

NBER WORKING PAPER SERIES

DYNAMIC TREATMENT EFFECTS OF JOB TRAINING

Jorge Rodríguez
Fernando Saltiel
Sergio S. Urzúa

Working Paper 25408
<http://www.nber.org/papers/w25408>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2018

We thank Matias Cattaneo, Xavier D'Haultfoeuille, Claudio Ferraz, Maurice Kugler, Lance Lochner, Oscar Mitnik, Gabriel Ulyssea for useful comments and suggestions. We are indebted to seminar participants at IZA/CREST/OECD Conference on Labor Market Policy Evaluation (2018), George Mason University (2018), Latin American Meeting of the Econometric Society (2017), RIDGE Impact Evaluation of Labor Market Policies Conference (2017), PUC-Chile (2017), Banco de Desarrollo de America Latina (2016), PUC-Rio (2016), and LACEA-LAMES (2016). We thank the Ministry of Finance of Chile for providing us access to the data used in this paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by Jorge Rodríguez, Fernando Saltiel, and Sergio S. Urzúa. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Dynamic Treatment Effects of Job Training
Jorge Rodríguez, Fernando Saltiel, and Sergio S. Urzúa
NBER Working Paper No. 25408
December 2018
JEL No. C31,C32,J24

ABSTRACT

This paper estimates the dynamic returns to job training. We posit a dynamic-discrete choice model of sequential training participation, where choices and earnings depend on observed and unobserved characteristics. We define treatment effects, including policy relevant parameters, and link them to continuation values. The empirical analysis is carried out using data combining job training records, matched employee-employer information, and pre-labor market ability measures from Chile. We document small positive average returns, large unobserved heterogeneity in responses, and dynamic substitutability of training investments. Our policy relevant treatment effects vary across dynamic response types, highlighting the relevance of our framework.

Jorge Rodríguez
Universidad de los Andes
San Carlos de Apoquindo 2200
Santiago
Chile
jrodriguez@uandes.cl

Sergio S. Urzúa
Department of Economics
University of Maryland
3115J Tydings Hall
College Park, MD 20742
and NBER
urzua@econ.umd.edu

Fernando Saltiel
University of Maryland
Tydings Hall, 3114 Preinkert Dr.
College Park, MD 20742
saltiel@econ.umd.edu

1 Introduction

The evolving complexity and uncertainty in the demand for skills in a changing labor market induces workers to constantly change their human capital investment decisions. Recent research has documented workers may possess different sets of competencies vis-à-vis those required in the workplace.¹ In this context, understanding the dynamics of training decisions and their associated returns has gained prominence over the past years: as fast technological progress shifts the set of skills that jobs demand, workers may participate in on- or off-the-job training multiple times in their careers.²

This paper estimates the returns to job training in a context of dynamic training choices and labor market outcomes. We use our framework to provide new insights into the static and dynamic effects of job training on earnings, including the identification of heterogeneous responses across different groups as well as an empirical assessment of dynamic complementary (substitutability) arising from repeated training participation.³ Ultimately, we not only document who benefits from training but also how the timing and sequence of the training decisions define potential and actual gains (and losses).

We posit a tractable dynamic-discrete choice model of job training. In the model, a worker must decide whether or not to take part in a training course across multiple periods, and workers may participate in training on multiple occasions. For a given training history, and conditional on firm characteristics, the agent chooses to participate in a job training course if the perceived net benefits are positive.⁴ Individual choices and outcomes depend on observed characteristics as well as on unobserved heterogeneity, which we interpret as the initial stock of skills. Using a measurement system of pre-labor market test scores, we are able to nonparametrically identify the distribution of

¹Guvenen et al. (2015), Postel-Vinay and Lise (2015), Lise et al. (2016), Buhrmann (2017), and Saltiel et al. (2018) discuss skill (mis)match in the labor market.

