal{L}}\right]. \tag{3.12}$$

This MTE measures the average difference in potential outcomes for individuals indifferent between option k and the next-best alternative–the choice individuals would have selected if option k were unavailable. Note that if we consider a model with three sectors, set $\mathcal{L}$ would encompass two distinct choices. We explore this scenario in more detail in the following subsection.

3.4.1 Three-sector MTE

In our specific application, we study the treatment effects of post-secondary educational choices. Agents choose one among three different choices. They could stop their education after high school graduation, staying only with a high school diploma (HS); they may pursue a two-year college degree (2YC); or they may enroll in a four-year college program to obtain the respective degree (4YC). These distinct sectors define a specific set $\mathcal{J} = \{$HS, 2YC, 4YC$\}$ containing only three elements. Therefore, each MTE will be defined as a weighted average of two pairwise MTEs. To exemplify this idea, let's explicitly derive the MTE for being a four-year graduate that is defined using the sets $\mathcal{K} = \{$4YC$\}$ and $\mathcal{L} = \{$HS, 2YC$\}$. Then we can express equation 3.12 as follows:

$$\begin{aligned}\Delta^{MTE}_{4YC,\mathcal{L}}(x, u_{D_{4YC}}) =& \Pr\left(C_{2YC} > C_{HS} | X = x, u_{D_{4YC}}, C_{4YC} = C_{2YC}\right) \\ & \times E\left[Y_{4YC} - Y_{2YC} | X = x, u_{D_{4YC}}, C_{4YC} = C_{2YC}\right] \\ & + \Pr\left(C_{2YC} \leqslant C_{HS} | X = x, u_{D_{4YC}}, C_{4YC} = C_{HS}\right) \\ & \times E\left[Y_{4YC} - Y_{HS} | X = x, u_{D_{4YC}}, C_{4YC} = C_{HS}\right].\end{aligned} \tag{3.13}$$

Then, the MTE for being a four-year graduate is determined by the probability that having a two-year degree would be preferred over a high school degree and being indifferent between a four- or two-year college degree, multiplied by the MTE for the binary case where staying as a high school graduate is not an option. This result is added to the probability that having a high school diploma would be preferred over a two-year college degree and being indifferent between a four-year college degree or a high school diploma, multiplied by the binary MTE where attaining a two-year college degree is no longer an option. These probabilities can be re-expressed as weights, and using Mountjoy (2022)'s nomenclature, we obtain:

$$\begin{aligned}\Delta^{MTE}_{4YC,\mathcal{L}}(x, u_{D_{4YC}}) =& \omega_{4YC}(x, u_{D_{4YC}}) \times MTE_{4YC\leftarrow 2YC}(x, u_{D_{4YC}}) \\ & + (1 - \omega_{4YC}(x, u_{D_{4YC}})) \times MTE_{4YC\leftarrow HS}(x, u_{D_{4YC}}),\end{aligned} \tag{3.14}$$

where the $k \leftarrow j$ subscript denotes the groups of individuals being indifferent between k and j, or who would switch from j to k if a policy aimed to increase (reduce) the utility (costs) of option k were implemented. Using the same rationale, we can equivalently define the MTE for attaining a two-year degree as follows:

$$\Delta^{MTE}_{2YC,\mathcal{L}}(x, u_{D_{2YC}}) = \omega_{2YC}(x, u_{D_{2YC}}) \times MTE_{2YC \leftarrow 4YC}(x, u_{D_{2YC}}) + (1 - \omega_{2YC}(x, u_{D_{2YC}})) \times MTE_{2YC \leftarrow HS}(x, u_{D_{2YC}}). \tag{3.15}$$

We will apply the decomposition into unobservables and unobservables, as shown in equation 3.7, to each of the pairwise MTEs in equations 3.14 and 3.15.

3.5 Roy Model with latent abilities

We employ a parametric generalized Roy model with latent abilities to analyze the returns to the three postsecondary choices that were explained in the previous section. Roy models have been extensively used to analyze returns to education (Heckman, Humphries and Veramendi, 2018; Heckman, Stixrud and Urzua, 2006; Prada and Urzúa, 2017) as well as in the marginal treatment effects literature (Carneiro, Heckman and Vytlacil, 2010; Heckman and Vytlacil, 2007, 2005; Heckman, Urzua and Vytlacil, 2006). In this model, individuals make schooling choices $j \in \mathcal{J} = \{\text{HS, 2YC, 4YC}\}$ to maximize their utility. Utility from choosing j is given by $C_j = Z_j'\gamma_j - V_j$ and earnings from choosing j by $Y_j = X_j'\beta_j + U_j$. V_j and U_j are unobserved from the perspective of the econometrician but not from the perspective of the decision maker. They have the following structure:

$$\begin{aligned} V_j &= \theta'\lambda_j^c + \varepsilon_j, \\ U_j &= \theta'\lambda_j^y + \epsilon_j \qquad \text{for } j \in \mathcal{J}, \end{aligned}$$

where θ is a vector of latent abilities that is predetermined before making any postsecondary decision. ε_j and ϵ_j are normally distributed and independent errors; where $\theta \perp\!\!\!\perp \{\varepsilon_j, \epsilon_j\}_{j=\{HS,2YC,4YC\}}$ and $\{\varepsilon_{j'}, \epsilon_{j'}\} \perp\!\!\!\perp \{\varepsilon_j, \epsilon_j\}$ for j' and j in {HS,2YC,4YC}. Despite the assumptions on the error terms, note that θ creates correlation across all unobserved components.

Moreover, we assume that $(X, Z) \perp\!\!\!\perp \{U_j, V_j\}_{j=\{HS,2YC,4YC\}}$. This condition guarantees non-parametric identification of all the MTE components that will be decomposed (Carneiro, Hansen and Heckman, 2003). Nonetheless, in practice, and relying on parametric identification, weaker assumptions can be considered. Despite all the assumptions on independence, X and Z may still be correlated. This fact will be crucial to interpret results in section 3.7

3.5.1 Identification of Latent Abilities

θ is a vector of latent abilities, $\theta = [\theta_c, \theta_s, \theta_m]$, encompassing cognitive (θ_c), socioemotional (θ_s), and mechanical (θ_m) dimensions. Previous studies have highlighted the importance of these dimensions in influencing post-secondary choices and their subsequent returns (Belzil, Hansen and Liu, 2022; Heckman, Stixrud and Urzua, 2006; Prada and Urzúa, 2017). We model abilities using a measurement system and factor variables following previous literature. The measurement system enables the factor variables to inherit the intended interpretation of the latent abilities and facilitates their identification. The latent variables resulting from this process aim to capture inherent abilities and their development within a specific environment with distinct resources until individuals are in high school and ready to determine their optimal post-secondary investments. The main idea behind the measurement system is to use tests related to each dimension to recover the free-of-measurement-error latent abilities. We impose the following structure for the measurement system:

$$
\begin{aligned}
c_r &= Q_c' \eta_{c_r} + \theta_c \lambda^c_{c_r} + \varepsilon_{c_r} && \text{for } r = 1, ..., R \\
s_l^* &= Q_s' \eta_{s_l} + \theta_s \lambda^s_{s_l} + \varepsilon_{s_l} && \text{for } l = 1, ..., L \\
m_p &= Q_m' \eta_{m_p} + \theta_m \lambda^m_{m_p} + \theta_c \lambda^c_{m_p} + \varepsilon_{m_p} && \text{for } p = 1, ..., P.
\end{aligned}
\tag{3.16}
$$

On the left-hand side of the system, we find test scores associated with cognitive (c_r), socioemotional (s_l^*), and mechanical traits (m_p). The main intuition behind considering a measurement system is to treat these tests only as noisy measures of each ability. Therefore, we impose structure and use factor variables to capture the information from several tests left after removing measurement errors. Q_c, Q_s, and Q_m are observed variables that affect test performance. Controlling for them allows us to remove any effect from background and family student characteristics that should not affect the actual ability and, hence, could generate non-classical measurement error. ε_{c_r}, ε_{s_l}, and ε_{m_p} are independent and normally distributed errors, capturing the fact that test scores are also subject to random error. $\eta_{c_r}, \eta_{s_l}, \eta_{m_p}, \lambda^c_{c_r}, \lambda^s_{s_l}, \lambda^m_{m_p}$, and $\lambda^c_{m_p}$ are parameters to be estimated. Note that for each type of test, there are several measures, each generating a different equation: R cognitive, L socioemotional, and P mechanical equations.

We impose several restrictions to achieve identification.[2] First, we restrict the relationship between test scores and θ. As the system shows, the mechanical measures inform about cognitive and mechanical abilities, while socioemotional (mechanical) measures are exclusively dedicated to informing about socioemotional (mechanical) abilities. Moreover, θ_m and θ_c are allowed to be correlated, while θ_s is assumed to be independent of other abilities. These assumptions are intro-

[2]See Carneiro, Hansen and Heckman (2003) for a deeper discussion about the identification of measurement systems.

duced after conducting an exploratory factor analysis on the observed measures.[3] Furthermore, we impose $\lambda^c_{c_1} = \lambda^s_{s_1} = \lambda^m_{m_1} = 1$ to set the scale of the factors to the scale of the first measure in each system of equations. Consistent with standard practice in the field, we also set $\lambda^c_{m_P} = 0$ to ensure that there is at least one dedicated measure for the mechanical ability, which helps identify the correlation between θ_c and θ_m.

In contrast to most of the previous literature analyzing abilities and returns to schooling, we derive θ_s from an ordered system. Observed socioemotional measures s_l are not cardinal, as they use a Likert scale. Therefore, respondents rank how much their personality aligns with a given trait or behavior. Previous work usually treats these measures as cardinal. Instead, we model the ordered nature of each question and assume that s^*_l, the associated latent variables, follow a normal process. Therefore, the socioemotional system is estimated as an Ordered Probit. We introduce standard assumptions to achieve identification in such a system. First, Q_s does not contain a constant term. Second, the variance of all the error terms associated with socioemotional measures is assumed to be one ($Var(\varepsilon_{s_l}) = 1$ for $l = 1, ..., L$). With these assumptions, we can estimate the cutoffs of the latent variables that generate the different levels in the observed measures (Muthén, 1984). Later, in subsection 3.6.7, we describe the empirical consequences of deviating from assuming cardinality.

3.6 Estimation - CMLE

The model is estimated using Conditional Maximum Likelihood Estimation (CMLE). This procedure defines two stages. In the first stage, we estimate the parameters associated with the

[3]The exploratory factor analysis is available upon request. We obtain the same conclusions using oblique or orthogonal rotations.

distribution of latent abilities and test scores. In the second stage, fixing the distribution of abilities, we get estimates for the parameters affecting earnings and schooling choices. As test scores are ideally observed before making any postsecondary decision and θ is ideally measuring endowments before college, this two-step procedure ensures that the factor variables are primarily identified based on theoretically proposed relationships with test scores rather than being influenced by postsecondary decisions and earnings.

To describe in detail the estimation procedure, consider the likelihood function associated with the model explained in Section 3.5:

$$\int_{\underline{\Theta}} \left\{ \prod_{t \in \mathbf{T}} f(t|B,\theta;\delta) \prod_{j \in \mathcal{J}} \left[f(Y_j|B,\theta;\delta) \Pr(D_{\mathcal{J},j}=1|B,\theta;\delta) \right]^{D_{\mathcal{J},j}} \right\} dF_\theta(\theta), \tag{3.17}$$

where $\mathbf{T}$ is set of test scores, B represents the set of explanatory variables, and δ the vector of all parameters. In addition, $\underline{\Theta}$ denotes the support of θ and $F_\theta(\cdot)$ is the joint cumulative distribution of θ.

The choice probabilities (Pr$(D_{\mathcal{J},j} = 1|B,\theta;\delta)$) do not have closed-form solution when more than two options are available to choose from. Our application considers three sectors, so we use the smoothed accept–reject method to estimate choice probabilities considering 100 evaluation points.

The distribution of the factors is allowed to follow a flexible distribution using mixtures of normal distributions as follows:

$$\theta_i \sim \sum_{j=1}^{h_i} p_j^i N(\mu_j^i, \sigma_j^{i2}) \qquad \text{for } i = c, s, u \tag{3.18}$$

where μ_j^i and σ_j^{i2} denote the means and variances for each component j of each factor θ_i. θ_u denotes the uncorrelated component of the mechanical system, as $\theta_m = \alpha\theta_c + \theta_u$. As shown by Prada and Urzúa (2017), α can be identified if we assume that one of the mechanical measures is a mechanical-dedicated test score, meaning it can only inform about the mechanical ability. Using Likelihood Ratio tests, we choose h_i, the number of mixture components. Furthermore, we restrict the mean of each factor to zero to allow the identification of the intercepts ($\sum_{j=1}^{h_i} p_j^i \mu_j^i = 0$). We solve the high-order integral over θ using Monte Carlo Simulation and 100 Halton draws per mixture-ability dimension pair.

To fully understand the two stages in the estimation problem, consider a partition of δ into δ_M and δ_R. δ_M contains all the parameters related to equations 3.16 and 3.18. Therefore, δ_M also includes all the parameters of the measurement system, including variances of errors when relevant, and all the parameters of the measurement system and mixtures of the factors. δ_R, on the other hand, contains all parameters related to utility and outcomes. In step one, the following likelihood function is maximized:

$$\int_{\underline{\Theta}} \left\{ \prod_{t \in \mathbf{T}} f(t|B, \theta; \delta_M) \right\} dF_\theta(\theta). \tag{3.19}$$

In the second step, we maximize the following function:

$$\int_{\underline{\Theta}} \left\{ \prod_{t \in \mathbf{T}} f(t|B, \theta; \hat{\delta_M}.\delta_R) \prod_{j \in \mathcal{J}} [f(Y_j|B, \theta; \delta_R)\Pr(D_{\mathcal{J},j} = 1|B, \theta; \delta_R)]^{D_{\mathcal{J},j}} \right\} d\hat{F}_\theta(\theta), \tag{3.20}$$

taking as given $\hat{\delta_M}$ and $\hat{F}_\theta$. To incorporate the uncertainty from both steps into the standard errors, we consider a one-iteration last step where the Hessian matrix is calculated. In this last step, all parameters are jointly estimated without constraints.

3.6.1 Data

We analyze data from the National Longitudinal Survey of Youth 1997 (NLSY97), a longitudinal study that follows individuals born between 1980 and 1984 and living in the United States during the first round (1997-1998). We focus on the cross-sectional sample of 6,748 men and women representative of the selected cohorts.[4] This data set offers rich information on background characteristics, annual earnings, postsecondary choices, and test scores. To measure cognitive and mechanical abilities, we use the Armed Services Vocational Aptitude Battery (ASVAB) test that is also available in the NLSY79 and has been extensively used to measure these abilities in previous studies (see Heckman, Urzua and Vytlacil, 2006, Prada and Urzúa, 2017, and Belzil, Hansen and Liu, 2022). The ASVAB test was administered from the summer of 1997 through the spring of 1998 when participants were at most 18 years old. Following previous literature, we consider six ASVAB components as cognitive measures: math knowledge, word knowledge, paragraph comprehension, numerical operations, coding speed, and arithmetic reasoning. We use mechanical comprehension, electronics, shop information, and auto information for the mechanical measures. Although we use the same measures available in the NLSY79, one difference is worth discussing. Shop information and auto information are available as one single measure in the longitudinal study of 1979, while separated in the NLSY97.

In assessing the dimensions of socio-emotional abilities, the National Longitudinal Survey of Youth 1997 (NLSY97) makes use of the Ten-Item Personality Inventory (TIPI) Scale, administered in round 12 (2009). This scale asks participants to self-evaluate their compatibility with

[4] We exclude the supplemental samples designed to oversample Hispanic, Latino, and Black populations.

various attributes. It is an effective tool for measuring the Big Five personality traits: extroversion, agreeableness, conscientiousness, emotional stability, and openness. A useful feature of the TIPI is its dual-question format for each trait, formulated in contrasting ways, which helps counteract the common survey issue of agreement bias. Thus, we adjust the order of the opposing questions for each trait and then compute their average, obtaining only one measure per trait.

In addition to the NLSY97 dataset, we incorporate data from the Integrated Postsecondary Education Data System (IPEDS) and the Current Population Survey (CPS) to construct geographic variables. IPEDS provides information on tuition across two- and four-year institutions, while CPS supplies data on wage rates and unemployment rates for various demographic groups. We will elaborate on the specific definitions of these variables later when outlining our empirical specification.

To define our sample, we exclude high school dropouts, focusing on individuals with a high school diploma or a GED certificate, which are prerequisites for pursuing postsecondary education. We include only those who had not begun their college education when they took the ASVAB test, ensuring that the cognitive and mechanical measures are unaffected by postsecondary education. In addition, earnings are measured when individuals are 28 to 30 years old, excluding those still in college at this age or with insufficient data in this age range. Furthermore, we omit observations with missing data for the variables considered in our specifications.[5] These criteria result in a sample comprising 1,181 men and 1,202 women. It is worth noting that Mountjoy (2022) focuses on individuals whose projected high school graduation year was between 2000 and 2004, just a few years later than the projected graduation years in the NLSY97. Ad-

[5]We perform imputations for earnings and the number of children. This imputation process is particularly beneficial for addressing missing information arising from the biannual survey format of the NLSY97 post-2011.

ditionally, earnings are measured over ages 28–30 in both studies. These decisions ensure some degree of comparability between his sample and ours.

Table 3.1 presents the summary statistics by group. We present the mean of the main variables employed in the analysis for men and women but also further divide each group depending on their postsecondary educational attainment. The proportion of Black and Hispanic individuals decreases as the level of education increases, particularly among men. The data from 1997 reveal consistent urban residency among the groups, except for two-year college graduates, who are at least ten percentage points more likely to have lived in rural areas. The proportion of individuals residing in the Southern United States in 1997 remains relatively stable across all educational groups, varying around 30%. The prevalence of individuals from single-parent homes declines with higher education, as does the number of siblings. As expected, family income and the mother's schooling in 1997 positively correlate with educational attainment.

Upon examining the cohorts within the data, it is observed that among men, the 1983 birth cohort is marginally overrepresented, whereas the 1980 cohort is marginally underrepresented. Similarly, for women, there is a slight overrepresentation of the 1981 cohort and a slight underrepresentation of the 1984 cohort. This pattern of representation remains consistent across the educational groups analyzed.

Current residency patterns, observed when individuals are between 28 and 30 years old, show a decline in rural living across all groups. This shift is most apparent among men with two-year degrees, a group that no longer shows a marked rural presence—an observation that contrasts their teenage years. The only exception to this trend is women with two-year degrees, who persist in urban living at a rate unchanged since 1997, with 58% residing in cities. Women show a higher tendency to live in the South across all educational categories, whereas the distribution of current residency for men is more balanced across regions.

The table also underscores disparities in earnings and family choices. Men's earnings significantly outpace women's across all education levels, but the gap narrows with higher education. Regarding family choices, the data show that women generally have more children and are more frequently married than men. The average number of children for both men and women declines as educational attainment increases. The data also show that the proportion of individuals married is highest among those with two-year degrees.

The final row of Table 3.1 presents the total number of individuals in each subgroup. The samples consist of 1,181 men and 1,204 women. Among these, high school graduates represent 59.86% of men and 51.82% of women. Those with two-year college degrees account for 8.21% of men and 8.22% of women. Lastly, individuals with four-year college degrees make up 31.92% of the male cohort and 39.95% of the female cohort. The cohorts from the NLSY97, which started their college education in the early 2000s, exhibit a trend documented in national statistics: women attend and graduate from college at higher rates than men.

3.6.2 Raw measures

Figures 3.2 and 3.3 display the cognitive and mechanical abilities distributions among the analyzed groups. In these figures, solid blue and dashed orange lines represent the distributions for women and men, respectively. The most significant group differences are observed for numerical operations, coding speed, and arithmetic reasoning. Women usually score relatively higher in numerical operations and coding speed. Yet, men's distributions show a thicker right tail in these areas, meaning they are more likely to get top scores.

Conversely, men tend to outperform women in arithmetic reasoning, particularly for scores near and above the overall mean. The disparities in other cognitive measures, math and word knowledge and paragraph comprehension, are comparatively less pronounced. Figure 3.3 indicates that, in mechanical measures, men consistently outperform women, showcasing larger disparities compared to cognitive measures. While both groups have similar scores at the lower end of the distribution, men's scores are predominantly higher, evident from their right-shifted distributions and thick right tails.

Figure 3.4 shows the raw socio-emotional measures. These measures combine the two questions measuring the same trait, and the histograms present the total count of answers for each value. Blue bars display the responses for women, while orange bars represent men's responses. The data shows that women generally exhibit higher levels of agreeableness, extroversion, and conscientiousness, as illustrated in panels A, B, and C. Conversely, men tend to display greater emotional stability and a slightly higher openness to new experiences, as evident in panels D and E.

In general, the substantial differences in the raw ability measures observed between women and men support the separate estimation of the model for each group.

3.6.3 Point estimates

The empirical specification presented in Table 3.2 delineates the variables included in each equation. All types of equations consider cohort, race, and ethnicity dummies. Schooling choice equations (Schooling) and measurement system equations (TIPI and ASVAB tests) incorporate additional variables designed to capture individuals' backgrounds during their teenage years before they make any college decisions. These variables encompass family income, years of schooling of the mother, and dummy indicators for residing in the South, living in urban areas, and being raised in single-parent households. All these variables were measured in 1997, the first year of the study, and when survey participants were between 12 and 17 years old. The number of siblings is also included in these equations. However, the NLSY97 started collecting this information only in 2011.

Schooling choice equations measure the utility of different sectors. We need to normalize one of the sectors to have zero utility to achieve identification in discrete choice models. We designate staying with a high school diploma (HS) as this alternative. The schooling choice equations also include three regional variables that are measured when individuals are 17 years old: average public tuition, unemployment relative to high school graduates, and wage rates relative to relative to high school graduates. Sector-specific in nature, these variables capture variations pertinent to either two-year or four-year college graduates or institutions. Acting as exogenous cost-shifters, they account for variations influencing schooling choices while remaining independent of individual characteristics. Unlike the schooling choice equations, the measurement system equations do not include regional variables. Instead, they contain information about the years of schooling of the student when they took each respective test, whether the ASVAB test or the TIPI scale test. Consistent with previous literature, the inclusion of these variables helps mitigate heterogeneity in derived abilities by accounting for differences in accumulated schooling when test scores are observed.

The earnings equations encompass information contemporary to the time earnings are observed, which typically falls between ages 20 and 30 for individuals. They incorporate dummies for the region of current residency, urban area residence, and for being married. Moreover, we also consider the number of biological children in the household.[6]

In Table 3.3, we present reduced form estimates that depart from the structural framework outlined in earlier sections. We use OLS estimations with earnings, measured in tens of thousands and in 2010 dollars, as the dependent variable in these specifications. We include dummies for schooling choices, with staying with a high school diploma as the omitted category. Additionally, we control for all variables mentioned in the empirical specification. To control for abilities, we include the standardized mean of test scores, creating one index for each type of measure: cognitive, mechanical, and socioemotional. We present specifications with and without ability proxies. Furthermore, estimates are presented separately for men and women. These results indicate that postsecondary decisions significantly impact annual earnings at age 30. Specifically, obtaining a four-year college degree increases annual earnings by between \$13,280 and \$17,610, and a two-year college degree increases earnings by between \$5,840 and \$9,400, relative to individuals who do not attain any further schooling degree after high school. Although estimates vary across groups, Wald tests on the differences between women and men suggest that the effects are statistically equal across groups.[7] These results also suggest that test scores are relevant in explaining earnings, except for the socioemotional ability's impact on men's earnings and the mechanical ability's impact on women's earnings.