²See Autor et al. (2003), SpitzOener (2006), Ingram and Neumann (2006), Acemoglu and Autor (2011), and Sanders and Taber (2012) for a literature on the changing returns to specific skills.

³Anticipating our analysis, if we let Y_t be earnings at the end of period t , the production function of Y_t exhibits dynamic complementarity (substitutability) if the return to I_t (training during present period) is higher (lower) conditional on prior training participation, i.e. $\frac{\partial^2 Y_t}{\partial I_{t-1} \partial I_t} \geq (\leq) 0$. We note that this result is captured by the same parameter, $\frac{\partial^2 Y_t}{\partial I_{t-1} \partial I_t}$, yet it is defined as a dynamic complementarity in the case of higher returns and as dynamic substitutability with lower returns to present training (Cunha and Heckman, 2006).

⁴Besides the index threshold-crossing property that determines the decision process, we do not model preferences and budget sets, thereby abstracting from strong assumptions about behavior and uncertainty—ubiquitous elements in the structural dynamic literature (Keane et al., 2011). Instead, we approximate choices and outcomes processes jointly, accounting for self-selection into program participation at each decision node.

latent skills and use it to identify the joint distribution of counterfactual earnings across potential training choices.⁵

We exploit a large-scale training program in Chile called “*Franquicia Tributaria*” (FT). FT fully subsidizes training courses at off-site providers for workers who are employed in a formal-sector firm. In the program, a worker can participate in a training course on multiple occasions, and a significant number of workers do so.⁶ We take advantage of administrative data on job training records for the population of workers and combine it with matched employee-employer data on labor market outcomes. We augment these data with measures on workers’ pre-labor market abilities coming from college admission test scores. We analyze the earnings returns to training for first-time labor market entrants, focusing on their first two years in the labor market.^{7,8}

Following the estimated model parameters, we document static and dynamic treatment effects of job training. In particular, we examine the effects of training on workers’ earnings in the first two years in the labor force. The static returns to training indicate that program participation in the first year raises average monthly earnings by 1.7%. While second-year participation positively impacts subsequent earnings, we find significant heterogeneity across first-period decisions, as this event increases earnings by 3.4% for workers not previously trained, while the equivalent effect for early trainees is smaller than 1%. The dynamic returns to training, meanwhile, capture the impact of first-period training on the present value of earnings in the first two periods in the labor market. The dynamic average treatment effect (DATE) indicates that first-period training increases the present value of earnings by 3.6%. We examine the mechanisms through which early training delivers positive medium-term impacts by decomposing the DATE parameter into the direct effect of training and its continuation value, which links human capital investment decisions and potential gains over time. We find that while the short- and medium-term direct effects are positive and

⁵We follow the models of Heckman and Navarro (2007) and Heckman et al. (2016), adapting them to the context of job training participation. We extend their analysis by considering the importance of dynamic complementarity/substitutability of human capital investments (training) and estimating dynamic policy relevant treatment effects.

⁶For workers in our estimation sample, 46% of those who take-up first period job training choose to do so again in their second year.

⁷We study training choices for labor market entrants to focus on workers with no prior job training experience. Furthermore, workers may find it especially valuable to take up job training as there is a longer time period over which they can capture the returns to training (Mincer, 1974).

⁸We measure earnings as the worker’s average monthly earnings in the first quarter of the year following the training event. The results are robust to other earnings measures. We note that we do not analyze two potential sources of gains from training: non-pecuniary benefits and effects on firm productivity which are not passed through onto wages.

significant, the continuation value is not statistically different from zero.⁹ We further analyze the factors driving the estimated continuation value by exploring the relationship between this parameter and dynamic complementarity (substitutability) of training decisions.¹⁰ In this context, we find evidence of dynamic substitutability, as training in the first period reduces the earnings returns to training in the second period. Furthermore, dynamic substitutability is stronger for high-skilled workers. We note that dynamic substitutability may be explained by the structure of the job training courses examined in this paper, or more generally by the production function of post-schooling human capital accumulation.¹¹

To examine the policy implications arising from these results, we estimate the effect of an increase in local course-hour availability. As this policy may affect workers' choices in both years, we define dynamic response types, defined by the responses to the policy change in each time period. For instance, we can identify a group of workers who are induced to participate in training in both time periods due to the policy change, despite being baseline never-participants. In this context, we contribute to the treatment effects literature by presenting and estimating dynamic policy-relevant treatment effects. We find that a 10% increase in course availability would increase the medium-term earnings of affected workers by 3%, with largely homogeneous effects across the skill distribution. Moreover, we find similar-sized impacts for larger program expansions, yet document that the effects are heterogeneous across dynamic response types, as workers induced to participate in both periods would enjoy the largest gains from the policy change.

Our paper contributes to an extensive literature on the effect of training programs on labor market outcomes. In a non-experimental context, the inherent identification challenge follows from potential self-selection into training.¹² To deal with this concern, various papers have relied on individual fixed effect estimators to account for unobserved individual heterogeneity.¹³ While [Heckman](#)

⁹This result stands in contrast with the positive continuation value of high school and college completion found by [Heckman et al. \(2016\)](#).

¹⁰In Section 5.3, we examine the difference between continuation values and dynamic complementarity (substitutability). We note that while the latter is informative about the production function of human capital of training across multiple periods, continuation values include workers' endogenous second-period participation decisions.

¹¹These results indicate that the post-school human capital production function differs from that of school-age children, for whom dynamic substitutability is positive ([Cunha et al., 2006](#))

¹²An extensive literature has identified the short-term effects of training through experimental variation ([Attanasio et al. \(2011\)](#), [Card et al. \(2011\)](#), [Alzúa et al. \(2014\)](#), [Attanasio et al. \(2015\)](#), [Ibarrarán et al. \(2015\)](#), among others). However, sample sizes in this literature tend to be small and the external validity of empirical estimates debatable.

¹³In the meta-analysis of [Card et al. \(2010\)](#), 51% of the studies implement some version of the fixed-effect estimator. See also [Ashenfelter \(1978\)](#), [Lynch \(1992\)](#), [Booth \(1993\)](#), [Veum \(1997\)](#), and [Lengermann \(1999\)](#).

et al. (1998, 1999) show that the standard fixed-effect estimator can effectively remove selection bias, this estimation strategy does not account for dynamic selection into training or estimate heterogeneous returns.¹⁴ As a result, we first contribute to this literature by considering dynamic sorting into job training participation.¹⁵ Moreover, we estimate heterogeneous returns to training across workers’ unobserved ability and allow for the effects to vary across training histories and time periods, noting these parameters have not been previously estimated in the job training literature. We additionally present the first estimates of the dynamic returns to training for employed workers and decompose the effect into direct effects and continuation values, while exploring the difference between continuation values and dynamic complementarity (substitutability).¹⁶ Lastly, we contribute to a growing literature exploring policy relevant treatment effects by defining dynamic response types and estimating the heterogeneous impacts of a policy change across each group (Heckman and Vytlacil, 2001; Carneiro et al., 2010; Mogstad et al., 2018; Mogstad and Torgovitsky, 2018).

The rest of the paper is organized as follows. Section 2 presents our model of sequential training participation with unobserved heterogeneity. Section 3 describes the institutional setup, data sources, sample characteristics and presents reduced-form estimates. Section 4 presents the estimated model parameters, considers goodness of fit, and documents the implied patterns of selection on unobservables. Section 5 defines the static and dynamic treatment parameters and presents evidence on the returns to job training. Section 6 discusses the simulated policy intervention, defines and identifies dynamic response types and presents evidence on the returns to the policy change. We conclude in Section 7.