In Table 3.4, we compare the estimated effects of abilities on schooling choices and earnings using OLS (Panel A) with the estimates for the same effects using structural estimation of the proposed Roy model (Panel B). Estimates related to the school choice equation refer to the point estimates of the utility derived from each sector. This table demonstrates that the conclusions

[6] Controlling for these variables implies that our model assumes fertility and marriage choices as exogenous.
[7] See Table 3.19.

drawn from these two methodologies differ. In most cases, using reduced-form techniques tends to underestimate the effect of abilities, both in terms of the absolute value of the effect and the statistical significance level. Additionally, there are notable differences between the estimates for women and men, regardless of the specific method considered. Although, in the case of the structural estimates, women's and men's estimates come from two distinct models, we test whether the effects of abilities are different across groups. We confirm significant statistical differences between both groups, especially regarding earnings effects.[8]

We have demonstrated that the conclusions drawn from the reduced-form estimation differ from those obtained from the structural model. Additionally, interpreting reduced-form estimates requires caution due to several factors. Firstly, there is selection bias, as individuals who choose different schooling paths are not randomly selected and could have made a choice based on unobservables. Secondly, test scores may be susceptible to measurement error, and actual abilities likely influence schooling decisions and earnings. These factors introduce bias into reduced-form estimates. Consequently, we will focus on the results derived from the structural model, which align with the discussion presented in section 3.5. The point estimates from the structural model are displayed from Table 3.5 to Table 3.14.

Tables 3.5 and 3.10 present the point estimates for the utility of sectors 2YC and 4YC, with sector HS normalized to zero utility. Similar variables influence utility for both men and women, ultimately shaping their schooling choices. Notably, living with both parents, family income, and the mother's years of schooling all positively impact the likelihood of choosing a four-year college degree. These factors are indicative of long-term credit constraints and correlate with investments that facilitate ability development, thereby fostering readiness for college enrollment.

[8] See Table 3.20 for the t-statistics from these tests.

Additionally, cognitive ability positively influences the likelihood of pursuing a four-year degree, while a negative effect is observed for mechanical ability, although only marginally significant for men. All these findings align with previous literature. However, a notable difference between women and men is that significant cohort effects influence the utility of women attaining a four-year college degree, which is not observed among men.

Fewer variables statistically influence the utility of two-year college degrees. Residing in an urban area decreases the utility of a two-year degree for both groups, although this effect is only marginally significant for men. Additionally, living with both parents increases the likelihood of attaining this degree for both groups. Notably, maternal years of schooling increase the likelihood of women obtaining a two-year college degree, whereas cognitive ability only increases this likelihood for men.

Tables 3.6 and 3.11 present the point estimates related to the earnings equations for men and women, respectively. Several differences between the groups are worth highlighting. Firstly, being black increases the earnings of male high school graduates, while neither race nor ethnicity plays a role in determining women's earnings. Secondly, the number of children decreases women's earnings as four-year or high school graduates, while for men, having more children seems to increase earnings as four-year graduates weakly. For men, being married increases earnings in all sectors, but this variable does not statistically affect women's earnings. Cognitive ability increases earnings for four-year graduates in all cases. However, there are several differences in how abilities define earnings between groups. Cognitive ability increases the earnings of female high school graduates but not men's.

Similarly, socioemotional ability increases the earnings of female high school graduates but does not affect men's earnings. Finally, mechanical ability increases earnings for all sectors except for attaining a four-year degree for men. For women, this ability decreases the earnings of a two-year college degree.

Tables 3.7 and 3.12 display the point estimates of the cognitive system for men and women, respectively. Across both groups, the loadings, which are the coefficients associated with the factor variable, are similar. The tests with the highest loadings are arithmetic reasoning, paragraph comprehension, and mathematical knowledge. Additionally, maternal education emerges as a crucial factor that increases cognitive scores for both women and men. However, some observed variables affect cognitive test scores differently across groups. Specifically, for men, coming from a two-parent household and family income seem to have less impact. In contrast, these characteristics consistently play crucial roles for women in all tests.

The estimates for the socioemotional system are shown in Tables 3.8 and 3.13. Once again, the loadings are similar across groups. The tests with the highest loadings are related to emotional stability and agreeableness, while the lowest loadings are for extroversion. These results need to be considered when interpreting the derived socioemotional factor variable. The point estimates for the mechanical system are displayed in 3.9 and 3.14. Despite mechanical measures coming from the same test as the cognitive measures, there are differences in terms of the role of observed characteristics. Mother's years of schooling continue to be an important explanatory factor but to a lesser extent than in the case of the cognitive system. Coming from a two-parent household only increases the mechanical knowledge and electronics information scores.

Interestingly, family income is not a significant factor except for men, where higher parental income decreases the shop information test scores. Loadings between groups differ for this system. The highest mechanical loading for men is associated with the auto and shop information tests, while for women, it is only the loading associated with auto information. Additionally, cognitive ability decreases the auto information score for women while having a null effect for men.

3.6.4 Marginal Densities

Figure 3.5 illustrates the marginal densities of latent abilities. Panel A presents the densities for men, while Panel B displays those for women.[9] All three densities show variations between men and women, supporting the decision to estimate separate models for each group. The cognitive and mechanical abilities depicted in blue and green exhibit a higher variance for men. In the case of cognitive ability, the higher variance is mainly explained by a thicker left tail, while for mechanical ability, men exhibit a thicker right tail. Therefore, a higher proportion of men have exceptionally high mechanical endowments and very low cognitive levels relative to women. The socioemotional densities are normally distributed but have a slightly higher variance for women. Refer to Figure 3.35 in the Appendix to directly compare ability types across groups.

The most notable distinction, however, lies in the correlation between cognitive and mechanical abilities. For men, we estimate a correlation of 0.46, which aligns with findings from a similar system estimated using the ASVAB test in the NLSY79 dataset, focusing solely on men's information (Prada et al., 2017). Conversely, for women, the estimated correlation almost doubles, reaching 0.86. This substantial difference carries significant implications for interpreting the results presented in this article. Specifically, women's mechanical ability appears to be highly dependent on cognitive ability, while men do not observe such dependency. Consequently, inferring that women's mechanical ability is more closely related to the ability required to succeed in

[9] Recall that the number of mixtures determining the distribution of the factor is chosen using likelihood ratio tests. We stop adding additional mixture components when likelihood ratio tests do not reject that the model's fit is equivalent to one using fewer mixtures.

STEM fields is plausible. In contrast, for men, the captured latent mechanical ability appears less reliant on cognitive endowments.

3.6.5 Variance Decomposition of the Measurement System

Figure 3.1 illustrates the variance decomposition of test scores within our measurement system to explain the relative contributions of observed variables (in darker blue), latent factors or abilities (in medium blue), and the error term (in light blue). Panel A and B present the decomposition for men and women, respectively. A consistent pattern across both groups is the lower proportion of variance explained by observed variables in socio-emotional measures; a trend also noted in studies using NLSY datasets. The variance attributable to abilities is relatively small for most socio-emotional measures, except for emotional stability. Notably, women's cognitive and mainly mechanical scores show a higher relative contribution of the error term. Additionally, mechanical measures for women exhibit lower variability, which is explained by abilities relative to men.

The presence of a significant error component across all measures for both groups underscores the importance of utilizing inferred distributions of unobserved abilities instead of relying solely on raw measures in our analysis.

3.6.6 Goodness of Fit

Measures

Figures 3.6-3.9 display the goodness of fit of the mechanical and cognitive predicted measures relative to the observed measures. Each figure follows a similar structure: a solid, light blue line represents the empirical Cumulative Distribution Function (CDF) of the observed measure, while a dashed, black line depicts the empirical CDF of the simulated measure—the measure predicted by the model post-estimation. We generate our simulated data using 1,000 replications of the original sample. Moreover, atop each graph, the p-value from a Kolmogorov-Smirnov test is displayed within parentheses. The test's null hypothesis proposes that both depicted CDFs come from the same distribution.

Figures 3.6 and 3.7 display the goodness of fit of cognitive measures for men and women, respectively. While the null hypothesis is rejected at a 95% confidence level for paragraph comprehension in both groups and for word knowledge and coding speed in men, none of them reject the null hypothesis at a 99% confidence level. Figures 3.8 and 3.9 depict the goodness of fit for the mechanical measures. In this case, all but one measure, mechanical comprehension, fails to reject the null under standard significance levels for both groups. Despite rejecting the Kolmogorov-Smirnov test at conventional significance levels, the empirical CDFs of the simulated measures closely resemble those of the observed measures in all cases.

Schooling Choices

Table 3.15 presents the goodness of fit of schooling choices on the left side. The first two columns display the fractions of individuals who choose each of the sectors, first for the simulated data and then for the observed data. Panel A shows these results for men at the top of the table, while Panel B displays these fractions for women. The fractions for men in each sector, both in the simulated and observed data, are very close. However, there is a slight overestimation of those staying as high school graduates by 1.7 percentage points, fully compensated by an underestimation of four-year graduates by the same magnitude. The same pattern applies to women, but the difference is larger, ranging from 3.5 to 3.7 percentage points. Although the fit for women is not as good as for men, in all cases, the fraction of individuals in each sector in the simulated data closely resembles those in the observed data.

Earnings

The second part of Table 3.15 presents some results describing the goodness of fit of earnings. The right side of this table once again compares summary statistics, now focusing on the mean and variance, between the simulated and observed earnings data, which, as a reminder, are measured in tens of thousands. For all sectors and groups, simulated earnings closely match the mean and variance of observed data.

Figure 3.10 presents additional information about the goodness of fit of earnings. This figure compares the density of observed (solid, light blue) and simulated (black, dashed) earnings. Additionally, the p-value of a Kolmogorov-Smirnov test is displayed in parenthesis above each graph. The goodness of fit is excellent for female and male two-year graduates, as the null hypothesis of simulated and observed earnings having the same distribution is not rejected in these cases. Although the null hypothesis is rejected for other sectors and groups, the simulated earnings can still replicate the general pattern observed in earnings.

3.6.7 Ordered vs Continuous System

This section discusses the empirical consequences of modeling or ignoring the ordered nature of socioemotional measures in the measurement system. To do so, we compare the results using two different assumptions or systems. Under the ordered system, we assume each socioemotional measure follows an Ordered Probit process, aligned with the assumptions considered in Section 3.5. This assumption captures the ordered nature of the measures surveyed using Likert scales. The second system, which we call a continuous system, assumes that the observed measures are cardinal. There, we assume that each test follows a normal distribution. Previous literature usually imposes this last assumption.

Table 3.16 presents the estimated loadings, or coefficients associated with the latent socioemotional ability, for men and women under different assumptions. Panel A shows the results after considering an ordered system, and Panel B shows the estimates from assuming a continuous system. We also show these estimates for models considering different numbers of mixtures, defining the distribution of the latent ability. Therefore, each column within a panel shows the results of one estimation of the measurement system. We can see that most loadings are relatively constant when we increase the number of mixtures, except in one case. This pattern is observed for stability in Panel B, which corresponds to the case where we ignore the ordered nature of the measures. Stability is the measure that has the highest correlation with the latent derived ability in all cases, as the loading is the highest. However, this table shows that under the continuous system, when the number of mixtures to model the distribution of the ability increases, the loading becomes much higher, especially in comparison to other loadings.

Number of Mixtures

Figures 3.11 and 3.12 display men's and women's socioemotional densities for different mixture specifications. Each specification assumes a different number of mixtures. We can see that the estimated densities are very different under a continuous and ordered system if we condition on the number of mixtures. This pattern is exhibited for both men and women.

However, someone could argue that the model with a higher number of mixtures could have better goodness of fit and, hence, be the "right" model. The legends in Figures 3.11 and 3.12 show p-values of a likelihood-ratio test comparing a given specification with the specification that uses only one additional mixture component. We can see that for men, the likelihood-ratio tests suggest using just one mixture—a normally distributed latent ability—under the ordered system. In contrast, the tests reject using even three mixtures for the continuous system. A similar pattern is observed for women in Figure 3.12, although for them, the model with two mixtures is rejected at 95% of confidence against the model that uses three mixtures under the ordered system. For the continuous system, a specification using three mixtures has the same fit as one using four.

Figure 3.13 compares the ordered and continuous system's goodness of fit for men. Figure 3.14 presents the results for women. We compare the CDFs of the observed (in solid, light blue) and simulated measures. Dashed black corresponds to CDFs of the predicted (or simulated) measures obtained under an ordered system, while solid red represents the predicted measures assuming a continuous system. We use one mixture for the ordered systems following the results of the likelihood-ratio tests. For the continuous system, we use four mixtures for men and three for women. We can see that the measures under the ordered system perfectly fit the observed data in all cases, while the ones derived from a continuous system do not.

Finally, Figures 3.15-3.18 show the variance decomposition of the socioemotional system under all specifications considered. Each color represents a different assumption. Dark blue corresponds to the decomposition obtained after assuming an ordered system, while light blue shows the results from assuming a continuous system. Each plot displays the result for a given measure or test, and each bar shows the decomposition of specifications using a given number of mixtures. Figure 3.15 shows the fraction of the test variance explained by the variance of the latent variable for men. We can observe that the fraction remains stable for all measures when we assume an ordered system. However, the fraction of the variance explained by the ability quickly increases to almost one in the case of the stability tests. The same pattern is observed for women in Figure 3.15.

Figures 3.17-3.18 present the relative participation of the error term in each of the socioemotional measures. Aligning with previous figures, we can see that the error variance drops quickly to almost zero for the continuous system and the stability measure. For the other measures, we do not see any unexpected behavior.

This section concludes that choosing between an ordered or continuous assumption for socioemotional measures holds significant implications, as the resulting distributions exhibit notable disparities. Adopting a continuous framework leads to a distribution heavily influenced by the test with the highest loading, a dependency exacerbated by increasing the number of mixtures. Given this trend and the superior fit observed with an ordered system, we opt for adopting an ordered approach.

3.7 Main Results

3.7.1 Ranking of Preferences

Analyzing preference ranking when more than two options are available is crucial to predicting the potential impact of policy interventions. In the following subsection, we will examine all the pairwise comparisons of treatment effect parameters. When designing policies targeting a specific sector, such as subsidizing two-year college tuition, understanding the likelihood of preferring a two-year college degree as the second option will inform how we should weigh the different calculated treatment effect pairs.

Table 3.17 presents the percentage of individuals by ranking of preferences for women and men. For both groups, the most preferred ranking, with more than 31% of women and men preferring this option, is staying as a high school graduate in the first place and obtaining a four-year college degree in second place. However, the second most preferred ranking differs between men and women. The second most preferred ranking for women, preferred by 25% of women, is having to get a four-year college degree as the first option and to stay with a high school diploma as the second best. Around 30% of men prefer staying as a high school graduate, and as their second preference, they would prefer attaining a two-year degree. For both groups, the least preferred ranking is achieving a two-year degree as the first preference and a four-year degree as the second preference.

The percentage of individuals ranking each alternative first, second, or third place is displayed in Table 3.18. Here, we can see that the most common least preferred degree or sector is a two-year college degree, with 58.7% of women and 53.3% ranking it as their third option. This table also shows that the most common second-best option differs between groups. For men, the most common second preferred degree is a two-year college degree (38.5%), while for women, it is a four-year one (36.6%). Evidence from these two tables can indicate the proportion of individuals that might *potentially* react when introducing incentives to attend one sector in the higher education system. However, not all of them will respond in the incentivized direction. For example, although we know that the second most preferred ranking among men is first getting a high school diploma, followed by attaining a two-year degree, this fact does not necessarily translate into a scenario where providing a subsidy for a two-year college results in all these male high school graduates obtaining such a degree. Their reaction will depend on how well the subsidy can offset the differences in utility between these two options.

3.7.2 Treatment Effects

We are analyzing three postsecondary schooling alternatives: $\mathcal{J} = \{$HS, 2YC,4YC$\}$. This section will focus on analyzing two sets of treatment effect parameters:

$$\Delta^{MTE}_{\mathcal{K},\mathcal{L}}(x, u_{D_{\mathcal{K}}}) = E\left[\Delta_{\mathcal{K},\mathcal{L}} \mid X = x, C_{\mathcal{K}} = C_{\mathcal{L}}\right] \tag{3.21}$$

$$\Delta^{TT}_{\mathcal{K},\mathcal{L}}(x) = E\left[\Delta_{\mathcal{K},\mathcal{L}} \mid X = x, C_{\mathcal{K}} > C_{\mathcal{L}}\right]. \tag{3.22}$$

The first equation corresponds to the Marginal Treatment Effect, which was discussed in Sections 3.3, 3.4, and 3.5, and the second equation refers to the Treatment on the Treated (TT) effect. As a reminder, $\mathcal{K} = \{k\}$ and $\mathcal{L} = \mathcal{J} \backslash \{k\}$ is the choice set with all available option but k. We will focus on characterizing the MTE and TT parameters for each sector k, using the second-best sector as a reference point.

Tables 3.17 and 3.18 presented evidence of heterogeneous preferences. There is no single hierarchy that satisfies everyone's preferences. Therefore, for each set $\mathcal{K}$ considered, there will be individuals selecting each element in $\mathcal{J}$ as their second-best. For this reason, we consider the following further decomposition of the MTE (and their TT equivalencies):

$$\begin{aligned}\Delta^{MTE}_{4YC,\mathcal{L}}(x,u_{D_{4YC}}) =& \omega_{4YC}(x,u_{D_{4YC}})\times MTE_{4YC\leftarrow 2YC}(x,u_{D_{4YC}})\\ &+(1-\omega_{4YC}(x,u_{D_{4YC}}))\times MTE_{4YC\leftarrow HS}(x,u_{D_{4YC}}).\\ \Delta^{MTE}_{2YC,\mathcal{L}}(x,u_{D_{2YC}}) =& \omega_{2YC}(x,u_{D_{2YC}})\times MTE_{2YC\leftarrow 4YC}(x,u_{D_{2YC}})\\ &+(1-\omega_{2YC}(x,u_{D_{2YC}}))\times MTE_{2YC\leftarrow HS}(x,u_{D_{2YC}}).\end{aligned} \tag{3.23}$$

Finally, each MTE on the right-hand side of these equations will also be decomposed into observable and unobservable contributions, as follows:

$$\Delta^{MTE}(u_{D_{\mathcal{K}}}) = \beta_{X|q(P_{\mathcal{K}}(Z),u_{D_{\mathcal{K}}})\approx 0} + \beta_{u_D|q(P_{\mathcal{K}}(Z),U_{D_{\mathcal{K}}})\approx 0} \tag{3.24}$$

Marginal Treatment Effects

Figures 3.19-3.28 present our estimates of the Marginal Treatment Effects and their decomposition. We first analyze a set of results for men, followed by the results for women.

Figure 3.19 displays the Marginal Treatment Effects for men. Panel A considers the MTE for a two-year college education (k =2YC), and Panel B for a four-year college education (k =4YC). The results of estimating equation 3.21 for different values of unobservables are represented with a solid, black line. Note the unobservables considered for each sector vary. We use the difference between the unobserved component of the sector being analyzed relative to the other non-zero utility sector. Recall that the utility of being a high school graduate is normalized to zero to achieve identification. The other lines in this figure reflect the different components of the system of equations 3.23. The red and blue line shows the MTE considering specific inflows of individuals, each representing men with different second-best alternatives. Red shows the inflow of individuals who would move from the other college sector, while blue captures the effect for those who would no longer stay with a high school diploma. The dashed, gray line shows the weights, which are the fraction of individuals with the other college sector as the second-best alternative.

For calculating the MTEs, we use the absolute value of the difference of utilities to define the distance between sectors. We use a tolerance level ϵ of 0.25. Reducing the tolerance level ϵ to 0.1 generates similar but noisier results (see Section 3.3 for more details about the practical implementation). The smaller the ϵ, the smaller the set of simulated individuals used to calculate each point of the graphs. We only show estimates for points of indifference with at least 0.25% of the simulated sample. Even without imposing this restriction, there would be unobserved component values where the calculation of the MTE is unfeasible.

Panel A in Figure 3.19 examines the MTE results for a two-year college education. Therefore, we can interpret these results as those of a policy that would reduce the costs of attending two-year college, attracting individuals from the other two sectors into two-year college institutions. We can see that the overall MTE associated with two-year college policies (black line) is usually positive. Still, it can be negative, especially for values of unobservables, where more four-year graduates are likely to react to such policies (higher values of the differentiated unobservables). A higher weight is associated with a negative overall MTE because the red MTE–measuring the effect on four-year graduates who would decide to get a two-year college degree because of the policy–is negative (ranging from -\$7,013 to -\$19,835). In contrast, the red MTE, measuring the impact for high school graduates, is positive (ranging from \$9,560 to -\$238).

The results on Panel A in Figure 3.19 can be compared with those by Mountjoy (2022), who considers an IV estimation and MTE decomposition of reducing the distance to two-year college institutions. Figure 3.36 in the Appendix shows his equivalent estimates. However, we need to take some precautions before conducting any comparisons. His results should be multiplied by four, as his estimation uses quarterly earnings, while ours considers an annual earnings frequency. Moreover and most importantly, the scale on the x-axis on both graphs is incomparable. We can not translate our unobservable components scale into distance to college institutions, as we do not have access to this type of information. The MTE calculated from those who would choose to attain only a high school diploma in the absence of a policy is within the confidence intervals calculated by Mountjoy (2022), although his point estimates are smaller. MTE for the inflow of students who would switch from a four-year degree to a two-year degree is also negative but, in our case, larger in magnitude. Additionally, Mountjoy (2022) obtains a flat MTE while we estimate a downward-slopping MTE. Other previous studies find a downward slope and argue this is an expected pattern (e.g., Carneiro, Heckman and Vytlacil, 2011). The reason is that those with high values for unobservables have a low probability of selecting the analyzed sector; hence, they probably have little to gain. Those with low levels of unobservables already have a very high likelihood of participating; thus, those are likely the ones that can benefit the most from selecting that sector. However, we will see that we obtain a positive slope for women for the pairwise MTEs. As Mountjoy (2022) estimates come from pooling both groups, an effect that nets out across groups could explain the flat MTEs in his study while reconciling the differences with our estimates.

Panel B in Figure 3.19 shows the estimates for the MTE of a four-year college. Note the red MTE, which is the effect for individuals who would switch from attaining a two-year college degree to getting a four-year college degree ($4yg \leftarrow 2yg$), is very similar to the negative of the red MTE in Panel A but with an inverted x-axis, which analyzes the inflow from four-year college into two-year college ($2yg \leftarrow 4yg$). Ultimately, both calculate the average difference between college earnings for the same group: those indifferent between a four-year college degree and a two-year college degree. The only difference relies on what the second-best alternative is. As a consequence, we see that this MTE ranges from $20,580 to $7,056. The blue MTE, calculated for those who would switch from staying as high school graduates to getting a four-year college, also exhibits positive values that closely follow those switching from a two-year college education (red MTE18), ranging from $18,154 to $11,308. In fact, for high values of unobservables ($V_{4yg} - V_{2yg}$), which is precisely when there is a lower proportion of those indifferent between a high school diploma and a four-year college degree (gray dashed line), the MTE for the inflow of high school graduates is higher than the overall MTE (black line). As in the case of Panel A, all MTEs in Panel B also have a negative slope.

Figure 3.20 decomposes men's MTE for those who would switch to get a two-year college from a four-year college degree (2y ← 4y), applying equation 3.24. The graph on the left shows the MTE as a solid line and how much is explained by each component cumulatively in the bars. Dark blue shows the cumulative explanation of observed characteristics. Light blue shows the part of the MTE explained by cognitive abilities, red by socioemotional, and green by mechanical abilities. On the right, lines represent each component's isolated contribution, without adding up, using the same color code. These graphs show that the MTE observed for those switching from a four-year to a two-year college education is mainly explained by cognitive ability and observed characteristics. We can see that the slope observed in the MTE is inherited from cognitive ability. The cognitive component of the MTE even becomes positive for low levels of the unobservables—those who are more likely to participate. However, the component explained by observables, although almost flat, is negative (varying from -\$8,465 to -\$10,676), generating a total MTE that is also negative.

Figure 3.21 decomposes the MTE that would be relevant for policies targeting the two-year college sector and attracting those who, in the absence of the policy, would havestayed with a high school degree (2y ← HS for men). The decomposition shows a similar pattern to that observed in Figure 3.20. Observed characteristics and cognitive ability mainly explain the MTE. The cognitive ability provides the slope, while the effect driven by observed characteristics is mostly stable across different levels of unobservables (around \$5,000). The main difference relative to results shown in Figure 3.20 is that the MTE and the effect of observables are both positive.