¹⁴Many of the commonly-used reduced-form estimators are unable to identify meaningful, or policy-relevant, causal parameters in dynamic contexts. See Callaway and Sant’Anna (2018), de Chaisemartin and D’Haultfoeuille (2018), and Goodman-Bacon (2018) for analyses showing how differences-in-differences estimators fail to identify treatment effects parameters in the dynamic context.

¹⁵Abbring and van den Berg (2003) establishes non-parametric identification of a duration model where training occurs randomly in time. However, their identification result is established assuming individuals do not anticipate (and thus do not act on) the treatment timing. Our framework does not rely in this no-anticipation assumption as individuals are allowed to choose when to participate in training. Abbring and Heckman (2007) and Lechner (2009) also discuss identification in the context of dynamic program participation.

¹⁶We note that Abbring and van den Berg (2003); Fitzenberger and Völter (2007); Fredriksson and Johansson (2008); Fitzenberger et al. (2016) have previously estimated dynamic returns to training for unemployed workers. However, analyzing the dynamic returns to training for employed workers is of particular interest to labor economists, as existing theoretical models show that both firms and workers should benefit from programs which provide employees with a broad set of skills required to perform the job (Acemoglu and Pischke, 1999).

2 Dynamic Discrete Choice Model

In this section, we introduce a conceptual framework characterizing the dynamics of training decisions and labor market outcomes. The model follows Heckman and Navarro (2007) and Heckman et al. (2016), yet it is adapted to the context of job training decisions. It allows earnings counterfactuals to vary freely across all potential histories of training choices. As discussed below, its non-parametric identification is secured through the combination of a matching-on-unobservables assumption and exclusion restrictions. We start this section by presenting the theoretical model and continue with its identification and estimation.

2.1 Model

The essence of the model involves an agent making training choices for many periods and earnings, which directly depend on training decisions. In any period t , potential earnings depend on her current training decision as well as on the entire history of training activities.¹⁷ In period t , the agent makes her optimal decision by potentially taking into account past, present, and future training decisions and associated labor market outcomes.¹⁸ The agent is allowed to participate in job training as frequently as desired.

We model the dynamic training decision as a tree of sequential binary decisions, where the individual chooses training in each stage $t \in \mathcal{T} \equiv \{1, \dots, T\}$. Figure 1 depicts this decision tree for our case, where $T = 2$. We define \mathcal{H}_t as the set of all possible training decisions histories up through time t . An element in that set, $h_t \in \mathcal{H}_t$, represents a specific training path. For instance, in period $t = 2$, $\mathcal{H}_2 = \{0, 1\}$, where $h_2 = 1$ if the agent was trained in $t = 1$ and 0 otherwise. Likewise, $\mathcal{H}_3 = \{(0, 0), (1, 0), (1, 0), (1, 1)\}$, where each element $h_t \equiv (i, j)$ denotes the training decision in period $t = 1$ (first component) and $t = 2$ (second component).¹⁹

At each choice node, agent i solves a simple benefit-cost analysis to make her next training

¹⁷We focus on training choices for workers who do not fall into unemployment. We thus abstract from modeling the employment decision, which could affect the probability of training. We note the model could generally accommodate unemployment as an outcome variable, however.

¹⁸As we explain below, we do not explicitly specify individual preferences and expectations formation. Hence, forward-looking behavior is not imposed in the model.

¹⁹At this point, we depart from Heckman and Navarro (2007) and Heckman et al. (2016). In these models, a particular education path restricts the choice set (for example, getting a GED implies that the individual cannot attain the high school graduate node). In contrast, our model allows for the option of training independently of previous decisions.