We could further decompose the observed effect of the MTE into how much each observed characteristic explains. Given the functional form assumed for earnings, note that the coefficients are fixed. As a consequence of this assumption, any change in the contribution of each observed component is explained by a change in the composition of those at the margin of indifference for each evaluation point of unobservables. The same interpretation applies to changes in the contribution of abilities, as the loadings are also fixed. These changes in composition highlight the interplay between observable variables and unobservables.

Figure 3.22 shows each observed variable's contribution to the two-year college MTE for men. Panel A shows this decomposition for those indifferent between a two-year college degree and a four-year college degree (2y ← 4y). We can see that the main driver of levels is urban residency. The relative importance of the different characteristics changes, too. Urban residency and cohort effects are less relevant for higher values of unobservables, while region dummies and the number of children become more important. All these three variables have a negative contribution to the MTE. Panel B displays the decomposition for those indifferent between a two-year college degree and staying as a high school graduate (2y ← HS). For this case, the constant term provides the highest contribution by far. The relative contributions of different characteristics do not vary much, except for the region of residency effects that have a negative impact and increase relevance for high values of unobservables (less likelihood of participation), becoming more negative.

To wrap up the analysis of men, Figure 3.23 shows the MTE decomposition in terms of observed factors and abilities for the inflow of those switching from staying as high school graduates to attaining a four-year college degree (4y ← HS). The same pattern observed for the previous decompositions for men still holds. The level of the total MTE is mainly explained by observed characteristics that are relatively stable around $15,000, exhibiting a slight negative slope (from $16,226 to $14,419). The cognitive contribution mainly explains the MTEs slope.

Figures 3.24-3.28 show the MTE estimations for women. We will see some patterns that are common across most of the graphs. Firstly, the role of observed characteristics is stronger for women. Secondly, some MTEs have a positive slope that seems to be explained by the effect of observed characteristics. Lastly, abilities play a less critical role relative to men's estimations, but mechanical is the most important ability to explain female MTEs.

Figure 3.24 displays the Marginal Treatment Effects for women. Panel A shows the MTE for the two-year college sector (2YC), and Panel B for the four-year college sector (4YC). There is one outstanding difference relative to Figure 3.19, which includes the equivalent results for men. In the first place, we observe that women's MTEs do not have a pronounced downward slope. The aggregate MTE has a small but negative slope, similar to results found by Mountjoy (2022) who pooled men and women in one estimation. Women's results are closer to those obtained by Mountjoy (2022); our point estimates are within his confidence intervals and, in most cases, close to his point estimates. We can see that the MTEs that characterize the comparison between a four-year college degree and a two-year college degree (red lines in both panels) and between a high school degree and a two-year college degree (blue line) are upward-sloping. A change in the weights mainly explains the negative slope in the overall MTE.

Panel A in Figure 3.24 shows that the overall MTE of a two-year college education (black line) is positive and primarily stable at approximately \$2,500 for most of the unobservables support, approaching zero for high values of the unobservables (low probability of choosing a two-year college degree). It ranges from \$3,377 to \$838. The MTE for the inflow of those switching a high school diploma for a two-year college degree (blue line) is also relatively flat at \$8,000 (varying from \$7,780 to \$10,302). The MTE for the inflow of four-year graduates (red line) has an upward slope and is negative, ranging from -\$12,082 to -\$3,243. A negative MTE means that if we were to create an incentive for women who, in the absence of a policy, decide to get a four-year college degree to get a two-year college degree instead, they would get lower earnings on average than in the four-year college sector. The upward slope means that those with higher values of unobservables and hence with a low likelihood of selecting a two-year college degree would receive higher earnings in that sector (2YC), although still lower than those with a four-year college education, given the negative MTE. We will explore the drivers of this pattern in the upcoming, more detailed decompositions.

Panel B in Figure 3.24 displays the MTE for women and a four-year college education. The patterns of each curve are very similar: a slight downward slope for the total MTE, an almost flat MTE for the inflow of high school graduates, and an upward-sloping MTE for those who would switch their two-year college degrees. However, the levels are always positive and higher. The total MTE (black line) and the MTE for the inflow of female high school graduates (blue line) start above \$15,000 and fluctuate slightly around this value (to \$14,892 and to \$18,583, respectively). The MTE for those switching a two-year college degree (red line) ranges from \$5,043 to \$12,062.

Figure 3.25 decomposes the MTE for women switching from four-year to two-year college (2y ← 4y). We can see that there are two main contributors to this MTE: observed variables and mechanical ability, although the former plays a more important role. In contrast to men's decomposition, cognitive ability is not relevant anymore, and ability effects do not explain the slope. The positive slope in the MTE is driven by a positive slope observed for the observables' contribution to the MTE. Figure 3.26 shows the results from decomposing the MTE of the inflow of women switching a high school diploma for a two-year college degree. Observables and mechanical ability are also the main explanatory factors behind the estimated MTE, but cognitive ability contributes a small degree to the MTE for this inflow. The effect solely explained by the mechanical ability has a negative slope that ends up offset by the positive slope of the observables and cognitive effects.

Decomposing the MTE in terms of the contribution of each observable variable allows us to determine which specific variable explains the positive slope observed in previous figures. Figure 3.27 decomposes the MTE into observables for women in the two-year college sector. In Panel A, we have the results from decomposing the MTE resulting from the inflow of women changing a four-year college degree (2y ← 4y). The upward-sloping MTE is explained by an increased positive contribution of the variable counting the number of children when the level of unobservables increases, or in other words when the likelihood of choosing a two-year college degree over a four-year college decreases. To understand this result better, recall that the difference between earnings coefficients is always constant, as the coefficients are also fixed. For this particular case, the difference in earnings coefficients between these two sectors is positive ($-0.33 - (-0.69) = 0.36$). Therefore, since the contribution of the number of children has a positive contribution to the MTE and its related effect increases with high values of unobservables, we conclude that women who are less likely to choose a two-year college degree over a four-year college degree (high differences in unobservables) have more children than those with higher probability. For men, we observed an increased contribution of this variable, too, but negative, in favor of the earnings someone with a four-year college degree would receive. Urban residency continues to play an important negative role, and actually, its contribution is also reduced when the level of unobservables increases. However, we observed the same pattern for the role of urban residency for this men's pairwise MTE, which, in contrast, exhibited a negative slope.

Panel B in Figure 3.27 decomposes the MTE for women who would move from the high school sector to the two-year college sector into each observed component contribution. This MTE is almost flat with a slightly positive slope, and the contribution of different observable variables is mostly stable. However, the positive contribution of region dummies increases slightly for higher values of unobservables. Relative to men's decomposition, the constant term plays a lesser role, and the residency region's effects have a positive rather than negative contribution. Additionally, cohort dummies are not relevant factors explaining this MTE for women, but they were for men. For this margin of decision (2YC-HS), the number of children, a crucial factor in Panel A, does not explain the MTE effect.

Finally, Figure 3.28 shows that the MTE for women who would switch from the high school sector to attaining a four-year college degree, which has a slightly positive slope, is almost entirely explained by variation in observable components. Figure 3.38 in the Appendix shows a negative and decaying effect of the number of children, explaining the change in slope. Additional graphs in the Appendix show the decomposition into observables and abilities of other pairwise MTEs that we decided not to discuss because they are very similar by construction to reflections of the MTEs already described.

Effects of Treatment on the Treated

The following paragraphs will consider the effects of TT by quintiles of unobserved ability, showing substantial heterogeneity in returns. We will highlight which groups, depending on their ability endowments, would gain from switching their choice. However, it is interesting to contrast these results with those obtained when calculating MTEs. We will see that the TTs and MTEs will sometimes exhibit opposing signs for the same margin of comparison. These results suggest that those at the margin of indifference might not be the same with the specific ability endowments that would benefit from switching their schooling choice.

Figures 3.29-3.34 present the effects of treatment on the treated. Negative TTs could mean opportunities for intervening, as a negative TT indicates a person would have earned more in their second-best alternative than the chosen one. Nevertheless, negative TTs will not contain information on whether individuals who could get higher earnings in sectors rather than the chosen are likely to react to a policy affecting their utility. Their likelihood to respond will depend on whether these individuals are close to indifferent between sectors.

Figure 3.29 presents the effects of obtaining a four-year college education (4YC) for men. This group can be divided into two categories: those whose next-best alternative is a two-year college degree (Panel A) and those whose next-best option is a high school diploma (Panel B). Each group is divided into quintiles based on cognitive, socioemotional, and mechanical abilities.

We calculate the average earnings difference between the 4YC sector and the next-best alternative for each quintile pair, normalizing by the mean of the next-best alternative. The results are presented using heat maps, where warmer colors represent higher earnings. The rest of the effects of TT are displayed in the same fashion.

In Panel A of Figure 3.29, we analyze four-year graduates who would opt to get a two-year college degree if the option of obtaining a four-year college degree were not available. There is a high probability of obtaining a high TT, especially for those with high cognitive and socioemotional endowments but low mechanical endowments, where the average earnings can even be 144% higher. However, belonging to the first cognitive quintile has a negative TT, where those who are also in the highest quintile of mechanical abilities can earn 34% less with a four-year college degree than if they had obtained a two-year college degree. Belonging to the second cognitive quintile and the highest mechanical quintile can still generate earnings that are 9% lower in comparison.

We see a very different panorama in Panel B, where we consider those who, if obtaining a four-year college education were not an option, would choose to stay with a high school diploma. For all quintile pairs, men earn more as four-year graduates on average, with 11% to 86% higher earnings. However, as expected, the difference relative to the earnings obtained with a high school diploma grows larger with cognitive ability and smaller with mechanical ability. This result aligns with previous findings in the literature showing that those with higher mechanical and low cognitive endowments are more likely to have positive returns from staying as a high school graduate (Montoya-Agudelo, 2024; Prada and Urzúa, 2017).

Figure 3.30 shows the normalized TT for women pursuing a four-year college degree. Regardless of the next-best alternative, female four-year graduates earn, on average, higher than they would have if they had chosen other sectors. Some quintile pair estimations are shown in black, meaning there is no TT estimation for that specific group. The absence of a TT arises from the high correlation between mechanical and cognitive abilities for women, causing some quintile pairs to be unpopulated and making the TT calculation unfeasible. Relative to a two-year degree (Panel A), women have higher returns if they have high mechanical and cognitive endowments but low socioemotional ability. Their earnings can be as much as 83% higher. However, their relative returns are not as high as those for male four-year graduates shown in Figure 3.29.

Panel B shows the normalized TT effects for female four-year graduates who would have preferred to stay with a high school diploma if being a four-year graduate was not an option. As in the case of men, we can see that all of the effects are positive; however, for women, they are higher. These returns range from 52% to 166%. As in Panel A, the TT effects are higher for female four-year graduates with high socioemotional and lower cognitive ability. However, mechanical ability is irrelevant in determining returns relative to a high school diploma.

The TT effects for those choosing to attain a two-year college degree are shown in Figures 3.31 and 3.32 for men and women. Before interpreting the results, note that these effects are not just the negative of the TT effects displayed for the two previous figures, as the schooling decisions are different. Panel A in Figure 3.31 shows the TT effects for male two-year graduates relative to the earnings that would have been observed had they chosen a four-year college education instead. The effects are usually negative, meaning male two-year graduates usually would have fared better with a four-year degree. Some of them, however, do earn more with their actual choice: those with low cognitive and socioemotional abilities but high mechanical abilities. For these individuals, the return to a two-year college degree is 4%-33% relative to a four-year college degree.

Panel B in Figure 3.31 shows the TT effects for men with a two-year college degree, whose next-best would be stopping their schooling after graduating from high school. In contrast to Panel A, these effects are usually positive, although negative for some cases. Specifically, negative TT effects are observed for male two-year graduates with high socioemotional and cognitive abilities but low mechanical abilities. Therefore, if these individuals had stayed with a high school diploma instead, their earnings would have been higher. Their actual earnings can be as low as 21% lower.

Figure 3.32 shows the TT effects for women holding a two-year degree. Panel A reveals that women whose second-best option is obtaining a four-year college degree would have higher earnings if they received a four-year degree. Average returns are as low as -44%. Positive returns are only observed for those in the lowest cognitive and socioemotional ability quintiles, and the maximum average is just 14%. Panel B compares the earnings of a two-year degree relative to the earnings of a high school diploma for two-year graduates. All but one quintile pair group exhibit positive returns. They increase with lower ability endowments in all dimensions. These returns are also higher than those observed for men, reaching an average return of 132% for the lowest socioemotional and cognitive quintiles.

Figures 3.33 and 3.34 display the TT results for men and women without college degrees. Generally, we see negative returns except for a handful of men with a two-year college degree as their next-best alternative (Panel B in Figure 3.33), high cognitive and socioemotional abilities, and low mechanical abilities. For men whose second-best is getting a four-year college degree (Panel A in Figure 3.33), average returns do not vary much with abilities, ranging from -0.14 to -0.47. We see even lower returns for women with the same preferences (Panel A in Figure 3.34), with relative TT effects lower than 50% in many cases. For female high school graduates whose second best is getting a two-year college degree, we also see that returns are usually negative, with only a few cases at the top of the mechanical quintile showing positive but low returns. For this group, the lowest returns are observed with women with the lowest endowments in all three abilities.

3.8 Conclusions

While policies targeting postsecondary education can yield positive outcomes, understanding the heterogeneity of individual returns and responses is crucial for maximizing their impact. In this study, we focus on analyzing the effects of marginal treatment effects (MTE) of three higher education choices. We use a Roy model, accounting for unobserved factors influencing earnings and decisions, specifically cognitive, socioemotional, and mechanical abilities, all well-documented sources of heterogeneity in schooling choices. We propose a decomposition of the MTE, elucidating the respective contributions of observed and latent ability components. Additionally, we estimate the model separately for women and men, focusing on describing how responses and treatment effects vary between these two groups.

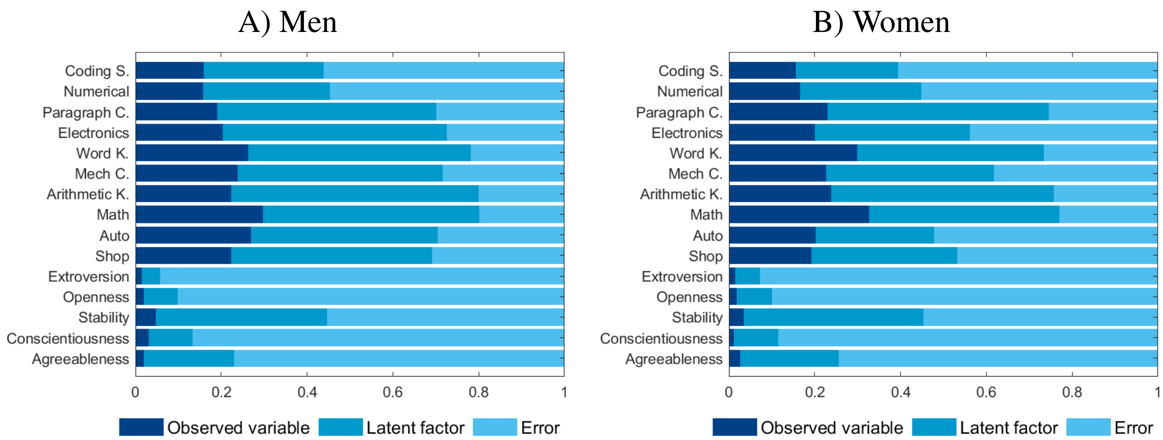
Our decomposition analysis shows that unobserved abilities are not always the primary factor that explains the level and slope of the MTE. For men, unobserved abilities are the main driver of a downward-sloping MTE pattern, although observables also play an important but constant role. On the other hand, for women, the primary driver is observables, with mechanical and cognitive abilities playing only a minor role. Additionally, women's pairwise MTEs exhibit an upward slope mainly driven by the observables' contribution.

A closer look at each observed variable's effect reveals that the contributions of variables such as the number of children and type or region of residency change through the support of unobservables. These changes in contributions are explained by a change in the composition of individuals who are indifferent when the level of unobservables changes to determine different indifference evaluation points. Therefore, this interaction between unobservables and observed variables determining earnings (X) arises from the fact that individuals who are indifferent at each evaluation point differ in terms of the observed variables that determine the educational choice (Z). These observed variables that determine choices also vary to satisfy the different points of indifference. Therefore, the relation between X and Z is key to understanding the interplay between observables and unobservables in shaping MTEs.

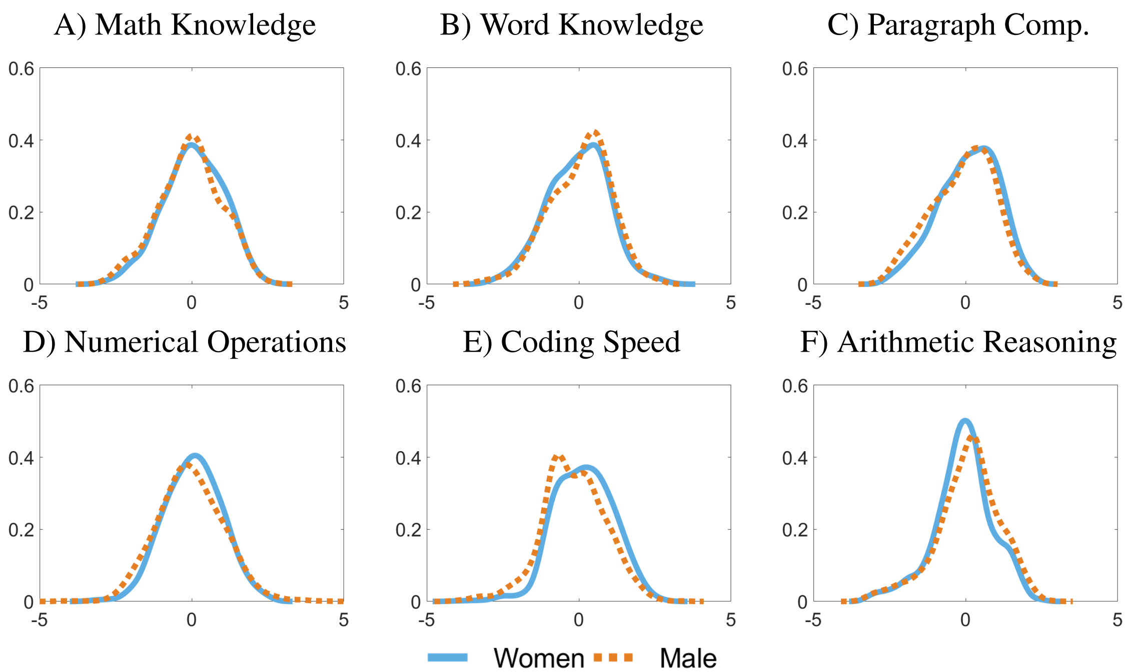
To conclude, we stress the importance of studying MTEs. When examining the treatment effects associated with attaining a two-year college degree, we find, aligned with previous literature, that the direction of the effect, whether positive or negative, depends on the alternative sector that an individual considers their second-best option. We observe that MTEs are positive for both women and men who would transition from staying with a high school diploma to pursuing a two-year college degree, although they are higher for women. Compared to the earnings potential associated with obtaining a four-year college degree, the MTEs for two-year college degrees are negative. On the other hand, we find evidence of gains in switching from a four-year college degree to a two-year college degree for both men and women with low cognitive abilities and high mechanical abilities when we examine the effects of treatment on the treated (TT). This fact contrasts with the negative MTE and highlights the importance of analyzing marginal effects when designing policy interventions: Although some individuals would benefit from attending a two-year institution instead of a four-year college, they are not necessarily the individuals who are at the margin of indifference. Those at the margin of indifference, who would react to minimal incentives, exhibit negative effects.

3.9 Figures and Tables

Figure 3.1: Variance Decomposition

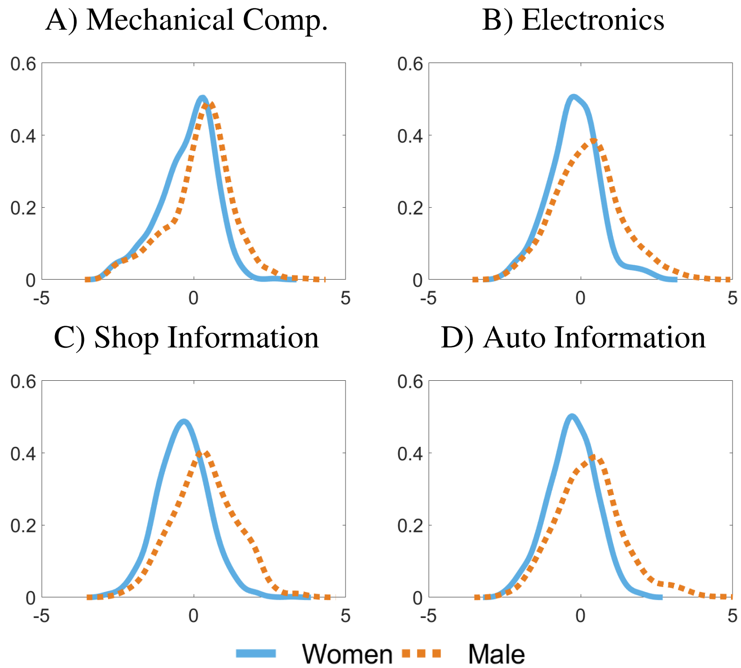


Note: The simulated data contains 1,000 replications of the sample. Coding S. = coding speed, Numerical = numerical operations, Paragraph C. = paragraph comprehension, Electronics = electronics information, Word K. = word knowledge, Mech C. = mechanical comprehension, Arithmetic K. = arithmetic reasoning, Math = math knowledge, Auto = auto information, Shop = shop information. The number of mixtures for each factor is based on Likelihood Ratio Tests comparing the fit in the measurement system. Cognitive factors have three mixtures, socioemotional factors have one mixture, and the uncorrelated components of the mechanical factors have two mixtures.



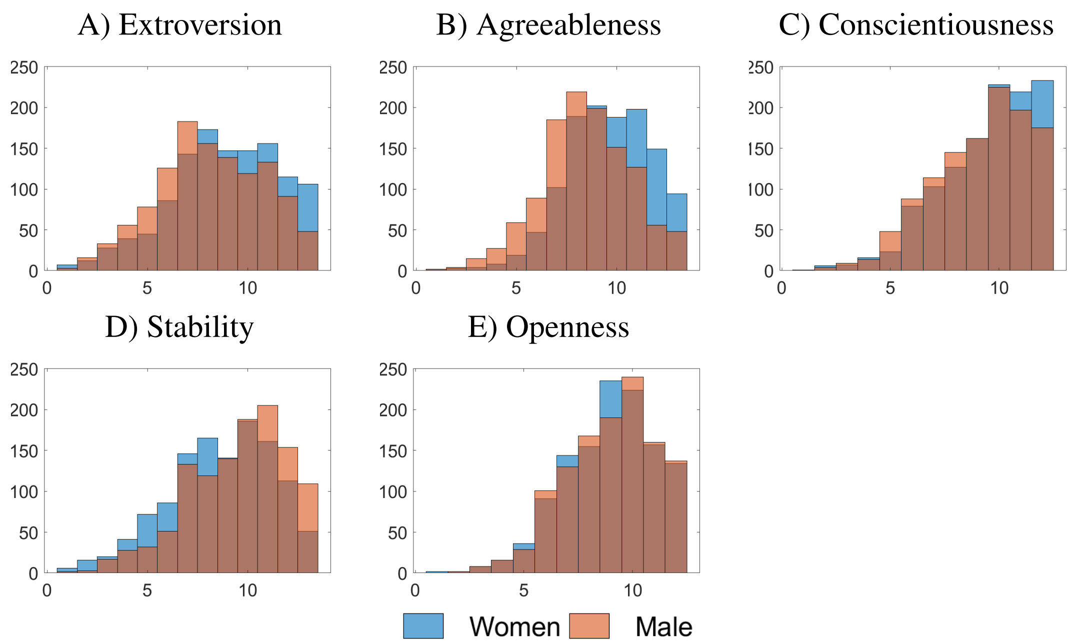
Note: Comparison between the densities of men's measures (orange, dashed line) and women's measures (blue, solid line). All measures are standardized in the pooled sample. Cognitive measures come from the ASVAB test, administered between 1997 and 1998. The simulated data contains 1,000 replications of the sample.

Figure 3.2: Raw Cognitive Measures



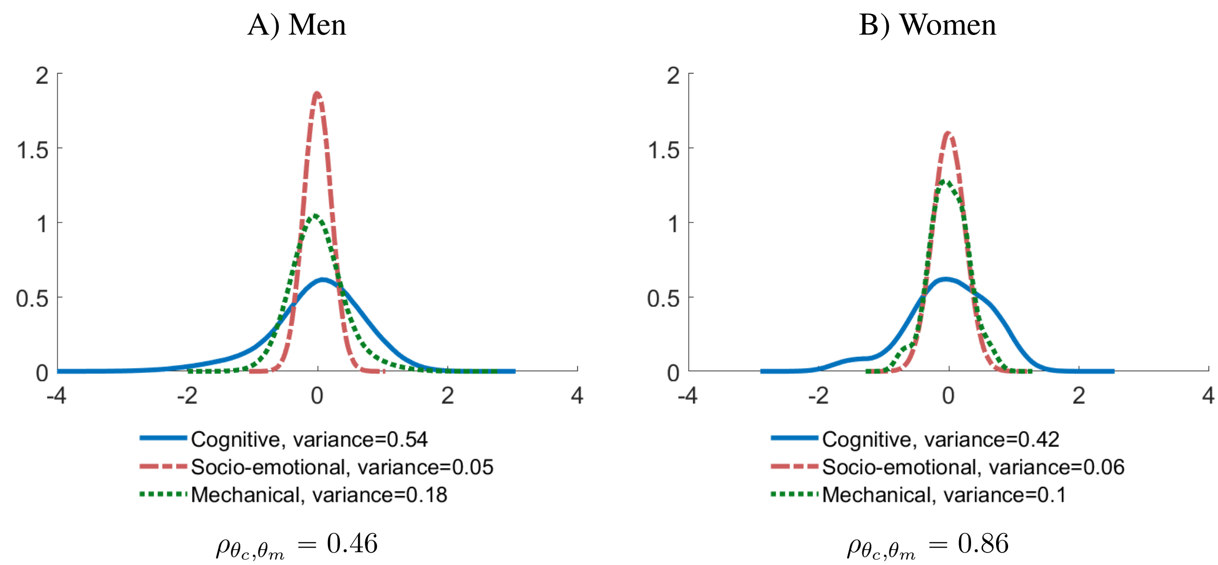
Note: Comparison between the densities of men's measures (orange, dashed line) and women's measures (blue, solid line). All measures are standardized in the pooled sample. Mechanical measures come from the ASVAB test, administered between 1997 and 1998. The simulated data contains 1,000 replications of the sample.

Figure 3.3: Raw Mechanical Measures



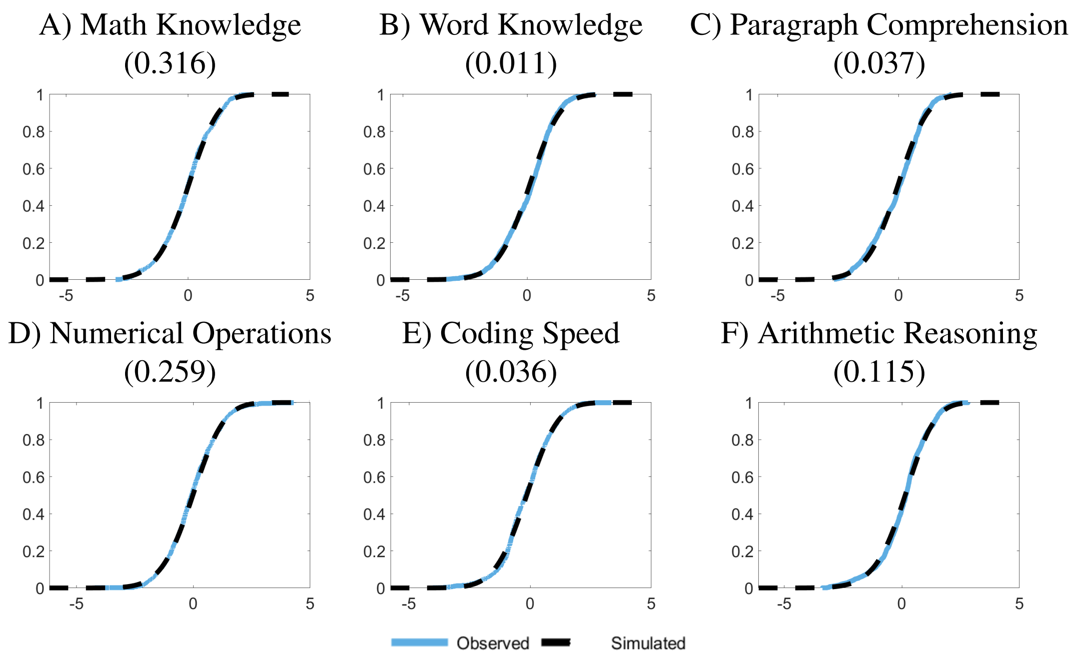
Note: Comparison between the densities of men's measures (orange bars) and women's measures (blue bars). All measures are standardized in the pooled sample. Socioemotional measures come from the TIPI scale, administered in 2008. Each big trait has two questions, which allows to control for acquiescence. The simulated data contains 1,000 replications of the sample.

Figure 3.4: Raw Socio-emotional Measures



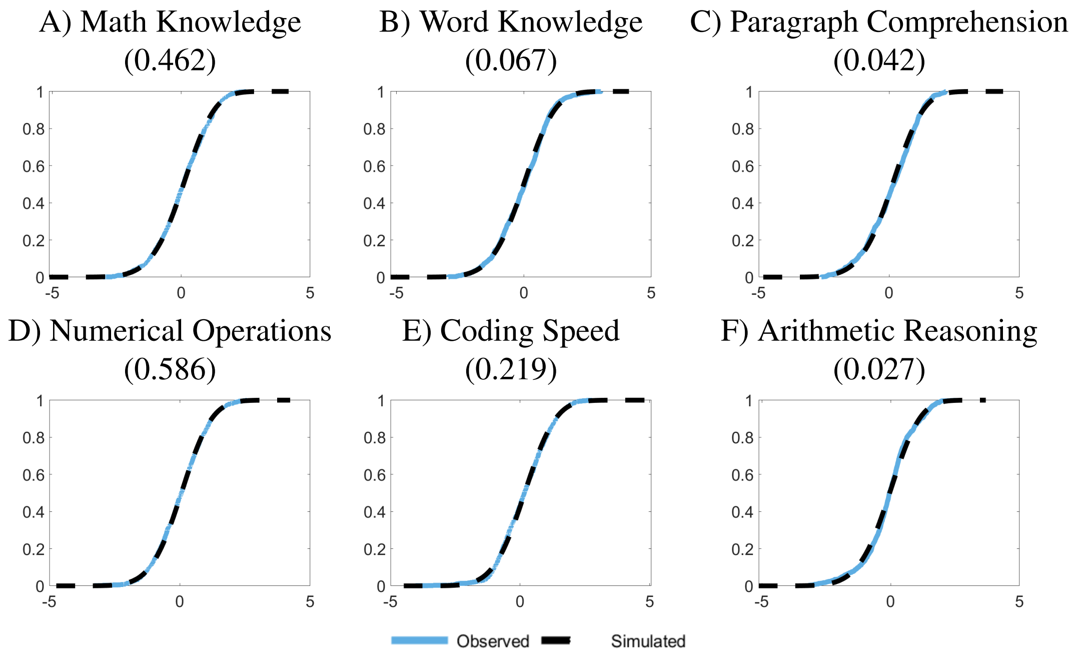
Note: The simulated data contains 1,000 replications of the sample. The number of mixtures for each factor is based on Likelihood-Ratio Tests comparing the fit in the measurement system. Cognitive factors have three mixtures, socioemotional factors have one mixture, and the uncorrelated components of the mechanical factors have two mixtures.

Figure 3.5: Latent Abilities' Marginal Densities.



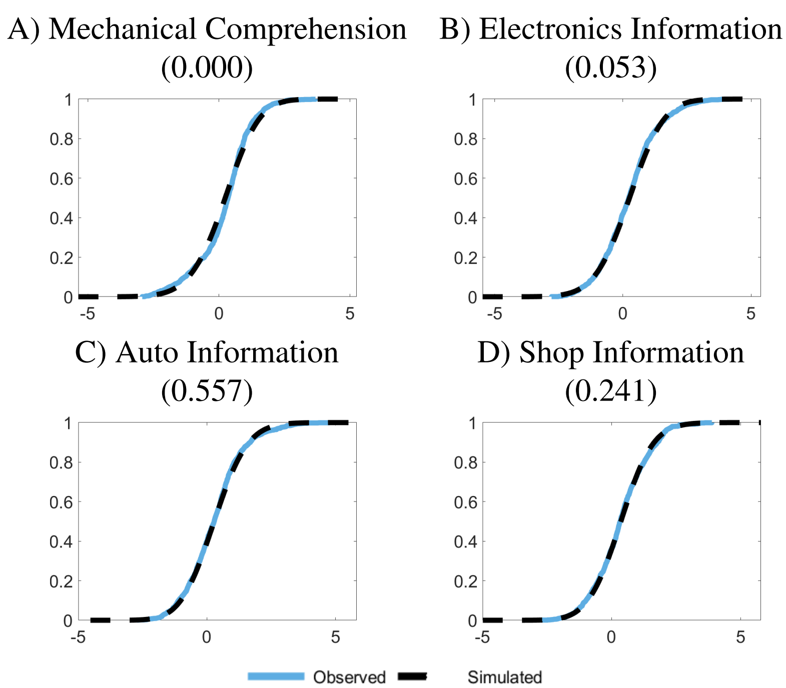
Note: Comparison between the CDF of the observed data (blue, solid line) and the CDF of the simulated data (black, dashed line). The p-value from a Kolmogorov–Smirnov test is shown in parenthesis at the top of each graph. The simulated data contains 1,000 replications of the sample.

Figure 3.6: Goodness of Fit: Cognitive System - Men.



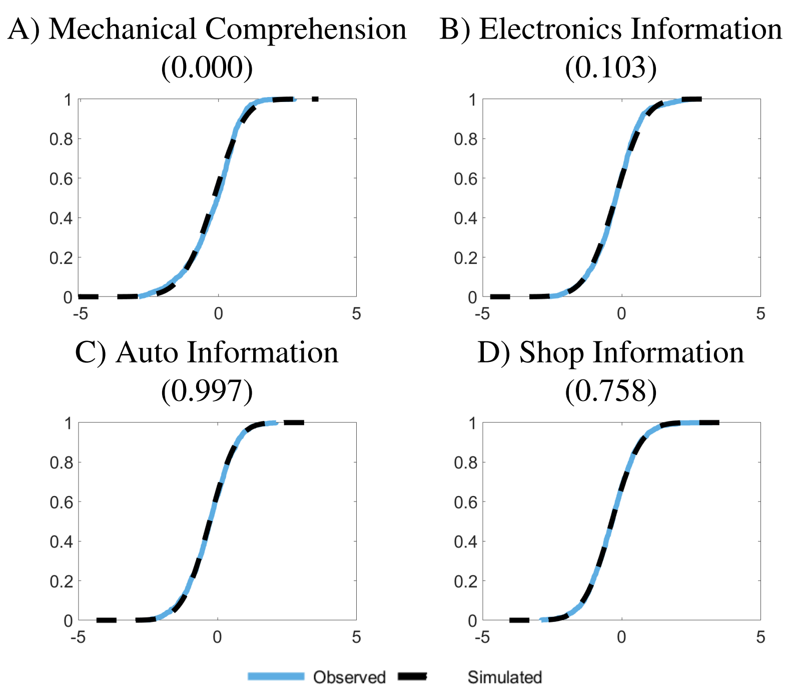
Note: Comparison between the CDF of the observed data (blue, solid line) and the CDF of the simulated data (black, dashed line). The p-value from a Kolmogorov–Smirnov test is shown in parenthesis at the top of each graph. The simulated data contains 1,000 replications of the sample.

Figure 3.7: Goodness of Fit: Cognitive System - Women.



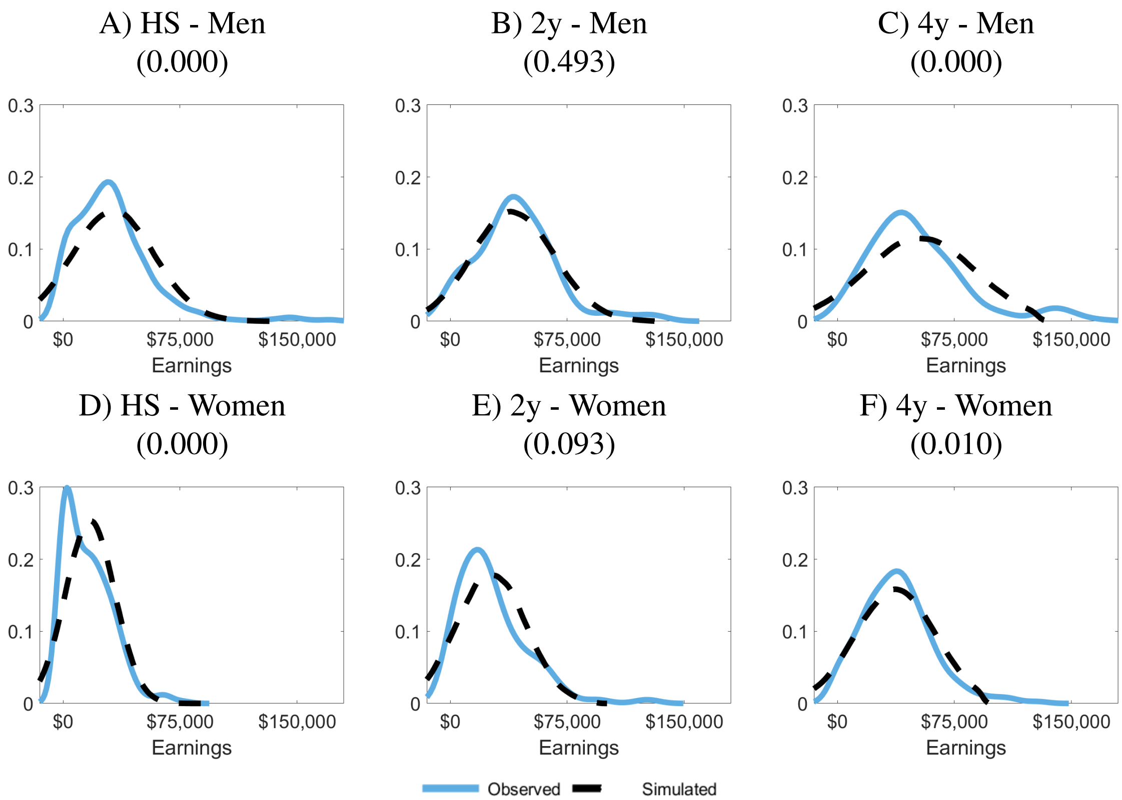
Note: Comparison between the CDF of the observed data (blue, solid line) and the CDF of the simulated data (black, dashed line). The p-value from a Kolmogorov–Smirnov test is shown in parenthesis at the top of each graph. The simulated data contains 1,000 replications of the sample.

Figure 3.8: Goodness of Fit: Mechanical System - Men.



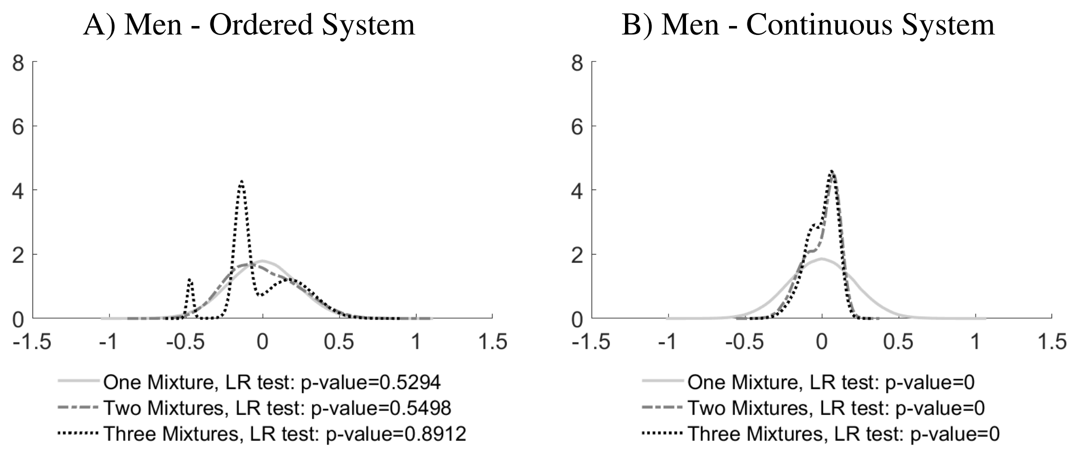
Note: Comparison between the CDF of the observed data (blue, solid line) and the CDF of the simulated data (black, dashed line). The p-value from a Kolmogorov–Smirnov test is shown in parenthesis at the top of each graph. The simulated data contains 1,000 replications of the sample.

Figure 3.9: Goodness of Fit: Mechanical System - Women.



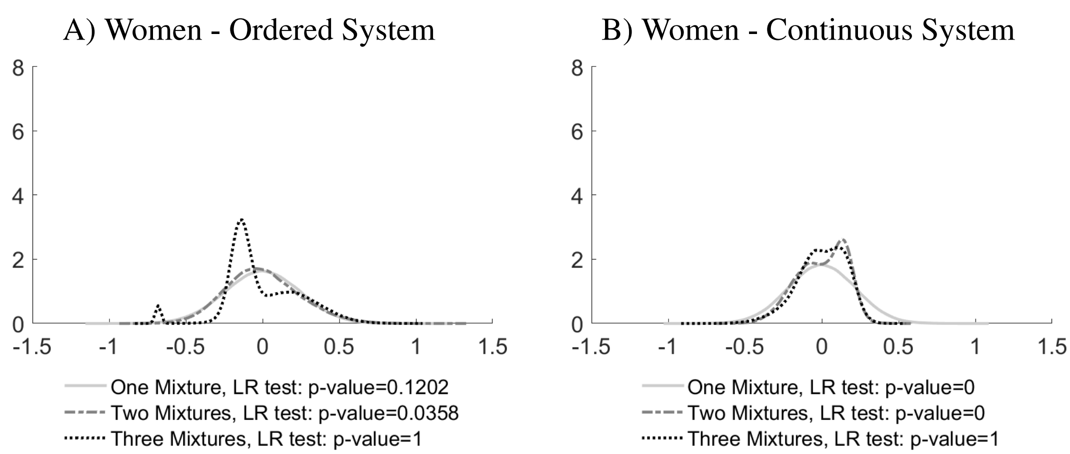
Note: Comparison between the densities of the observed data (blue, solid line) and the densities of the simulated data (black, dashed line). The p-value from a Kolmogorov–Smirnov test is shown in parenthesis at the top of each graph. Values in the bottom and top 0.5% of the unconditional distribution are not considered. The simulated data contains 1,000 replications of the sample.

Figure 3.10: Goodness of fit: Earnings



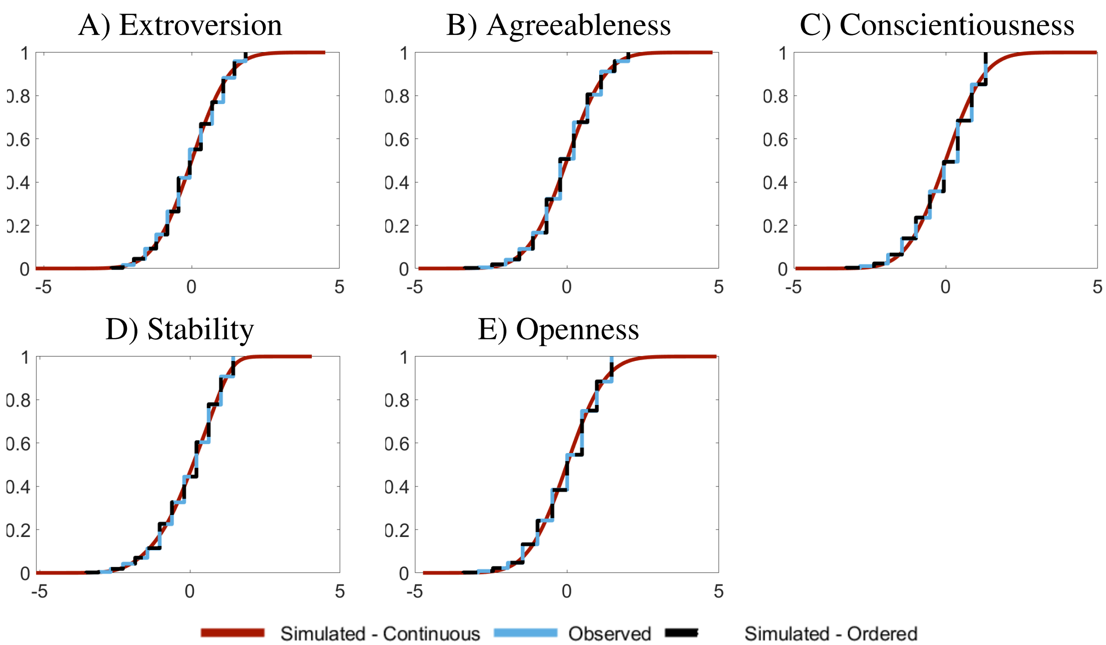
Note: Comparison of socioemotional factor distribution densities under different mixture assumptions (ordered and continuous). P-values from Likelihood-Ratio tests inform our choice of mixtures: one for the ordered system and four for the continuous system. Each p-value test considers the plotted assumption (restricted model) relative to a model where one additional mixture is considered. We consider a multi-start algorithm for each specification.

Figure 3.11: Socio-emotional ability distributions under different specifications - Men



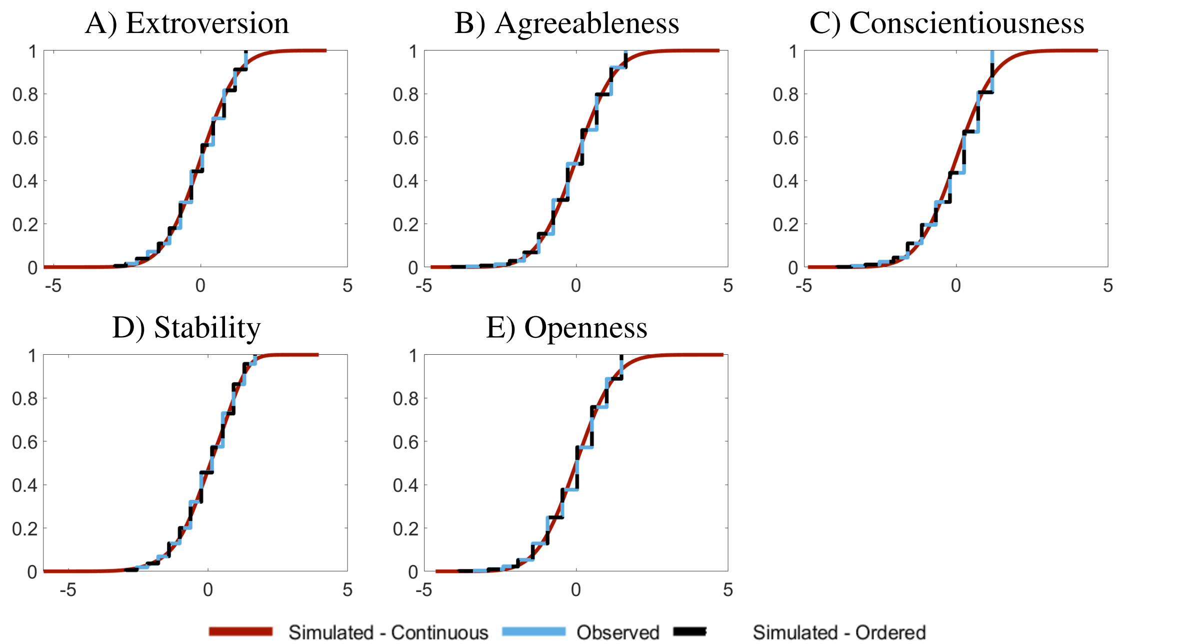
Note: Comparison of socioemotional factor distribution densities under different mixture assumptions (ordered and continuous). P-values from Likelihood-Ratio tests inform our choice of mixtures: one for the ordered system and three for the continuous system. Each p-value test considers the plotted assumption (restricted model) relative to a model where one additional mixture is considered. We consider a multi-start algorithm for each specification.