choice. Let $D_{it}(h_t)$ be her training decision in period t , for history $h_t \in \mathcal{H}_t$. $D_{it}(h_t)$ equals one if she decides to participate in a training program and zero otherwise. Her optimal choice is given by:

$$D_{it}(h_t) = \begin{cases} 1 & \text{if } I_{it}(h_t) \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad h_t \in \mathcal{H}_t, t \in \mathcal{T} \quad (1)$$

where $I_{it}(h_t)$ denotes the net value of training in period t for a given history $h_t \in \mathcal{H}_t$. $I_{it}(h_t)$ could incorporate non-pecuniary benefits and costs of training. In principle, expression 1 provides a general framework. It can accommodate, for example, forward-looking agents anticipating and acting based on present and future benefits of training in period t , who are uncertain with respect to the true model that generates counterfactual earnings. We model the decision process using a linear-in-the-parameters equation to approximate the individual net benefit of training:

$$I_{it}(h_t) = X_i^I \beta^I(h_t) + \theta_i \lambda^I(h_t) + \varepsilon_{it}^I(h_t) \quad h_t \in \mathcal{H}, t \in \mathcal{T}, \quad (2)$$

where X_i^I is a vector of individual characteristics (observed by the econometrician and the agent), θ_i (with $E(\theta_i) = 0$) represents a fixed, latent ability endowment known by the agent but not the econometrician, and $\varepsilon_{it}^I(h_t)$ is an unobserved measurement error term. The effects of observed characteristics and unobserved ability on the value of training vary across time periods and training histories.²⁰

The agent progresses through each node after making training choices, and for each possible choice and training history, there is an associated labor market outcome. Let $Y_{it}(h_t, j)$ be potential earnings for a training decision $j \in \{0, 1\}$ made by worker i at time t with history h_t . Following Heckman et al. (2016), the potential outcome associated with $j \in \{0, 1\}$ and $h_t \in \mathcal{H}_t$ is given by:

$$Y_{it}(h_t, j) = X_i^Y \beta^Y(h_t, j) + \lambda^Y(h_t, j) \theta_i + \varepsilon_{it}^Y(h_t, j) \quad j \in \{0, 1\}, h_t \in \mathcal{H}, t \in \mathcal{T}, \quad (3)$$

where X_i^Y is a vector of observed characteristics. $\lambda^Y(h_t, j) \theta_i + \varepsilon_{it}^Y(h_t, j)$ is latent productivity,

²⁰This specification can accommodate any type of preferences. For example, it can be considered as an approximation of the value function of a forward-looking agent looking to maximize the present value of the earnings streams (Heckman et al., 2016).

which is unobserved by the econometrician. This unobserved component is separable into an ability component known by the agent at all times (θ_i) and an idiosyncratic shock to productivity that the agent cannot anticipate ($\epsilon_{it}^Y(h_t, j)$). As with the training choice equation, the parameters of the earnings equation are allowed to vary with training counterfactual choices.

Equations (2) and (3) imply that a common (and constant) factor, θ_i , drives the endogeneity of choices and outcomes. The rest of the unobserved components of the model are independent across choices and outcomes as well as across time and training histories.²¹ Thus, conditional on X_i^Y and X_i^I , θ_i generates all cross-correlations of outcomes and choices: θ_i produces the so-called “ability bias” (Card, 1999), as high-ability individuals may have larger returns from training than low-ability workers. Since ability is unobserved by the econometrician, OLS estimates of equation (3) are biased. Our model extends Card’s logic to a dynamic context: the earnings return to different paths of training choices (h_t, j) vary across the latent ability distribution.

Following a large body of literature on static and dynamic treatment effects, we extract the latent ability factor using a measurement system.²² We do not observe θ_i , but approximate it through a set of variables measuring an initial ability endowment, as these measures of observed prior to labor market entry and training decisions. Let T_k represent a measure of ability, for $k \in \mathcal{K} \equiv \{1, 2, 3, 4\}$. The measurement system is given by:

$$T_{ik} = X_{ik}^T \beta_k^T + \theta_i \lambda_k^T + \epsilon_{ik}^T \quad k \in \mathcal{K}, \quad (4)$$

where X_{ik}^T is a vector of exogenous control variables and $\epsilon_{it}^I(h) \perp \epsilon_{it}^T$ for all $(h, k) \in \mathcal{H} \times \mathcal{K}$, $\epsilon_{it}^Y(j, h) \perp \epsilon_{it}^T$ for all $(t, j, k) \in \mathcal{T} \times \{0, 1\} \times \mathcal{K}$, and $\epsilon_k^T \perp \epsilon_v^T$ for all $k, v \in \mathcal{K}$.