Figure 3.12: Socio-emotional ability distributions under different specifications - Women



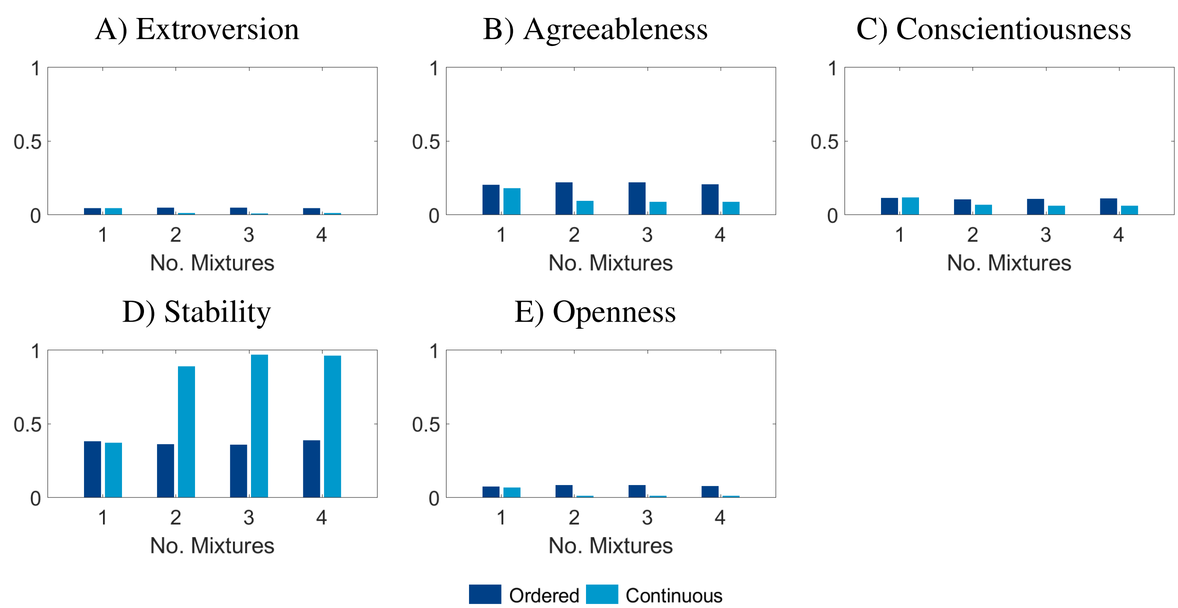
Note: Comparison between the CDFs of the observed data (blue, solid line), the CDFs of the simulated data assuming an ordered system (black, dashed line) and a continuous system (red, solid line) for men. Ordered system assumes one normal mixture for the socioemotional factor, while the continuous measure uses four mixtures. The simulated data contains 1,000 replications of the sample.

Figure 3.13: Predicted socio-emotional measures under different specifications - Men



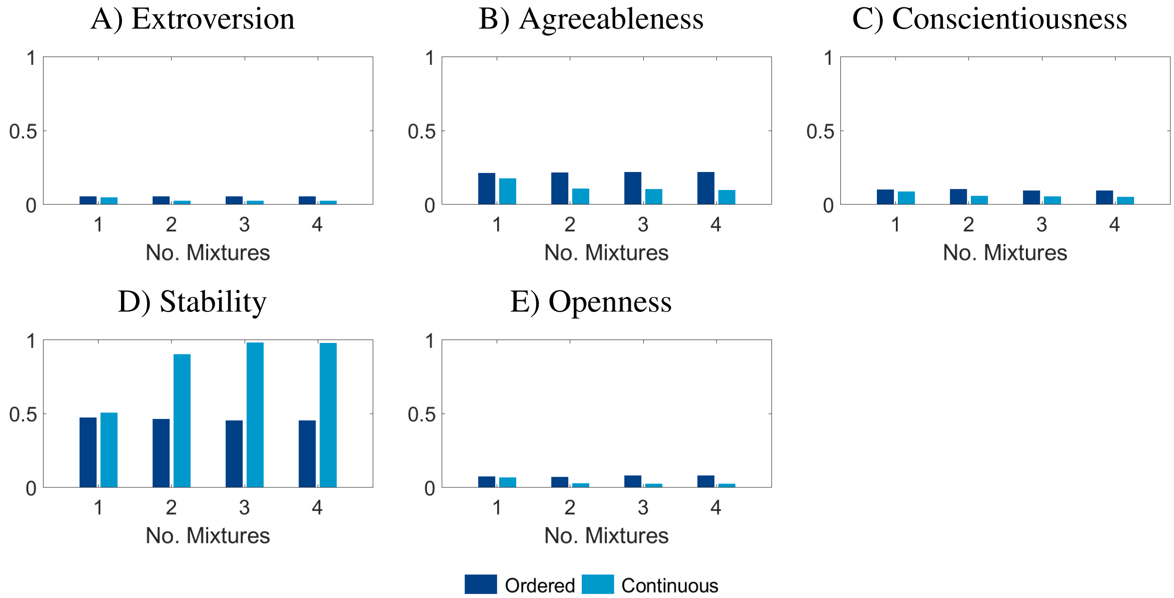
Note: Comparison between the CDFs of the observed data (blue, solid line), the CDFs of the simulated data assuming an ordered system (black, dashed line) and a continuous system (red, solid line) for men. Ordered system assumes one normal mixture for the socioemotional factor, while the continuous measure uses three mixtures. The simulated data contains 1,000 replications of the sample.

Figure 3.14: Predicted socio-emotional measures under different specifications - Women



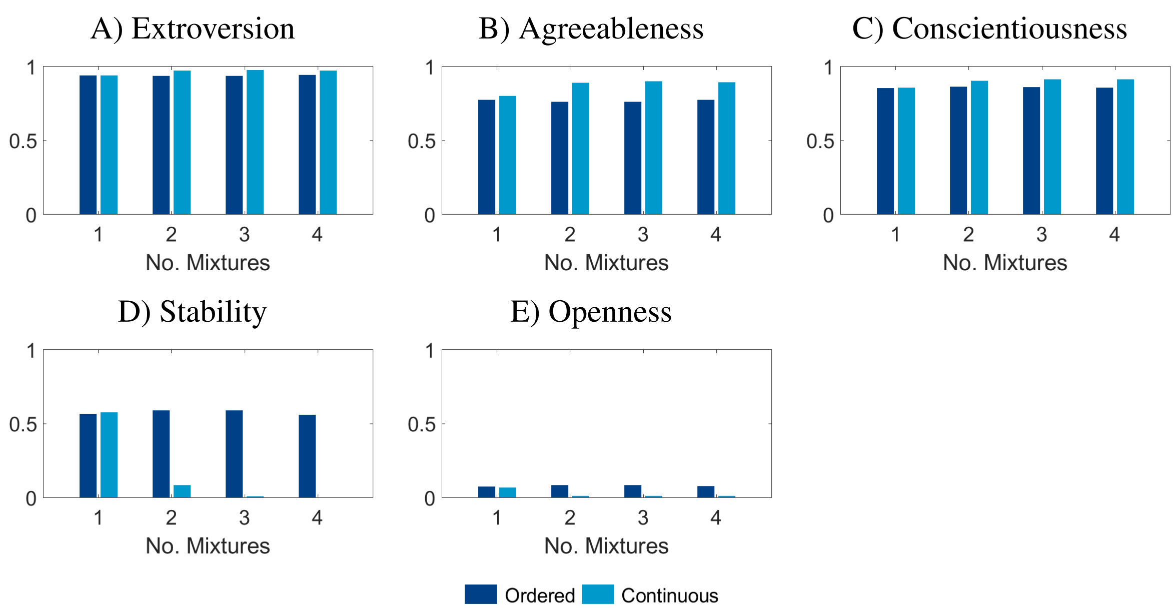
Note: We present what fraction of the variance of each measure is explained by the variance of the socioemotional factor. Darker blue represents the specifications where we consider the ordered nature of the measures, while light blue represents those assuming continuous measures.

Figure 3.15: Ability's contribution to socio-emotional measure by specification - Men



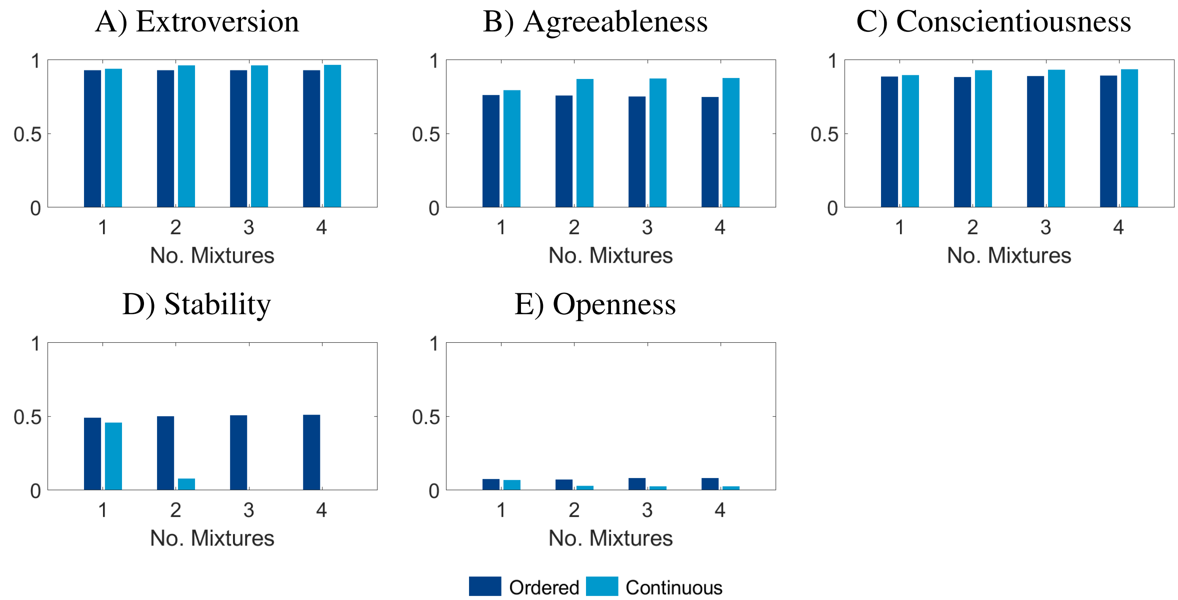
Note: We present what fraction of the variance of each measure is explained by the variance of the socioemotional factor. Darker blue represents the specifications where we consider the ordered nature of the measures, while light blue represents those assuming continuous measures.

Figure 3.16: Ability's contribution to socio-emotional measure by specification - Women



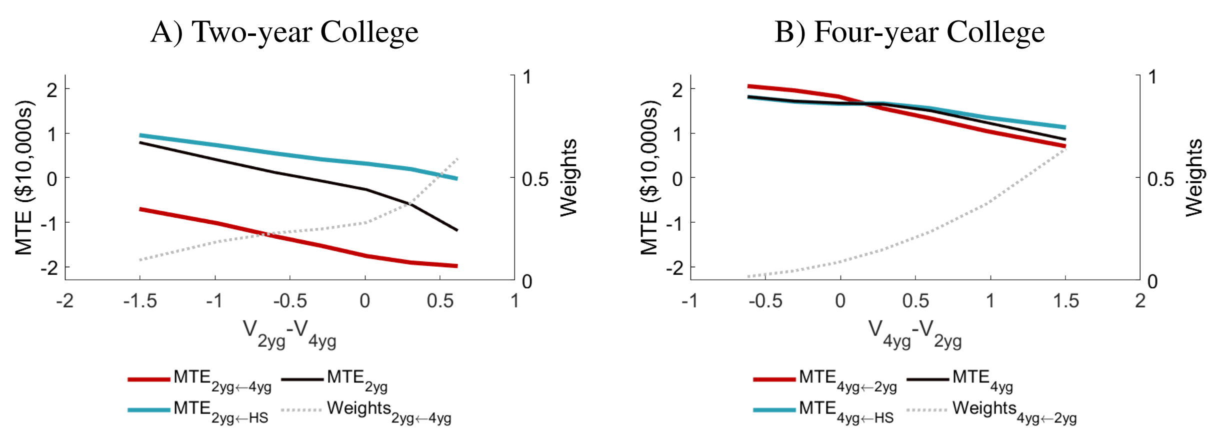
Note: We present what fraction of the variance of each measure is explained by the error term. Darker blue represents the specifications where we consider the ordered nature of the measures, while light blue represents those assuming continuous measures.

Figure 3.17: Error term's variance by socio-emotional measure and model specification - Men



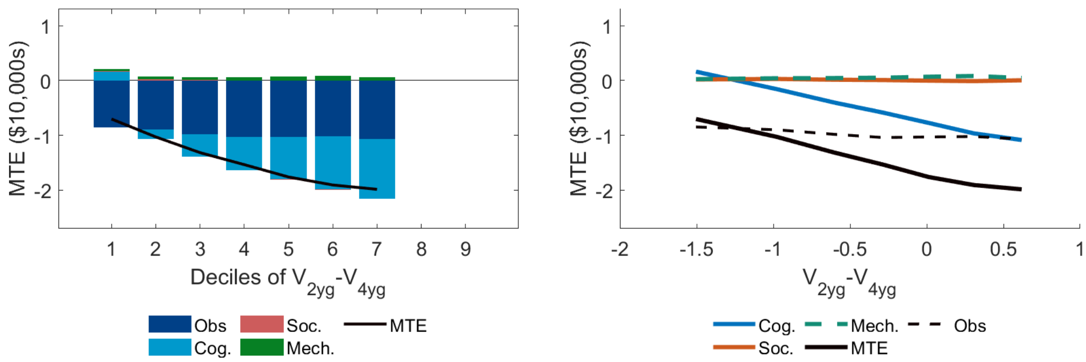
Note: We present what fraction of the variance of each measure is explained by the error term. Darker blue represents the specifications where we consider the ordered nature of the measures, while light blue represents those assuming continuous measures.

Figure 3.18: Error term's variance by socio-emotional measure and model specification - Women



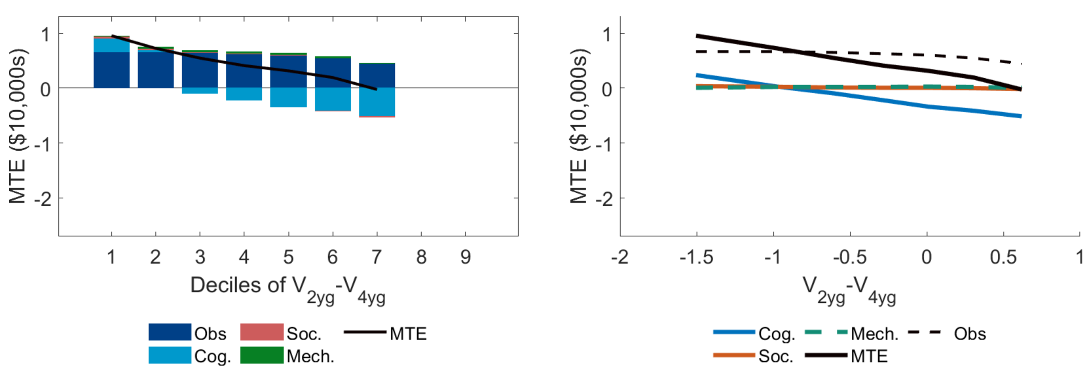
Note: The simulated data contains 1,000 replications of the sample. We use $|C_j - C_k| < \epsilon$ to determine indifference, where $\epsilon = 0.25$. Results are similar but noisier considering $\epsilon = 0.1$. Calculations are displayed for points with at least 0.25% of the simulated sample. This threshold does not change the calculations for populated-enough grids, but prevent outlier-driven evaluations. Even without this restrictions, we have some grid points without any simulated individual in the indifference margin.

Figure 3.19: Marginal Treatment Effect for Four- and Two-year College - Men



Note: The simulated data contains 1,000 replications of the sample. We use $|C_j - C_k| < \epsilon$ to determine indifference, where $\epsilon = 0.25$. Results are similar but noisier considering $\epsilon = 0.1$. Calculations are displayed for points with at least 0.25% of the simulated sample. This threshold does not change the calculations for populated-enough grids, but prevent outlier-driven evaluations. Even without this restrictions, we have some grid points without any simulated individual in the indifference margin.

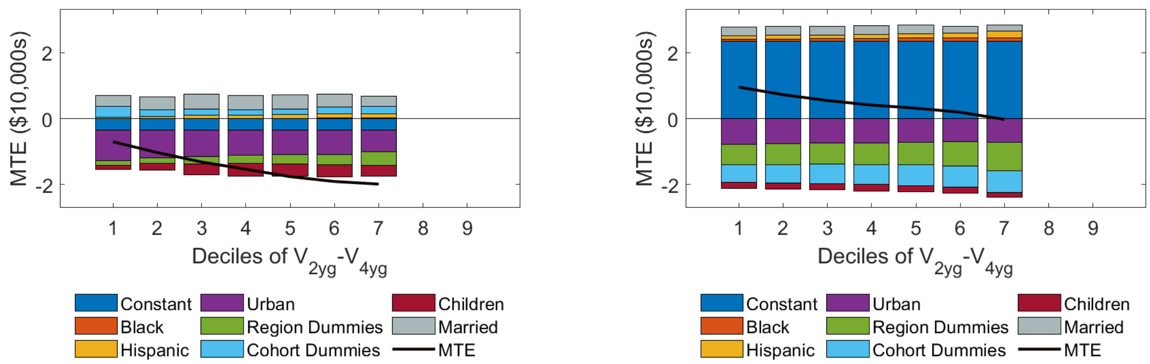
Figure 3.20: MTE Decomposition of Four-year Graduates Inflow into Two-year College - Men



Note: The simulated data contains 1,000 replications of the sample. We use $|C_j - C_k| < \epsilon$ to determine indifference, where $\epsilon = 0.25$. Results are similar but noisier considering $\epsilon = 0.1$. Calculations are displayed for points with at least 0.25% of the simulated sample. This threshold does not change the calculations for populated-enough grids, but prevent outlier-driven evaluations. Even without this restrictions, we have some grid points without any simulated individual in the indifference margin.

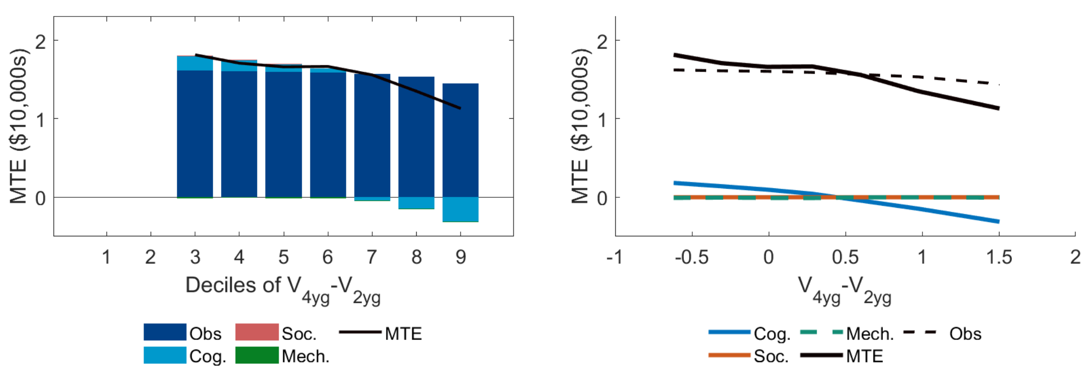
Figure 3.21: MTE Decomposition of High School Graduates Inflow into Two-year College - Men

A) Two-year College ← Four-year College B) Two-year College ← High School



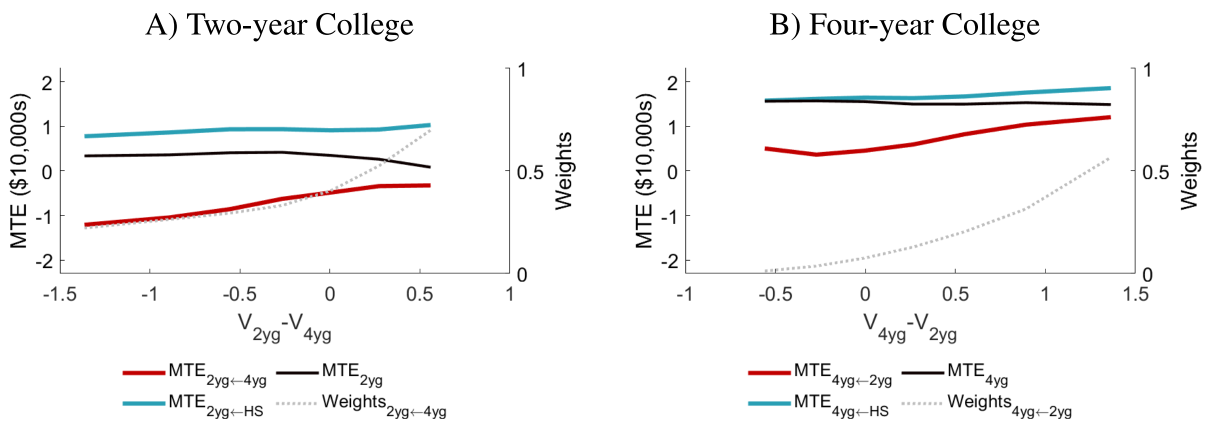
Note: The simulated data contains 1,000 replications of the sample. We use $|C_j - C_k| < \epsilon$ to determine indifference, where $\epsilon = 0.25$. Results are similar but noisier considering $\epsilon = 0.1$. Calculations are displayed for points with at least 0.25% of the simulated sample. This threshold does not change the calculations for populated-enough grids, but prevent outlier-driven evaluations. Even without this restrictions, we have some grid points without any simulated individual in the indifference margin.

Figure 3.22: MTE Decomposition of Observables for Two-year College - Men



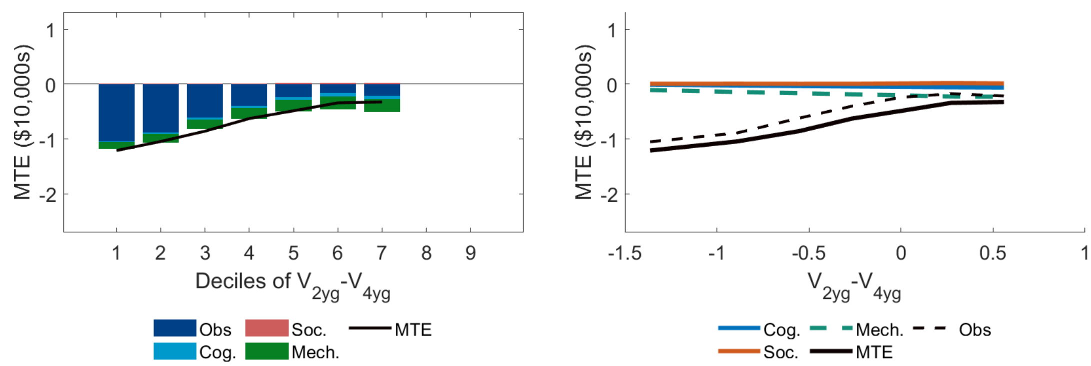
Note: The simulated data contains 1,000 replications of the sample. We use $|C_j - C_k| < \epsilon$ to determine indifference, where $\epsilon = 0.25$. Results are similar but noisier considering $\epsilon = 0.1$. Calculations are displayed for points with at least 0.25% of the simulated sample. This threshold does not change the calculations for populated-enough grids, but prevent outlier-driven evaluations. Even without this restrictions, we have some grid points without any simulated individual in the indifference margin.

Figure 3.23: MTE Decomposition of High School Graduates Inflow into Four-year College - Men



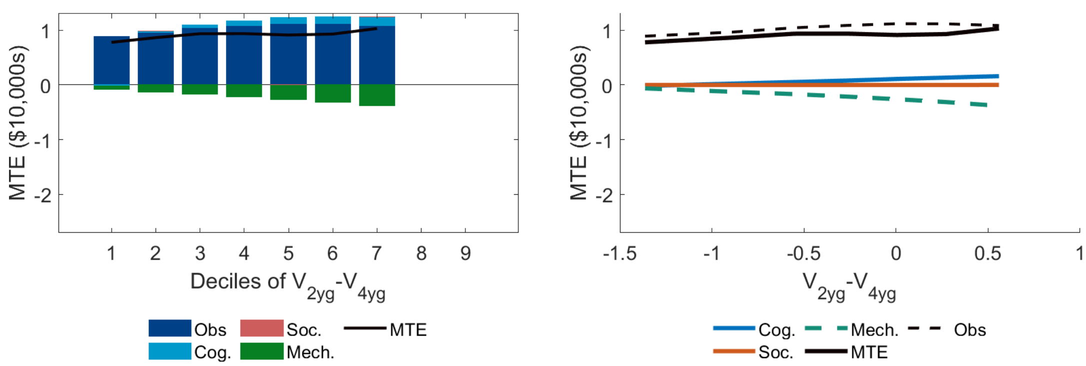
Note: The simulated data contains 1,000 replications of the sample. We use $|C_j - C_k| < \epsilon$ to determine indifference, where $\epsilon = 0.25$. Results are similar but noisier considering $\epsilon = 0.1$. Calculations are displayed for points with at least 0.25% of the simulated sample. This threshold does not change the calculations for populated-enough grids, but prevent outlier-driven evaluations. Even without this restrictions, we have some grid points without any simulated individual in the indifference margin.

Figure 3.24: Marginal Treatment Effect for Four- and Two-year College - Women



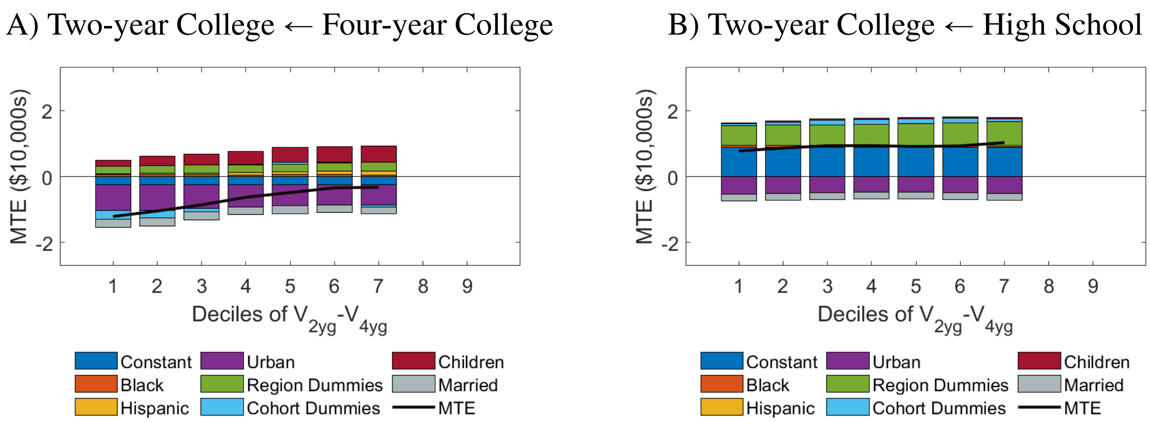
Note: The simulated data contains 1,000 replications of the sample. We use $|C_j - C_k| < \epsilon$ to determine indifference, where $\epsilon = 0.25$. Results are similar but noisier considering $\epsilon = 0.1$. Calculations are displayed for points with at least 0.25% of the simulated sample. This threshold does not change the calculations for populated-enough grids, but prevent outlier-driven evaluations. Even without this restrictions, we have some grid points without any simulated individual in the indifference margin.

Figure 3.25: MTE Decomposition of Four-year Graduates Inflow into Two-year College - Women



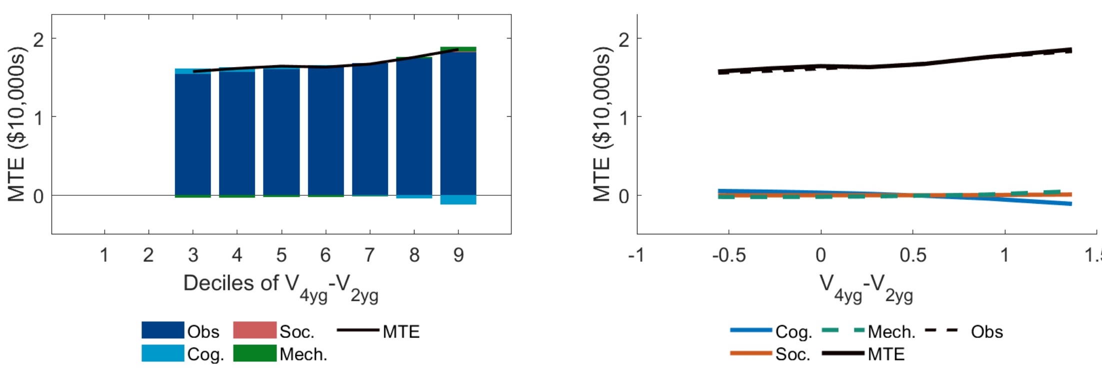
Note: The simulated data contains 1,000 replications of the sample. We use $|C_j - C_k| < \epsilon$ to determine indifference, where $\epsilon = 0.25$. Results are similar but noisier considering $\epsilon = 0.1$. Calculations are displayed for points with at least 0.25% of the simulated sample. This threshold does not change the calculations for populated-enough grids, but prevent outlier-driven evaluations. Even without this restrictions, we have some grid points without any simulated individual in the indifference margin.

Figure 3.26: MTE Decomposition of High School Graduates Inflow into Two-year College - Women



Note: The simulated data contains 1,000 replications of the sample. We use $|C_j - C_k| < \epsilon$ to determine indifference, where $\epsilon = 0.25$. Results are similar but noisier considering $\epsilon = 0.1$. Calculations are displayed for points with at least 0.25% of the simulated sample. This threshold does not change the calculations for populated-enough grids, but prevent outlier-driven evaluations. Even without this restrictions, we have some grid points without any simulated individual in the indifference margin.

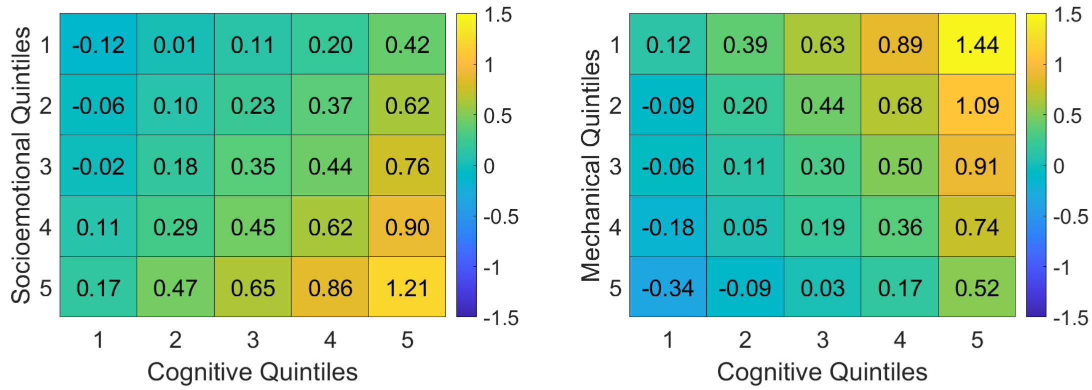
Figure 3.27: MTE Decomposition of Observables for Two-year College - Women



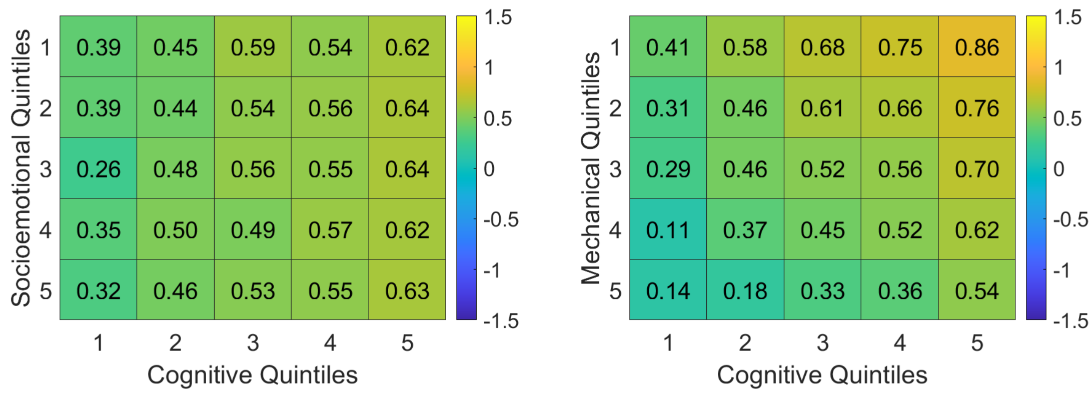
Note: The simulated data contains 1,000 replications of the sample. We use $|C_j - C_k| < \epsilon$ to determine indifference, where $\epsilon = 0.25$. Results are similar but noisier considering $\epsilon = 0.1$. Calculations are displayed for points with at least 0.25% of the simulated sample. This threshold does not change the calculations for populated-enough grids, but prevent outlier-driven evaluations. Even without this restrictions, we have some grid points without any simulated individual in the indifference margin.

Figure 3.28: MTE Decomposition of High School Graduates Inflow into Four-year College - Women

A) Four-year College Relative to Two-year College

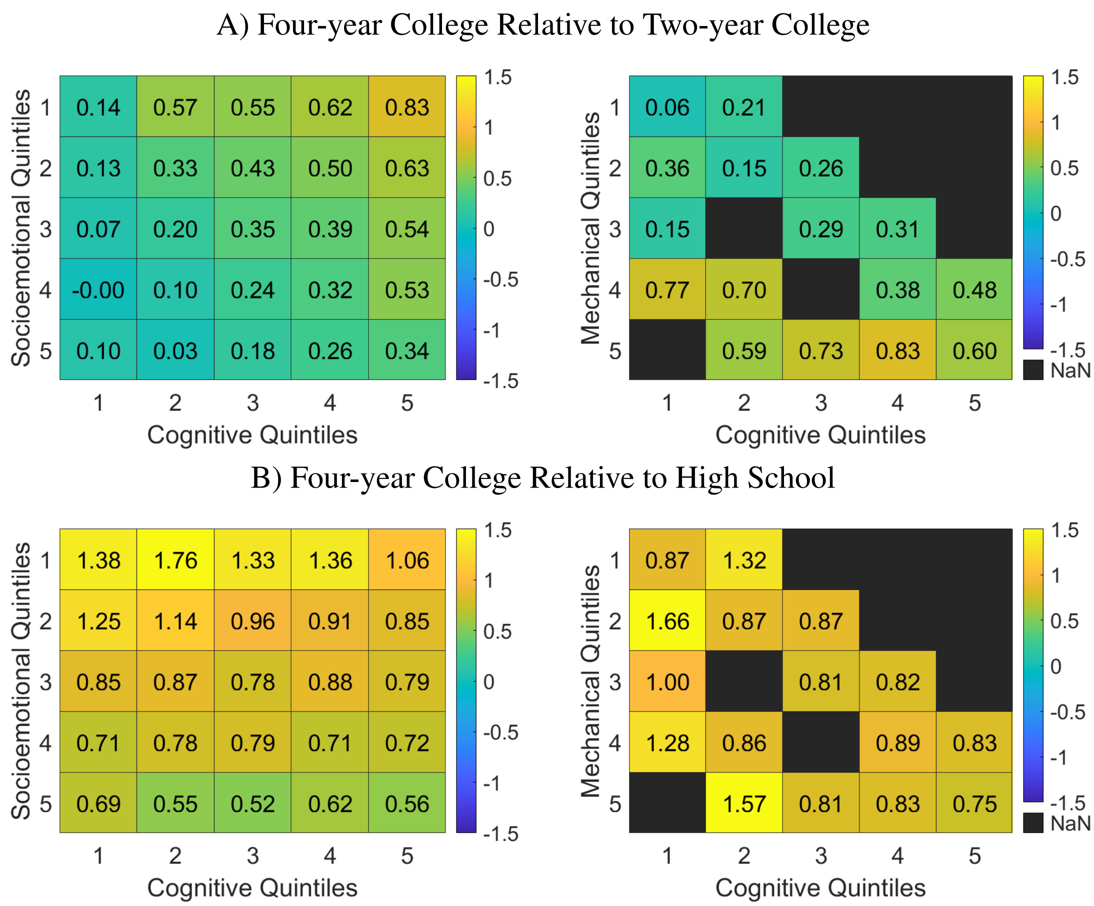


B) Four-year College Relative to High School



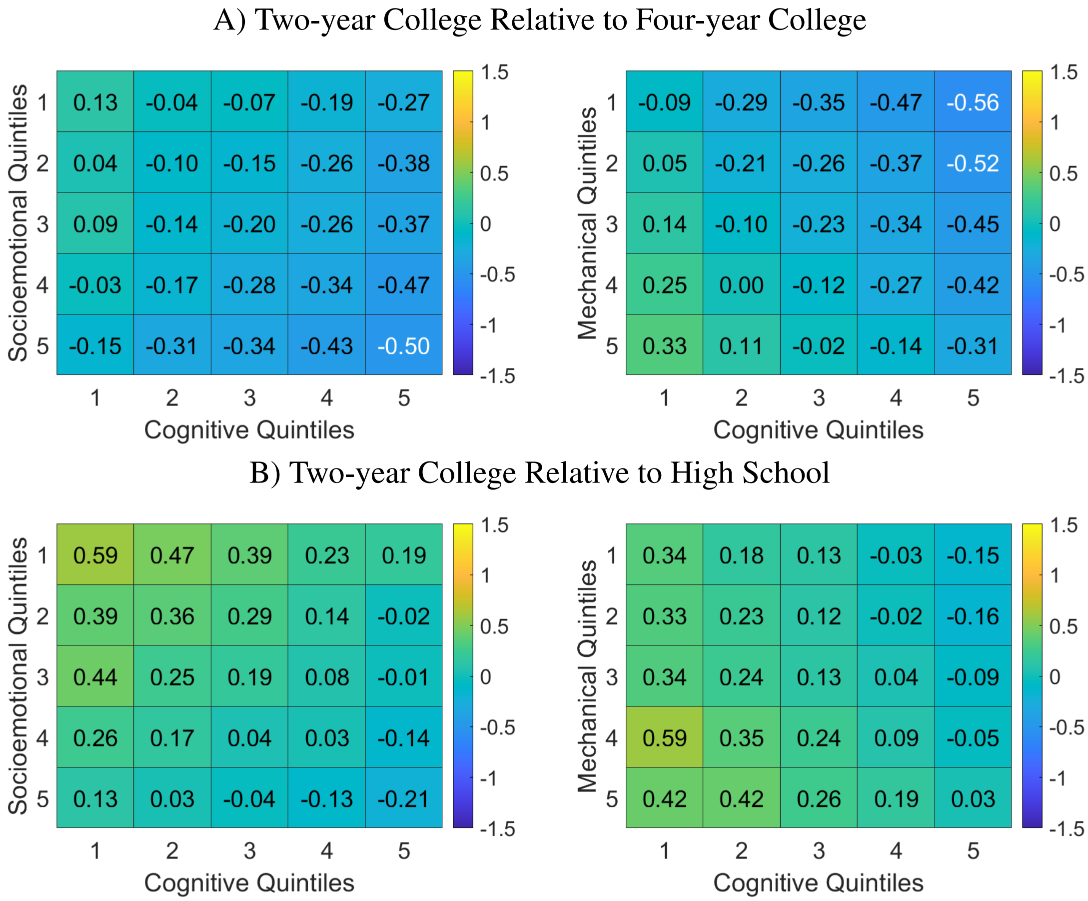
Note: The simulated data contains 1,000 replications of the sample. We compute the average treatment effect relative to the next-best alternative, normalized by the mean of the next-best alternative at each grid point

Figure 3.29: Treatment on the Treated for Four-Year Graduates - Men



Note: The simulated data contains 1,000 replications of the sample. We compute the average treatment effect relative to the next-best alternative, normalized by the mean of the next-best alternative at each grid point

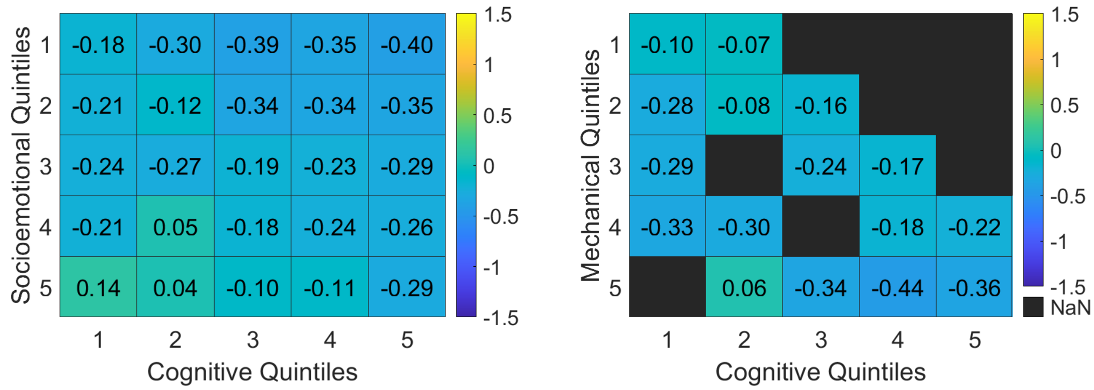
Figure 3.30: Treatment on the Treated for Four-Year Graduates - Women



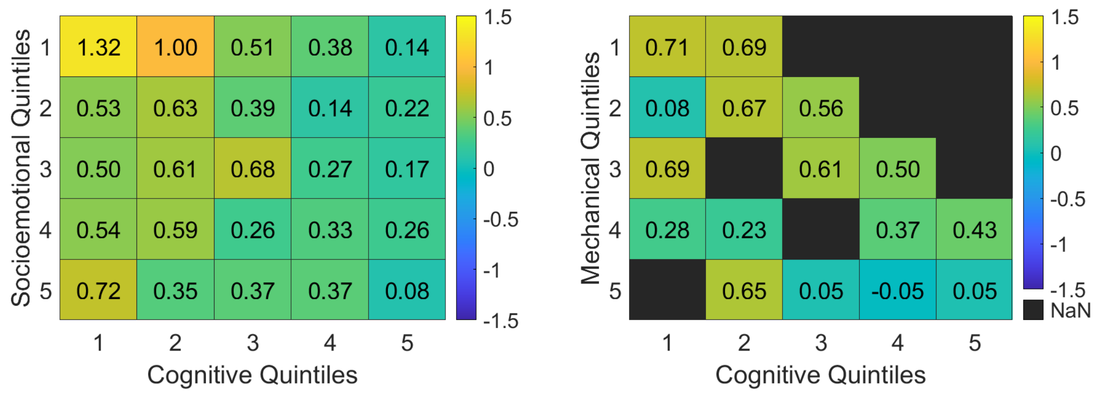
Note: The simulated data contains 1,000 replications of the sample. We compute the average treatment effect relative to the next-best alternative, normalized by the mean of the next-best alternative at each grid point

Figure 3.31: Treatment on the Treated for Two-Year Graduates - Men

A) Two-year College Relative to Four-year College



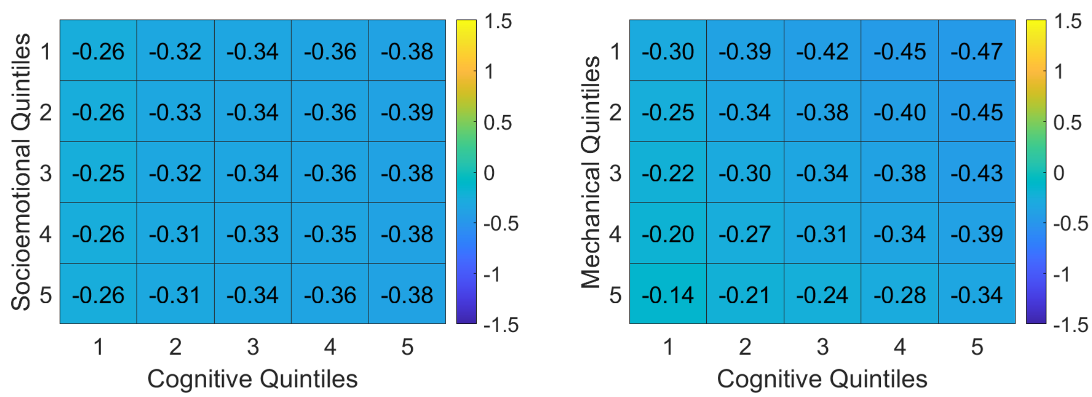
B) Two-year College Relative to High School



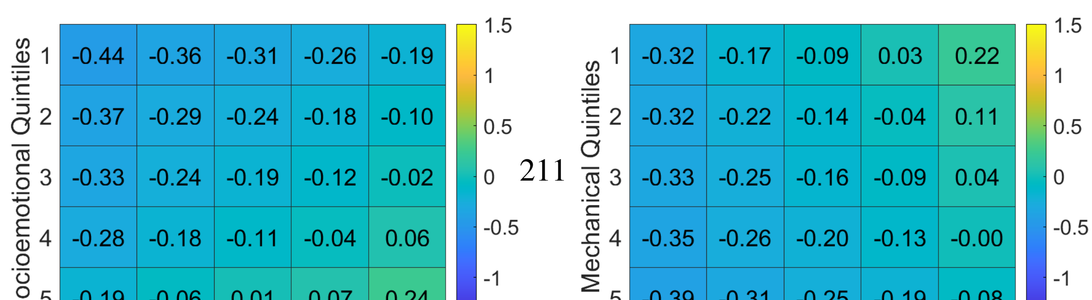
Note: The simulated data contains 1,000 replications of the sample. We compute the average treatment effect relative to the next-best alternative, normalized by the mean of the next-best alternative at each grid point