2.2 Identification and Estimation

The identification of the joint distribution of counterfactual outcomes and choices is achieved in three steps.²³ First, we identify the distribution of θ using the measurement system. In this step, the

²¹Formally, we assume that $\epsilon_{it}^I(h) \perp \epsilon_{it}^I(h')$ for all $(t, h) \in \mathcal{T} \times \mathcal{H}$, $\epsilon_{it}^I(h) \perp \epsilon_{it}^Y(j, h)$ for all $(t, j, h) \in \mathcal{T} \times \{0, 1\} \times \mathcal{H}$, and $\epsilon_{it}^Y(j, h) \perp \epsilon_{it}^Y(j', h')$ for all $(t, j, h) \in \mathcal{T} \times \{0, 1\} \times \mathcal{H}$

²²Carneiro et al. (2003), Hansen et al. (2004), Heckman et al. (2006), Heckman and Navarro (2007), Heckman et al. (2013), Attanasio et al. (2015), Agostinelli and Wiswall (2016a), Agostinelli and Wiswall (2016b), and Heckman et al. (2016) are just a few examples.

²³Our identification argument follows by applying Theorems 1 and 2 of that Heckman and Navarro (2007) to any particular node. See also Heckman et al. (2016).

distribution of the factor is nonparametrically identified up to a normalization (which usually takes the form of setting one factor loading $\alpha_k = 1$), and by assuming that unobserved ability is orthogonal to $\mathbf{X}_i \equiv (X_i^Y, X_i^I, \{X_{ik}^T\}_{k \in \mathcal{K}})$. By identifying the joint distribution of counterfactual outcomes, we can also identify the distribution of the treatment effect parameters of interest. Second, we invoke that, conditional on θ and \mathbf{X} , all choices and outcomes are independent—a “matching” assumption, which allows for matching on unobserved dimensions. Third, identification is aided with independent variation in the choice and outcome equations to apply an identification-at-infinity argument. We achieve this condition by using instruments with enough support. We use the average training hours at the firm and all firms within a certain geographical location (“comuna”) where the individual is currently working. Hence, we use these variables as node-specific instruments. An implicit assumption behind these exclusion restrictions is that an individual does not alter her behavior—in a way that could affect her earnings—due to working in a firm that is more or less likely to invest in training.²⁴ We argue that it is unlikely that individuals sort into firms based on the propensity of that firm to let workers participate in training hours, especially since average training hours at the firm is not public information.²⁵ Moreover, Heckman et al. (2016) argue that, for an instrument to meet the standard exclusion and exogeneity restrictions, agents cannot anticipate the value of the instrument in future periods. In our context, this argument implies that hours at the firm and comuna level are revealed to workers at the time of each training decision, rather than in the initial period.

For estimation purposes, we define the sample likelihood as follows. Let Ψ be the vector that collects the set of structural parameters. Given our independence assumptions, the likelihood for a set of \mathcal{I} individuals is given by:

$$\mathcal{L}(\Psi | \mathbf{X}, Y) = \prod_{i \in \mathcal{I}} \left[\int_{\theta} \prod_{k \in \mathcal{K}} f(T_{ik} | X_{ik}^T, \theta) \prod_{t \in \mathcal{T}} \prod_{h_t \in \mathcal{H}_t} \left\{ f(Y_{it} | X_i^Y, \theta) Pr(H_t = 1 | X_i^I, \theta) \right\}^{H_t} dF(\theta) \right],$$

²⁴Related to this issue, Ba et al. (2017) discuss how identification of the effects of training programs in an experimental setting breaks down when individuals anticipate having a subsidized training in future periods and so they change behavior today (for example, by lowering their present employment search intensity).

²⁵For the first training node, we use the number of hours the employing firm used two years prior to the worker’s participation decision (at $t = -1$ in the timing of the model). Similarly, for the second-period decision, we use average hours two years before, at $t = 0$. The correlation between these two variables is 0.397, indicating that a non-negligible share of the workers in our sample change firms in their first two years in the labor market. We explore differential returns to training for firm-stayers and switchers in Appendix C.

where we assume that idiosyncratic shocks in the choice process (equation 2), earnings regression (equation 3), and measurement system (equation 4) follow independent normal distributions. $F(\theta)$ represents the unobserved ability’s cdf. We adopt a flexible approach to estimate $F(\theta)$:

$$\theta \sim \rho_1 N(\tau_1, \sigma_1^2) + \rho_2 N(\tau_2, \sigma_2^2).$$

Given the numeric complexity in estimating the likelihood, we estimate the model by Markov Chain Monte Carlo (MCMC). Using the estimated model, we simulate 20 samples from the original sample, each new sample associated to a different draw from the posterior of distribution of structural parameters.

3 Data Sources, Descriptive Analysis, and Institutional Details

3.1 Data Sources and Institutional Characteristics

In this paper, we examine the effects of a nationwide funding scheme for job training programs called *Franquicia Tributaria* (FT) in Chile. The program funds training courses for a significant number of workers in the country in any given year through a large-scale subsidy for training expenditures undertaken by firms.²⁶ As a result, the program targets formal-sector workers. While all workers are theoretically eligible for the program, the design of the funding scheme implies that those at larger firms are more likely to participate in training through FT.²⁷ Training courses are held off-the-job, in centers managed by private providers.²⁸ There are three types of training courses covered by FT ([Comisión Revisora del Sistema de Capacitación e Intermediación Laboral, 2011](#)): (i) short-term courses, including industry-specific programs (such as learning to operate heavy machinery), general-skills courses (such as Microsoft Excel courses), as well as programs focused on soft skills; (ii) short-term degrees leading to specialization in specific disciplines; (iii) professional

²⁶In 2010, FT funded training courses for 920,688 workers, 12% of the labor force in that year. Using information from Chile’s national accounts and data from [Carrillo et al. \(2018\)](#), we estimate the total cost of FT to be 0.08% of the GDP.

²⁷FT subsidizes firms through a tax exemption, with a cap set at 1% of the firms’ annual payroll. [Carrillo et al. \(2018\)](#) show that the government subsidizes 80% of the total cost of courses, making FT courses relatively inexpensive for firms.

²⁸There are over 15,000 providers in the market. Courses are generally scheduled after-work and on weekends, ensuring they do not interfere with regular work schedules.

conferences.²⁹ The predominant role played by the private sector in training course provision under FT resembles that of the Workforce Investment Act (WIA) in the United States. One of the main goals of the WIA (which replaced the Job Training Partnership Act), was to strengthen the role of the private sector (Barnow and Smith, 2016). In practice, the WIA defined individual training accounts (vouchers for training) by which individuals can choose short-term, off-the-job, training courses held in private providers.³⁰ While the workers targeted by the FT program are different than those in the WIA, we note that our analysis speaks to the effectiveness of short-term courses that are also commonly used in other contexts.