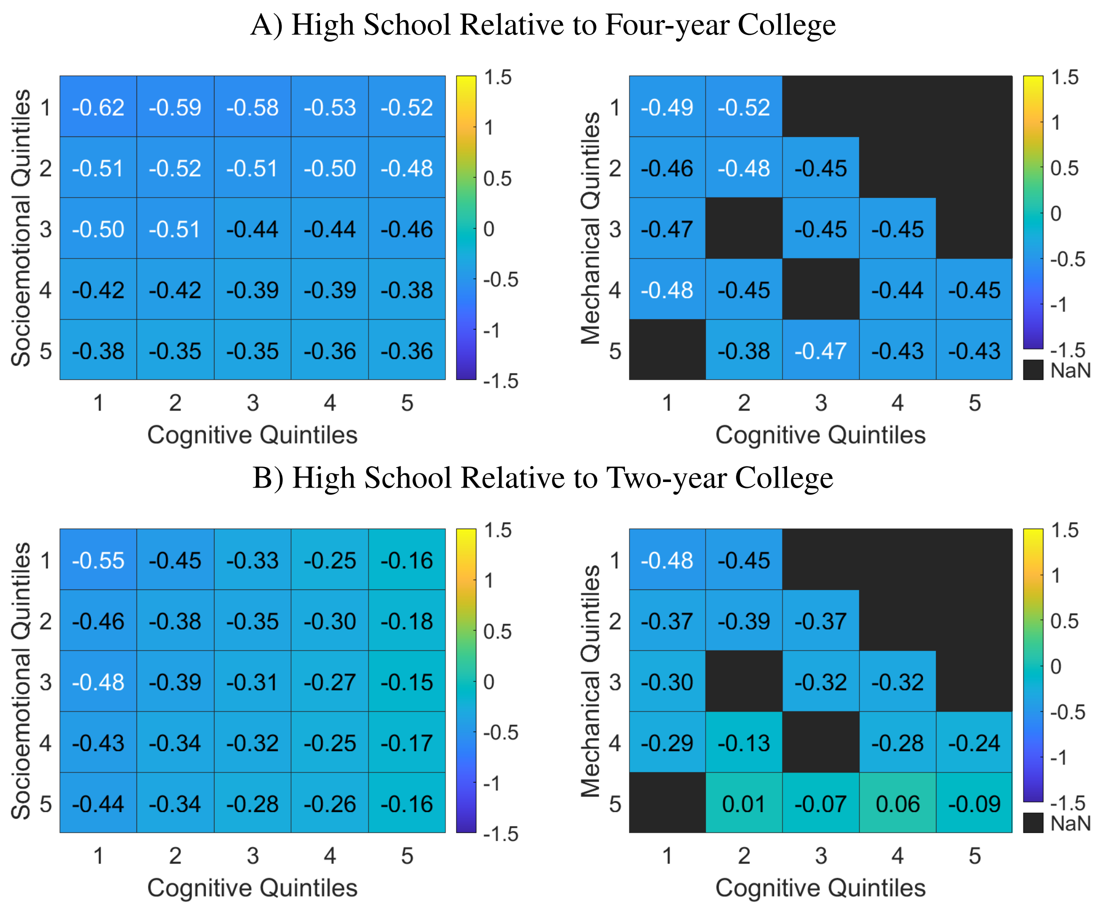
Figure 3.32: Treatment on the Treated for Two-Year Graduates - Women

A) High School Relative to Four-year College



B) High School Relative to Two-year College





Note: The simulated data contains 1,000 replications of the sample. We compute the average treatment effect relative to the next-best alternative, normalized by the mean of the next-best alternative at each grid point

Figure 3.34: Treatment on the Treated for High School Graduates - Women

Table 3.1: Summary Statistics

	Men				Women			
	All	HS	2y	4y	All	HS	2y	4y
Black	0.12	0.15	0.08	0.06	0.16	0.19	0.09	0.12
Hispanic	0.11	0.14	0.08	0.07	0.12	0.16	0.13	0.06
Living in the South in 1997	0.31	0.32	0.32	0.27	0.33	0.35	0.31	0.31
Living in an urban area in 1997	0.70	0.70	0.60	0.71	0.70	0.72	0.58	0.71
Single-parent home in 1997	0.40	0.51	0.36	0.21	0.44	0.57	0.42	0.28
Family income 1997 ($1,000s)	56.72	47.07	52.40	75.92	54.72	41.80	51.36	72.18
No. siblings[a]	2.19	2.38	2.05	1.85	2.34	2.64	2.29	1.96
Mother's years of schooling in 1997	13.30	12.66	13.15	14.54	13.19	12.33	12.93	14.36
Cohort 1981	0.21	0.22	0.20	0.20	0.23	0.22	0.22	0.23
Cohort 1982	0.19	0.20	0.16	0.20	0.21	0.19	0.27	0.23
Cohort 1983	0.23	0.22	0.26	0.23	0.20	0.21	0.12	0.20
Cohort 1984	0.19	0.17	0.23	0.20	0.17	0.17	0.13	0.19
Earnings ($1,000s)[b]	39.30	31.78	42.10	52.70	26.36	16.21	26.66	39.47
Living in an urban area	0.79	0.75	0.77	0.85	0.77	0.76	0.58	0.81
Region of current residency: South	0.33	0.35	0.32	0.30	0.38	0.38	0.36	0.37
Region of current residency: North Central	0.28	0.26	0.30	0.31	0.25	0.24	0.29	0.26
Region of current residency: West	0.23	0.22	0.22	0.24	0.22	0.23	0.20	0.21
No. Children	0.46	0.55	0.56	0.27	0.96	1.34	1.05	0.46
Married	0.37	0.34	0.45	0.41	0.45	0.42	0.52	0.47
Observations	1,181	707	97	377	1,204	624	99	481

Note: [a]2011 is the most recent year in which sibling information is available in the NLSY97. [b]Annual earnings are measured in real 2010 USD, at ages 28-30. Cohort 1980 is the baseline category for cohort dummies. East is the baseline category for region dummies. HS = High school graduates, 2y = two-year graduates, and 4y = four-year graduates. If a person enrolled but did not graduate, she is classified as a high school graduate.

Table 3.2: Model Specification

	Schooling	Earnings	TIPI tests	ASVAB tests
Cohort dummies	✓	✓	✓	✓
Race dummy (being Black)	✓	✓	✓	✓
Ethnicity dummy (being Hispanic)	✓	✓	✓	✓
Living in the South in 1997	✓		✓	✓
Living in an urban area in 1997	✓		✓	✓
Single-parent home in 1997	✓		✓	✓
Family income in 1997	✓		✓	✓
Number of siblings[a]	✓		✓	✓
Mother's years of schooling in 1997	✓		✓	✓
Regional variables at age 17[b]	✓			
Years of schooling when TIPI was taken			✓	
Years of schooling when ASVAB was taken				✓
Living in an urban area		✓		
Region of current residency (dummies)		✓		
Number of biological children in the household		✓		
Dummy variable for being married		✓		

Note: [a]2011 is the most recent year in which sibling information is available in the NLSY97. [b]Regional variables include average public tuition, unemployment relative to HS graduates, and wage rate relative to HS graduates. Regional variables are specific to each sector (2YC and 4YC). TIPI is the Ten-Item Personality Inventory Scale, which provides the socioemotional measures. ASVAB is the Armed Services Vocational Aptitude Battery test, which provides information for the cognitive and mechanical measures.

Table 3.3: Reduced-form Results: OLS Results

	Earnings ($10,000s)					
	All	All	Men	Men	Women	Women
Dummy for being a 2y grad.	0.865***	0.727***	0.697***	0.584**	0.940***	0.786***
	(0.180)	(0.181)	(0.267)	(0.267)	(0.230)	(0.227)
Dummy for being a 4y grad.	1.761***	1.487***	1.592***	1.328***	1.733***	1.450***
	(0.135)	(0.137)	(0.225)	(0.231)	(0.145)	(0.149)
Cognitive		0.279***		0.224**		0.488***
		(0.072)		(0.102)		(0.089)
Socioemotional		0.141***		0.108		0.168***
		(0.052)		(0.089)		(0.052)
Mechanical		0.215***		0.293***		-0.148
		(0.076)		(0.098)		(0.102)

Note: Standard errors are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Reduced-form estimates come from OLS specifications. Annual earnings are measured in $10,000s and in real 2010 USD, at ages 28-30. Abilities are measured as the standardized average of the same measures used in the measurement system. Staying as a high school graduate is the baseline category. Individuals who enrolled at some point in college but did not graduate are classified as high school graduates.

Table 3.4: Reduced-form vs. Structural Estimates

	Men					Women				
	School Choice		Earnings			School Choice		Earnings		
	2y	4y	HS	2y	4y	2y	4y	HS	2y	4y
Panel A: Reduced-Form Estimates:										
Cognitive	0.22**	0.76***	-0.00	-0.14	0.59**	0.32***	0.71***	0.27***	0.59	0.75***
	(0.09)	(08)	(0.11)	(0.37)	(0.24)	(0.12)	(0.10)	(0.09)	(0.40)	(0.17)
Socioemotional	0.09	0.14**	0.15*	-0.17	0.03	0.07	0.20***	0.22***	0.37*	0.06
	(0.07)	(0.07)	(0.09)	(0.33)	(0.19)	(0.08)	(0.06)	(0.06)	(0.21)	(0.10)
Mechanical	0.09	0.08	0.37***	0.08	0.29	0.25*	0.21*	0.03	-0.85**	-0.27
	(0.09)	(0.07)	(0.11)	(0.25)	(0.25)	(0.14)	(0.11)	(0.11)	(0.38)	(0.20)
Panel B: Structural Estimates:										
Cognitive	0.27***	0.95***	0.09	-0.71*	0.69**	0.13	0.79***	0.42***	0.66	0.74**
	(0.08)	(0.09)	(0.13)	(0.38)	(0.29)	(0.15)	(0.15)	(0.15)	(0.48)	(0.33)
Socioemotional	-0.02	0.03	0.04	-0.49	0.04	0.02	0.02	0.28***	0.34	0.03
	(0.09)	(0.08)	(0.13)	(0.39)	(0.25)	(0.08)	(0.08)	(0.09)	(0.30)	(0.16)
Mechanical	-0.07	-0.35***	0.30**	0.72**	-0.04	0.22	-0.24*	-0.22	-0.89**	-0.49
	(0.08)	(0.08)	(0.13)	(0.32)	(0.28)	(0.15)	(0.14)	(0.16)	(0.43)	(0.33)

Note: Standard errors are presented in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Coefficients are standardized. Reduced-form estimates come from OLS specifications in the case of earnings and from a multinomial probit estimation in the case of schooling choices. Annual earnings are measured in real 2010 USD, at ages 28-30. HS = High school graduates, 2y = two-year graduates, and 4y = four-year graduates. If a person enrolled but did not graduate, she is classified as a high school graduate.

Table 3.5: Point Estimates - Men, Utility

	HS → 2y Grad.	HS → 4y Grad.
Constant	-0.47	0.23
	(0.93)	(0.60)
Black	-0.30	-0.25
	(0.21)	(0.18)
Hispanic	-0.30	-0.25
	(0.20)	(0.18)
Living in the South in 1997	0.03	0.03
	(0.15)	(0.13)
Living in an urban area in 1997	-0.20*	0.06
	(0.12)	(0.12)
Single-parent home in 1997	-0.36***	-0.77***
	(0.13)	(0.12)
Family income 1997	-0.01	0.21***
	(0.07)	(0.05)
No. siblings	-0.07	-0.03
	(0.07)	(0.06)
Mother's years of schooling in 1997	0.11	0.43***
	(0.07)	(0.06)
Cohort 1981	0.09	-0.15
	(0.19)	(0.18)
Cohort 1982	-0.07	-0.14
	(0.23)	(0.20)
Cohort 1983	0.12	-0.07
	(0.18)	(0.18)
Cohort 1984	0.24	0.17
	(0.19)	(0.18)
Tuition	-0.16	0.14
	(0.46)	(0.10)
Unemployment Difference (relative to HS)	1.46	3.52
	(3.80)	(4.39)
Wage Ratio (relative to HS)	-0.13	-0.34
	(0.50)	(0.23)
Cognitive factor	0.27***	0.95***
	(0.08)	(0.09)
Socioemotional factor	-0.02	0.03
	(0.08)	(0.09)
Mechanical factor	-0.07	-0.35***
	(0.08)	(0.08)

Note: Standard errors are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Family income in 1997, siblings, and mother's years of schooling are standardized. Tuition is in $10,000s. Loadings are also standardized.

Table 3.6: Point Estimates - Men, Earnings.

	HS	2y Grad.	4y Grad.
Constant	2.37***	4.72***	5.08***
	(0.35)	(0.86)	(0.74)
Black	-1.06***	-0.41	-0.54
	(0.27)	(0.95)	(0.73)
Hispanic	-0.22	0.89	-0.59
	(0.27)	(0.85)	(0.69)
Living in an urban area	0.25	-0.76	0.28
	(0.22)	(0.59)	(0.50)
Region of current residency: South	0.27	-0.85	0.00
	(0.27)	(0.69)	(0.55)
Region of current residency: North Central	-0.14	-0.07	-0.59
	(0.29)	(0.71)	(0.55)
Region of current residency: West	0.36	-0.83	-0.31
	(0.30)	(0.76)	(0.57)
Cohort 1981	0.19	-0.50	-1.11*
	(0.29)	(0.79)	(0.58)
Cohort 1982	-0.02	0.10	-1.21**
	(0.30)	(0.82)	(0.58)
Cohort 1983	-0.04	-1.23*	-0.88
	(0.29)	(0.73)	(0.56)
Cohort 1984	0.41	-0.30	-0.07
	(0.31)	(0.76)	(0.59)
Children in the Household	0.19	-0.18	0.53*
	(0.12)	(0.31)	(0.29)
Married	1.48***	2.19***	1.13***
	(0.22)	(0.56)	(0.40)
Cognitive factor	0.09	-0.71*	0.69**
	(0.13)	(0.38)	(0.29)
Socioemotional factor	0.04	-0.49	0.04
	(0.13)	(0.39)	(0.25)
Mechanical factor	0.30**	0.72**	-0.04
	(0.13)	(0.32)	(0.28)

Note: Standard errors are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Earnings are measured in $10,000s. Loadings are also standardized.

Table 3.7: Point Estimates - Men, Cognitive System.

	Math	Word	Para	Num	Cod	Arith
Constant	0.07	0.30***	0.09	-0.17*	-0.26***	0.15*
	(0.08)	(0.09)	(0.09)	(0.10)	(0.10)	(0.09)
Black	-0.64***	-0.71***	-0.62***	-0.29***	-0.38***	-0.92***
	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.08)
Hispanic	-0.30***	-0.43***	-0.33***	-0.35***	-0.15*	-0.43***
	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.08)
Living in the South in 1997	0.05	-0.03	-0.05	0.10	0.12**	0.04
	(0.05)	(0.05)	(0.06)	(0.06)	(0.06)	(0.05)
Living in an urban area in 1997	0.07	0.06	0.02	0.10*	0.12**	0.09*
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)
Single-parent home in 1997	-0.12**	-0.05	-0.10*	-0.03	-0.12**	-0.11**
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)
Family income 1997	0.06**	0.06**	0.05*	0.06**	0.04	0.05**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
No. siblings	-0.00	-0.04*	-0.03	-0.07**	-0.07**	0.02
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Mother's years of schooling in 1997	0.23***	0.22***	0.19***	0.12***	0.16***	0.20***
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
Cohort 1981	0.13	0.01	0.09	0.16*	0.15	0.13
	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.08)
Cohort 1982	0.08	-0.09	-0.08	0.23**	0.09	0.10
	(0.09)	(0.09)	(0.10)	(0.11)	(0.10)	(0.10)
Cohort 1983	-0.02	-0.21*	-0.05	0.18	0.15	0.08
	(0.11)	(0.11)	(0.12)	(0.13)	(0.12)	(0.11)
Cohort 1984	-0.24**	-0.37***	-0.11	0.04	0.07	0.02
	(0.12)	(0.13)	(0.13)	(0.15)	(0.14)	(0.13)
Years of schooling when the ASVAB was taken	0.35***	0.22***	0.23***	0.35***	0.32***	0.26***
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)
Cognitive factor	1	0.87***	1.01***	0.80***	0.75***	1.09***
	-	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)

Note: Standard errors are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Family income in 1997, siblings, mother's years of schooling, and the years of schooling when the ASVAB test was taken are standardized. Arith = arithmetic reasoning, Math = math knowledge, Word = word knowledge, Para = paragraph comprehension, Cod = coding speed, Num = numerical operations.

Table 3.8: Point Estimates - Men, Socioemotional System.

	Extro	Agree	Consc	Stab	Open
Black	-0.11	-0.05	0.39***	0.16	0.34***
	(0.10)	(0.11)	(0.11)	(0.13)	(0.10)
Hispanic	0.00	0.10	0.16	0.05	0.10
	(0.10)	(0.11)	(0.11)	(0.13)	(0.10)
Living in the south in 1997	0.00	-0.09	0.09	-0.04	-0.01
	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)
Living in an urban area in 1997	-0.04	-0.06	-0.07	-0.26***	0.02
	(0.07)	(0.07)	(0.07)	(0.09)	(0.07)
Single-parent home in 1997	-0.01	-0.09	-0.12*	-0.19**	0.07
	(0.07)	(0.07)	(0.07)	(0.09)	(0.07)
Family income 1997	0.10***	-0.03	0.01	0.02	0.01
	(0.03)	(0.04)	(0.04)	(0.04)	(0.03)
No. siblings	0.02	0.07**	0.02	0.01	0.00
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)
Mother's years of schooling in 1997	0.01	0.03	-0.04	0.05	0.06*
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Cohort 1981	0.04	-0.15	-0.10	-0.00	-0.11
	(0.10)	(0.11)	(0.10)	(0.12)	(0.10)
Cohort 1982	0.14	-0.22**	-0.02	-0.08	-0.12
	(0.10)	(0.11)	(0.10)	(0.13)	(0.10)
Cohort 1983	0.17*	-0.21**	-0.26***	0.04	-0.17*
	(0.10)	(0.10)	(0.10)	(0.12)	(0.10)
Cohort 1984	0.04	-0.06	-0.23**	-0.07	0.04
	(0.10)	(0.11)	(0.10)	(0.13)	(0.10)
Years of schooling when the TIPI was taken	-0.02	0.09**	0.02	0.20***	0.03
	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)
Socioemotional factor	1	2.46***	1.61***	3.99***	1.39***
	-	(0.60)	(0.39)	(1.14)	(0.32)

Note: Standard errors are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Family income in 1997, siblings, mother's years of schooling, and the years of schooling when the TIPI test was taken are standardized. Extro = extroversion, Stab = emotional stability, Open = openness to experience, Agree = agreeableness, Consc = conscientiousness.

Table 3.9: Point Estimates - Men, Mechanical System.

	Mech	Elect	Auto	Shop
Constant	0.55***	0.70***	0.93***	1.11***
	(0.09)	(0.10)	(0.09)	(0.10)
Black	-0.97***	-0.69***	-1.01***	-0.74***
	(0.08)	(0.09)	(0.08)	(0.09)
Hispanic	-0.47***	-0.55***	-0.65***	-0.45***
	(0.08)	(0.09)	(0.09)	(0.09)
Living in the South in 1997	0.00	-0.04	-0.07	0.07
	(0.05)	(0.06)	(0.06)	(0.06)
Living in an urban area in 1997	-0.13**	-0.05	-0.18***	-0.32***
	(0.06)	(0.06)	(0.06)	(0.06)
Single-parent home in 1997	-0.16***	-0.13**	0.01	-0.08
	(0.05)	(0.06)	(0.05)	(0.06)
Family income 1997	0.02	-0.00	-0.04	-0.07**
	(0.03)	(0.03)	(0.03)	(0.03)
No. siblings	-0.03	-0.00	-0.05*	-0.05*
	(0.03)	(0.03)	(0.03)	(0.03)
Mother's years of schooling in 1997	0.16***	0.13***	0.05*	0.06**
	(0.03)	(0.03)	(0.03)	(0.03)
Cohort 1981	0.18**	-0.05	0.04	-0.14
	(0.08)	(0.09)	(0.09)	(0.09)
Cohort 1982	-0.04	-0.16	-0.16	-0.49***
	(0.10)	(0.11)	(0.10)	(0.11)
Cohort 1983	-0.03	-0.36***	-0.40***	-0.70***
	(0.12)	(0.13)	(0.12)	(0.12)
Cohort 1984	-0.16	-0.41***	-0.61***	-0.81***
	(0.14)	(0.15)	(0.14)	(0.15)
Years of schooling when the ASVAB was taken	0.21***	0.22***	0.04	0.15***
	(0.04)	(0.05)	(0.05)	(0.05)
Mechanical factor	1	1.29***	1.63***	1.64***
	-	(0.09)	(0.11)	(0.12)
Cognitive factor	0.60***	0.50***	0.02	0
	(0.04)	(0.04)	(0.05)	-

Note: Standard errors are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Family income in 1997, siblings, mother's years of schooling, and the years of schooling when the ASVAB test was taken are standardized. Mech = mechanical comprehension, Auto = auto information, Elect = electronics information.

Table 3.10: Point Estimates - Women, Utility

	HS → 2y Grad.	HS → 4y Grad.
Constant	-0.49	-0.23
	(0.61)	(0.55)
Black	-0.24	0.03
	(0.19)	(0.15)
Hispanic	-0.03	-0.19
	(0.18)	(0.16)
Living in the South in 1997	-0.14	0.05
	(0.18)	(0.12)
Living in an urban area in 1997	-0.28**	-0.00
	(0.13)	(0.11)
Single-parent home in 1997	-0.26**	-0.65***
	(0.12)	(0.10)
Family income 1997	0.11	0.31***
	(0.07)	(0.06)
No.siblings	-0.03	-0.06
	(0.06)	(0.05)
Mother's years of schooling in 1997	0.17***	0.48***
	(0.07)	(0.05)
Cohort 1981	-0.09	0.26*
	(0.23)	(0.16)
Cohort 1982	0.19	0.50***
	(0.21)	(0.16)
Cohort 1983	-0.37	0.22
	(0.23)	(0.17)
Cohort 1984	-0.13	0.34**
	(0.19)	(0.16)
Tuition	-0.51	0.11
	(0.41)	(0.08)
Unemployment Difference (relative to HS)	-3.30	-0.35
	(4.39)	(4.22)
Wage Ratio (relative to HS)	0.06	-0.09
	(0.27)	(0.16)
Cognitive factor	0.13	0.79***
	(0.15)	(0.15)
Socioemotional factor	0.02	0.02
	(0.08)	(0.08)
Mechanical factor	0.22	-0.24*
	(0.15)	(0.14)

Note: Standard errors are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Family income in 1997, siblings, and mother's years of schooling are standardized. Tuition is in \$10,000s. Loadings are also standardized.

Table 3.11: Point Estimates - Women, Earnings.

	HS	2y Grad.	4y Grad.
Constant	2.37***	3.25***	3.49***
	(0.24)	(0.79)	(0.48)
Black	-0.09	0.40	-0.23
	(0.17)	(0.81)	(0.37)
Hispanic	0.16	0.21	-0.28
	(0.17)	(0.66)	(0.47)
Living in an urban area (currently)	0.20	-0.50	0.48
	(0.15)	(0.45)	(0.30)
Region of current residency: South	-0.29	0.40	0.16
	(0.19)	(0.68)	(0.34)
Region of current residency: North Central	-0.21	0.35	-0.15
	(0.20)	(0.71)	(0.36)
Region of current residency: West	-0.36*	0.54	0.47
	(0.20)	(0.78)	(0.37)
Cohort 1981	-0.31*	0.38	0.15
	(0.18)	(0.65)	(0.37)
Cohort 1982	-0.30	0.39	-0.28
	(0.19)	(0.60)	(0.37)
Cohort 1983	-0.18	-0.63	0.01
	(0.19)	(0.78)	(0.38)
Cohort 1984	0.02	-1.14	0.61
	(0.20)	(0.73)	(0.39)
Children in the Household	-0.36***	-0.33	-0.69***
	(0.05)	(0.21)	(0.15)
Married	0.14	-0.34	0.16
	(0.13)	(0.48)	(0.25)
Cognitive factor	0.42***	0.65	0.74**
	(0.15)	(0.48)	(0.33)
Socioemotional factor	0.28***	0.34	0.03
	(0.09)	(0.30)	(0.16)
Mechanical factor	-0.22	-0.89**	-0.49
	(0.16)	(0.43)	(0.33)

Note: Standard errors are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Earnings are measured in $10,000s. Loadings are also standardized.

Table 3.12: Point Estimates - Women, Cognitive System.

	Math	Word	Para	Num	Cod	Arith
Constant	0.22***	0.17**	0.34***	-0.01	0.33***	0.07
	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.08)
Black	-0.56***	-0.56***	-0.55***	-0.15**	-0.30***	-0.69***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.07)
Hispanic	-0.13*	-0.30***	-0.14*	-0.13*	0.02	-0.19**
	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Living in the South in 1997	0.08	0.01	-0.02	0.07	0.09*	-0.05
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Living in an urban area in 1997	0.04	0.02	-0.06	0.08	0.02	0.03
	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	(0.05)
Single-parent home in 1997	-0.25***	-0.18***	-0.16***	-0.22***	-0.15***	-0.22***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Family income 1997	0.10***	0.08***	0.07**	0.09***	0.08***	0.06**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
No. siblings	-0.03	-0.05*	-0.02	-0.01	-0.07***	0.01
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
Mother's years of schooling in 1997	0.17***	0.24***	0.22***	0.09***	0.09***	0.20***
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)
Cohort 1981	0.09	0.08	0.07	0.14*	-0.01	0.09
	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
Cohort 1982	0.11	-0.02	0.03	0.16*	-0.07	0.07
	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.08)
Cohort 1983	-0.05	-0.11	-0.08	0.18*	-0.17	0.09
	(0.09)	(0.10)	(0.10)	(0.10)	(0.11)	(0.10)
Cohort 1984	-0.30***	-0.14	-0.19	-0.01	-0.30**	0.05
	(0.11)	(0.11)	(0.12)	(0.12)	(0.13)	(0.11)
Years of schooling when the ASVAB was taken	0.30***	0.26***	0.18***	0.29***	0.19***	0.23***
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Cognitive factor	1	0.97***	1.06***	0.76***	0.71***	1.05***
	-	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)

Note: Standard errors are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Family income in 1997, siblings, mother's years of schooling, and the years of schooling when the ASVAB test was taken are standardized. Arith = arithmetic reasoning, Math = math knowledge, Word = word knowledge, Para = paragraph comprehension, Cod = coding speed, Num = numerical operations.

Table 3.13: Point Estimates - Women, Socioemotional System.

	Extro	Agree	Consc	Stab	Open
Black	-0.05	-0.08	0.11	0.22*	0.25***
	(0.09)	(0.10)	(0.10)	(0.12)	(0.09)
Hispanic	0.04	-0.07	-0.02	0.37***	0.05
	(0.10)	(0.11)	(0.10)	(0.13)	(0.10)
Living in the south in 1997	-0.02	0.06	-0.06	0.03	-0.03
	(0.07)	(0.07)	(0.07)	(0.09)	(0.07)
Living in an urban area in 1997	0.07	0.07	0.08	0.06	0.13*
	(0.07)	(0.08)	(0.07)	(0.09)	(0.07)
Single-parent home in 1997	-0.06	-0.03	0.01	-0.14	-0.06
	(0.07)	(0.07)	(0.07)	(0.09)	(0.07)
Family income 1997	0.04	-0.04	-0.01	-0.00	0.03
	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)
No. siblings	0.02	-0.04	-0.03	-0.01	-0.01
	(0.03)	(0.04)	(0.03)	(0.04)	(0.03)
Mother's years of schooling in 1997	0.06*	0.06	-0.06	0.05	0.07*
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Cohort 1981	-0.06	-0.04	-0.01	-0.25**	-0.00
	(0.09)	(0.10)	(0.10)	(0.12)	(0.10)
Cohort 1982	0.00	-0.13	-0.06	-0.14	0.01
	(0.10)	(0.11)	(0.10)	(0.12)	(0.10)
Cohort 1983	0.02	-0.29***	-0.07	-0.13	0.04
	(0.10)	(0.11)	(0.10)	(0.12)	(0.10)
Cohort 1984	0.09	0.07	-0.12	-0.05	0.12
	(0.10)	(0.11)	(0.10)	(0.13)	(0.10)
Age when the TIPI was taken	0.03	0.10**	0.10***	0.17***	0.01
	(0.03)	(0.04)	(0.04)	(0.05)	(0.04)
Socioemotional factor	1	2.24***	1.37***	3.53***	1.21***
	-	(0.49)	(0.30)	(0.76)	(0.25)

Note: Standard errors are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Family income in 1997, siblings, mother's years of schooling, and the years of schooling when the TIPI test was taken are standardized. Extro = extroversion, Stab = emotional stability, Open = openness to experience, Agree = agreeableness, Consc = conscientiousness.

Table 3.14: Point Estimates - Women, Mechanical System.

	Mech	Elect	Auto	Shop
Constant	0.08	0.10	0.02	0.14*
	(0.08)	(0.07)	(0.07)	(0.07)
Black	-0.79***	-0.37***	-0.63***	-0.56***
	(0.07)	(0.06)	(0.06)	(0.06)
Hispanic	-0.27***	-0.26***	-0.31***	-0.23***
	(0.07)	(0.07)	(0.07)	(0.07)
Living in the South in 1997	-0.09*	0.02	-0.10**	0.02
	(0.05)	(0.05)	(0.05)	(0.04)
Living in an urban area in 1997	-0.09*	-0.15***	-0.20***	-0.12***
	(0.05)	(0.05)	(0.05)	(0.05)
Single-parent home in 1997	-0.11**	-0.11**	-0.05	-0.05
	(0.05)	(0.04)	(0.05)	(0.04)
Family income 1997	0.05*	0.01	0.01	0.03
	(0.03)	(0.02)	(0.02)	(0.02)
No. siblings	-0.01	-0.03	0.01	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Mother's years of schooling in 1997	0.12***	0.13***	0.12***	0.10***
	(0.02)	(0.02)	(0.02)	(0.02)
Cohort 1981	-0.02	-0.06	-0.03	-0.19***
	(0.07)	(0.07)	(0.07)	(0.07)
Cohort 1982	0.10	-0.03	0.01	-0.20***
	(0.08)	(0.08)	(0.08)	(0.07)
Cohort 1983	0.04	-0.18**	-0.04	-0.34***
	(0.09)	(0.09)	(0.09)	(0.09)
Cohort 1984	-0.08	-0.33***	-0.19*	-0.36***
	(0.11)	(0.10)	(0.10)	(0.10)
Years of schooling when the ASVAB was taken	0.17***	0.17***	0.10***	0.10***
	(0.03)	(0.03)	(0.03)	(0.03)
Mechanical factor	1	1.24***	2.12***	1.30***
	-	(0.28)	(0.45)	(0.35)
Cognitive factor	0.42***	0.20**	-0.36**	0
	(0.12)	(0.10)	(0.15)	-

Note: Standard errors are in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Family income in 1997, siblings, mother's years of schooling, and the years of schooling when the ASVAB test was taken are standardized. Mech = mechanical comprehension, Auto = auto information, Elect = electronics information.

Table 3.15: Goodness of Fit: Schooling Choices and Earnings

	Schooling Choices		Earnings (in 10,000s)			
	Fraction of individuals		Mean		Variance	
	Simulated	Observed	Simulated	Observed	Simulated	Observed
Panel A. Men						
HS grad.	0.616	0.599	3.226	3.178	6.795	6.845
2y grad.	0.082	0.082	4.066	4.210	6.816	6.641
4y grad.	0.302	0.319	5.322	5.270	12.367	12.262
Panel B. Women						
HS grad.	0.553	0.518	1.728	1.565	2.476	2.528
2y grad.	0.084	0.082	2.594	2.649	4.944	4.936
4y grad.	0.363	0.400	3.719	3.973	6.581	6.502

Note: The simulated data contains 1,000 replications of the sample.

Table 3.16: Estimated loading parameters by socio-emotional measure: Ordered vs. Continuous Systems

	Men				Women			
	Number of Mixtures				Number of Mixtures			
	1	2	3	4	1	2	3	4
Panel A: Ordered System								
Extroversion	1	1	1	1	1	1	1	1
Agreeableness	2.32	2.36	2.39	2.41	2.18	2.23	2.23	2.24
Conscientiousness	1.65	1.53	1.58	1.68	1.38	1.42	1.33	1.33
Stability	3.71	3.43	3.48	3.91	4.06	4.03	3.89	3.90
Openness	1.31	1.33	1.37	1.38	1.16	1.15	1.24	1.24
Panel B: Continuous System								
Extroversion	1	1	1	1	1	1	1	1
Agreeableness	1.98	2.71	2.96	2.95	1.95	2.03	2.05	2.05
Conscientiousness	1.59	2.34	2.48	2.46	1.37	1.49	1.48	1.48
Stability	2.86	8.46	9.88	9.88	3.30	5.94	6.45	6.44
Openness	1.21	1.01	1.08	1.07	1.19	1.05	1.02	1.02

Note: Comparison of socioemotional system loadings under two assumptions: ordered (with 13 distinct levels) and continuous. We consider a multi-start algorithm for each specification.

Table 3.17: Ranking of Preferences by Group

Ranking		Group Percentage	
First Option	Second Option	Men	Women
HS degree	2y degree	30.2%	22.1%
HS degree	4y degree	31.4%	33.2%
2y degree	HS degree	5.4%	4.9%
2y degree	4y degree	2.8%	3.4%
4y degree	HS degree	22.0%	25.5%
4y degree	2y degree	8.3%	10.9%

Note: Each column shows the percentage of simulated individuals the indicated preference ranking by group. Columns sum up to 100%.

Table 3.18: Sector Preferences by Group

	Men			Women		
	First Option	Second Option	Third Option	First Option	Second Option	Third Option
HS degree	61.6%	27.3%	11.1%	55.3%	30.4%	14.3%
2y degree	8.2%	38.5%	53.3%	8.4%	33.0%	58.7%
4y degree	30.2%	34.2%	30.2%	36.3%	36.6%	22.1%

Note: Each column displays the percentage of simulated individuals in each group ranking each alternative in the place indicated by the headline, unconditionally of other alternative's places.

3.10 Additional Figures and Tables

Table 3.19: Wald Tests for Differences on the Effect of Decisions on Earnings between Women and Men in the Reduced Form Estimation

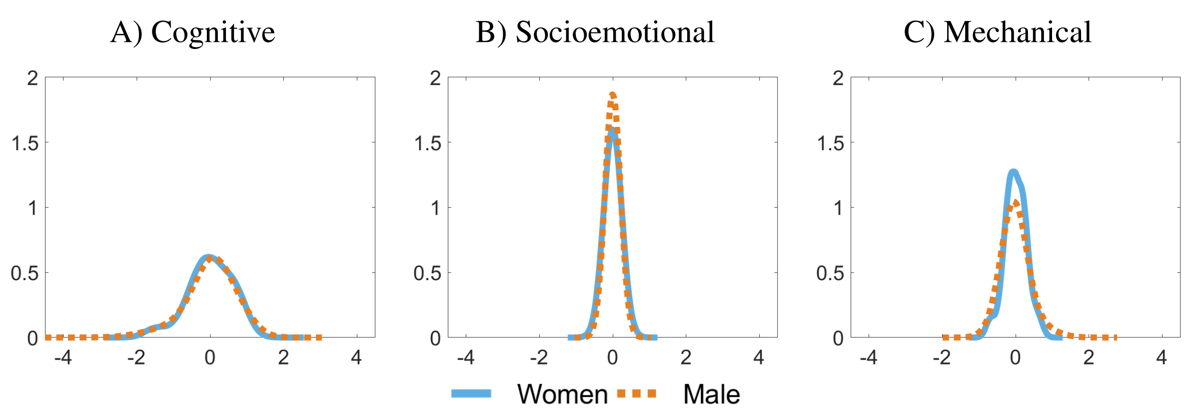
	F-statistic	P-value
Dummy for being a 2y grad.	0.33	0.565
Dummy for being a 4y grad.	0.20	0.655

Note: Wald tests are conducted on a fully-interacted model to pool estimates from Columns 4 and 6 in Table 3.3.

Table 3.20: T-tests for Differences on Loadings between Women and Men in the Structural Estimation

	Utility		Earnings		
	2y	4y	HS	2y	4y
Cognitive	0.829	0.942	-1.669	**-2.215**	-0.098
Socioemotional	-0.311	0.052	-1.528	-1.673	0.022
Mechanical	-1.719	-0.663	**2.454**	**2.979**	1.049

Note: The t-statistic is calculated as follows: $\hat{t} = \frac{\hat{\alpha}_M - \hat{\alpha}_W}{\sqrt{\hat{SE_{\alpha_M}}^2 + \hat{SE_{\alpha_W}}^2}}$.



Note: Comparison between the densities of men's abilities (orange, dashed line) and women's abilities (blue, solid line). The simulated data contains 1,000 replications of the sample. The number of mixtures for each factor is based on Likelihood-Ratio Tests comparing the fit in the measurement system. Cognitive factors have three mixtures, socioemotional factors have one mixture, and the uncorrelated components of the mechanical factors have two mixtures.

Figure 3.35: Factors - Densities, Comparisons between Women and Men.

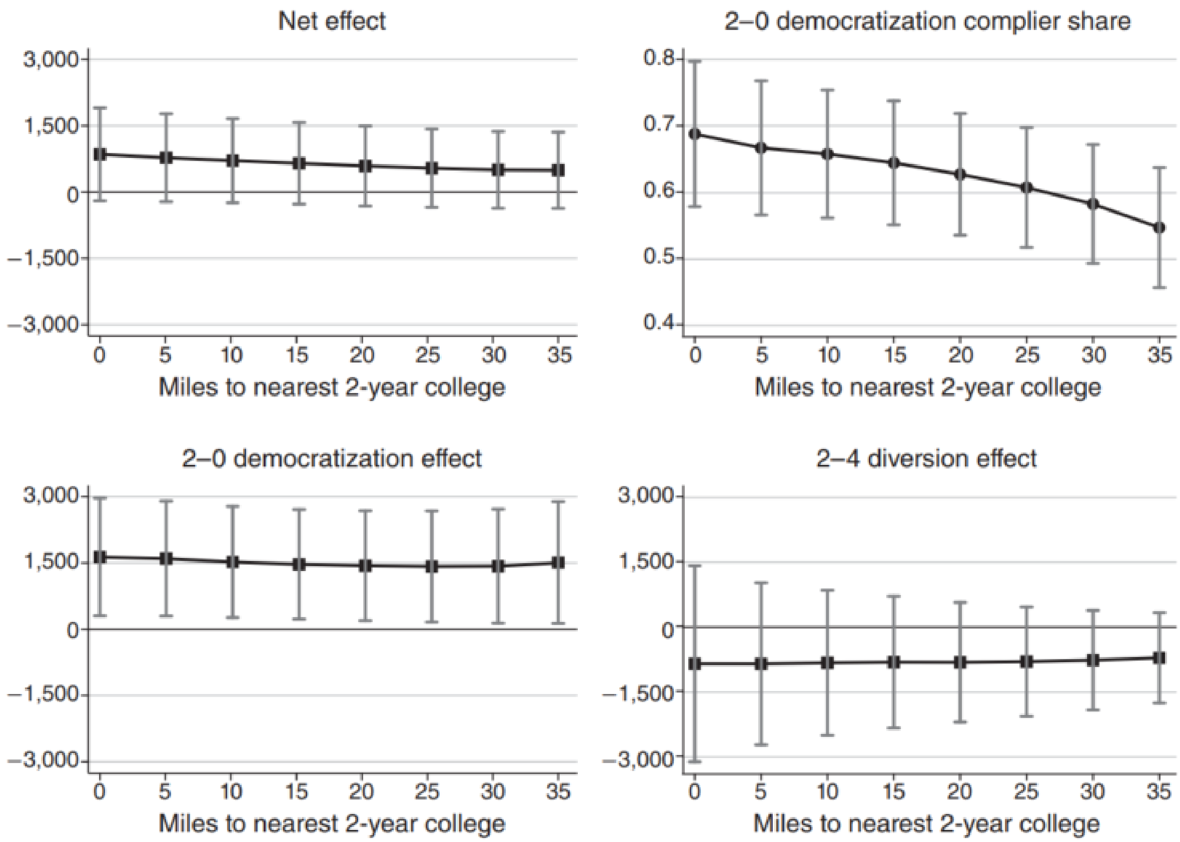
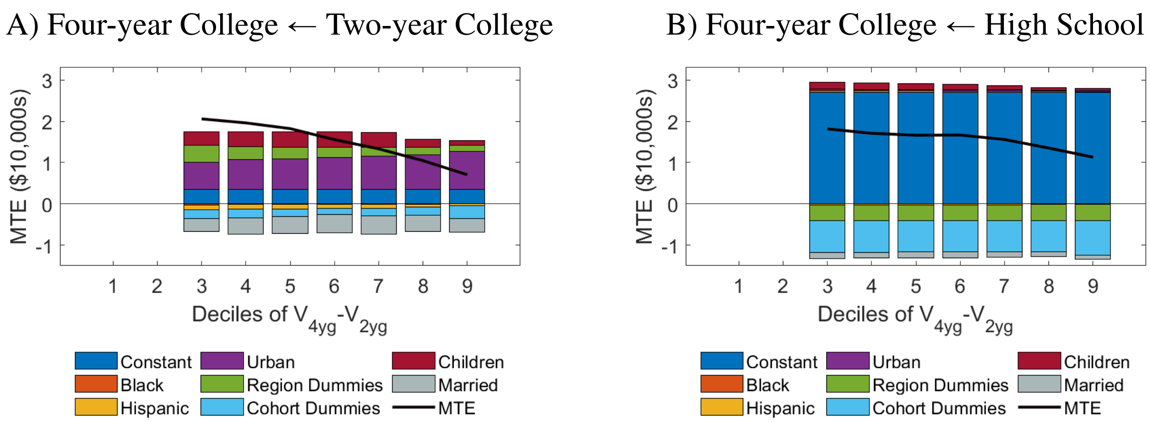


FIGURE 7. SELECTION PATTERNS AND EXTERNAL VALIDITY

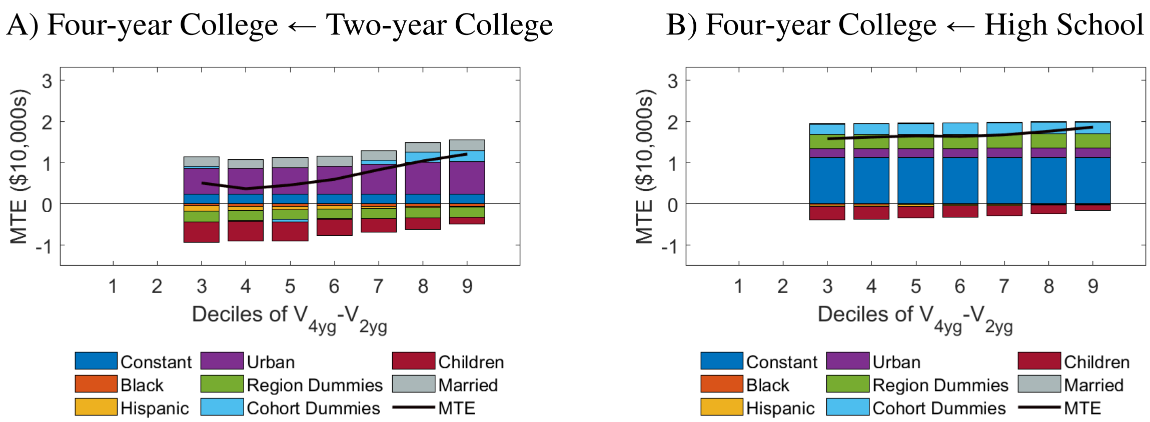
Notes: Each estimate is evaluated at a given two-year distance value holding four-year distance fixed at its mean. Quarterly earnings are measured in real 2010 US dollars and averaged within person over ages 28–30. Ninety-five percent confidence intervals are estimated via block bootstrap at the high school campus by cohort level.

Figure 3.36: Mountjoy (2022)'s Results.



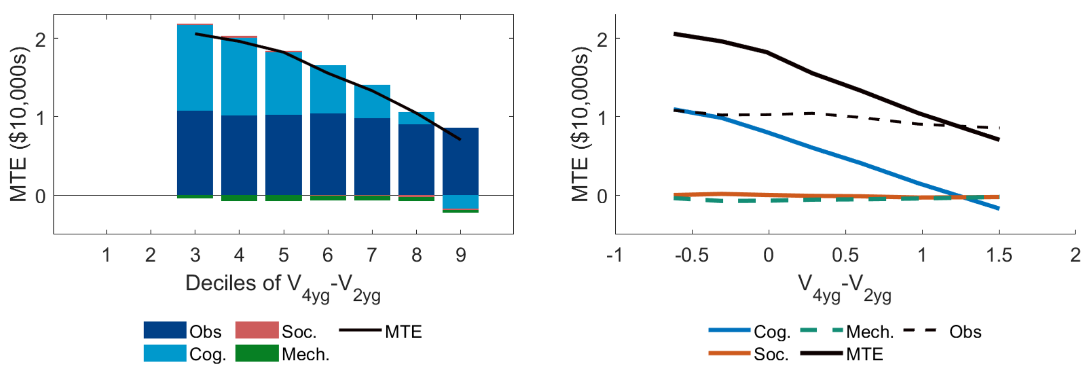
Note: The simulated data contains 1,000 replications of the sample. We use $|C_j - C_k| < \epsilon$ to determine indifference, where $\epsilon = 0.25$. Results are similar but noisier considering $\epsilon = 0.1$. Calculations are displayed for points with at least 0.25% of the simulated sample. This threshold does not change the calculations for populated-enough grids, but prevent outlier-driven evaluations. Even without this restrictions, we have some grid points without any simulated individual in the indifference margin.

Figure 3.37: MTE Decomposition of Observables for Four-year College - Men



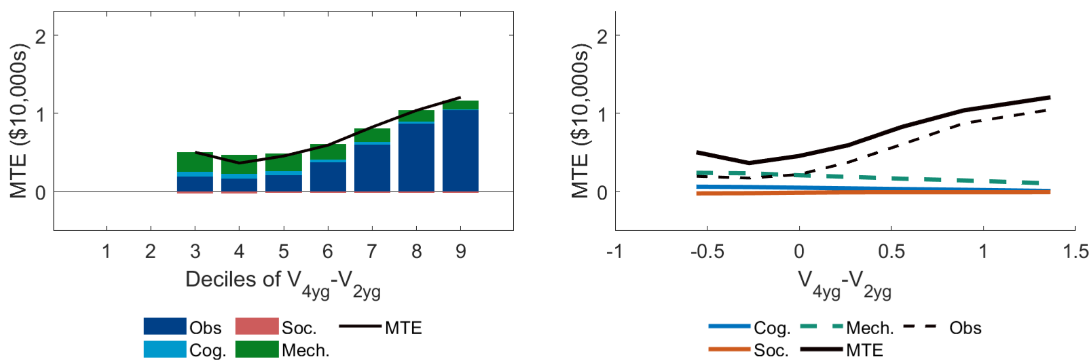
Note: The simulated data contains 1,000 replications of the sample. We use $|C_j - C_k| < \epsilon$ to determine indifference, where $\epsilon = 0.25$. Results are similar but noisier considering $\epsilon = 0.1$. Calculations are displayed for points with at least 0.25% of the simulated sample. This threshold does not change the calculations for populated-enough grids, but prevent outlier-driven evaluations. Even without this restrictions, we have some grid points without any simulated individual in the indifference margin.

Figure 3.38: MTE Decomposition of Observables for Four-year College - Women



Note: The simulated data contains 1,000 replications of the sample. We use $|C_j - C_k| < \epsilon$ to determine indifference, where $\epsilon = 0.25$. Results are similar but noisier considering $\epsilon = 0.1$. Calculations are displayed for points with at least 0.25% of the simulated sample. This threshold does not change the calculations for populated-enough grids, but prevent outlier-driven evaluations. Even without this restrictions, we have some grid points without any simulated individual in the indifference margin.

Figure 3.39: MTE Decomposition: 4y ← 2y - Men



Note: The simulated data contains 1,000 replications of the sample. We use $|C_j - C_k| < \epsilon$ to determine indifference, where $\epsilon = 0.25$. Results are similar but noisier considering $\epsilon = 0.1$. Calculations are displayed for points with at least 0.25% of the simulated sample. This threshold does not change the calculations for populated-enough grids, but prevent outlier-driven evaluations. Even without this restrictions, we have some grid points without any simulated individual in the indifference margin.

Figure 3.40: MTE Decomposition: 4y ← 2y - Women.