To recover training histories and associated labor market outcomes, we construct a novel database that merges three different sources of information. First, we take advantage of administrative records from *Franquicia Tributaria*. Using this data source, we construct workers' training histories by observing their participation in FT-subsidized courses from 1998 through 2010. We analyze labor market outcomes using information from Chile's Unemployment Insurance (UI) system. UI data registers workers' monthly earnings and the firm of employment for all workers with formal sector contracts. We focus on the worker's main employment stint in each quarter and examine earnings in the first quarter of the year following the training event. For notational simplicity, we refer to the outcome variable as contemporaneous with the training event at time t . Lastly, our final source of information comes from performance in a college-entry examination (PSU), which is a mandatory test for all students who wish to enter a post-secondary institution. We observe PSU scores for all high school graduates who took the test between 2000 and 2007. Using individual identifiers, we recover PSU scores of workers to supplement our data of labor market outcomes and training choices. We work with standardized PSU test scores (computed separately by year). The PSU database also includes information on student's observable characteristics, such as gender, age, parental education, family size, and parental employment at the time of the test.

To circumvent threats to identification and for computational tractability, we restrict our sample in several ways. First, we focus our attention on the returns to multiple job training courses for young workers who are first-time labor entrants. We impose this restriction given that we do not

²⁹Courses of type (i) and (iii) must be of five hours at least while courses type (ii) must be over 100 hours. In 2009, the duration of the average job training course was 19.3 hours (Comisión Revisora del Sistema de Capacitación e Intermediación Laboral, 2011).

³⁰Andersson et al. (2013) find moderate effects of receiving WIA training (plus possibly other types of job-related services) on employment and earnings.

observe training histories before 1998; if training choices depend on prior training decisions, then we would be omitting a relevant variable—past training—in the choice equation.³¹ We use the sample of young workers, identify their first year of employment, and follow their labor market history thereafter—by definition, their job training history in our first period is zero. Second, for tractability, we restrict our analysis to training stints during their first two years in the labor force and examine extensive-margin training decisions on a yearly basis. As a result, workers are trained at most twice during our period of interest. Third, we restrict the sample to individuals who are eligible to participate in training financed by FT—that is, individuals who work in the formal sector. As our analysis of worker self-selection into training requires workers to be able to take part in courses each year if they want to, we limit the sample to individuals who are employed for at least nine months in each of their first two years in the labor force in firms with at least ten employees.³² In this way, we analyze a group of workers who are effectively eligible for training each year. By doing so, we abstract from analyzing effects of training on employment.³³ We focus on labor market entrants from years 2003-2008 and their training choices two years after entry. Our final sample consists of 37,089 workers who meet all of the above criteria.

3.2 Descriptive Statistics

Table 1 presents summary statistics for our sample. 54% of individuals in our sample are women. The age at the time of taking the college entrance exam is in line with the average for the country (close to 18 years old). The average GPA and PSU scores in math and language are largely in line with the national average. The average monthly salary in the first quarter of the second year in the labor force equals 551 dollars, reaching 624 dollars after the second year in the labor market.

Table 2 reports summary statistics across workers’ training participation in their first two years in the labor force. A training group is denoted as (h, h') , where the first and second entries denote training participation in the first and second period, respectively ($h, h' \in \{0, 1\}$). The never-trained group is by far the largest in our sample, representing 61 % of all individuals. This group has a lower

³¹This omission could influence the estimation of the distribution of unobserved heterogeneity. See Heckman (1981) and Arellano and Honore (2001) for a related discussion.

³²The structure of the FT subsidy implies that the cost to medium- and large-sized firms is significantly smaller than for firms with less than ten employees. Since UI data indicates that 90% of formal-sector employment in Chile is at firms with at least ten employees, this restriction is not necessarily binding.

³³Our focus on the earnings dimension makes our paper comparable in scope to the on-the-job training literature.

PSU and GPA than workers in all other groups, with the largest difference appearing relative to always-trained workers. The unconditional earnings differential between these two groups reaches 48 and 45% for the first and second period, respectively. We highlight earnings differentials across workers trained only in the first year relative to second year trainees, as in a world of constant returns to training, earnings differentials should not appear for these workers. In this context, later-trained workers earn higher salaries than early-trainees, despite similar test score performance. These intertemporal earnings differences across groups with similar stock of training suggest the presence of different treatment effects of job training over time.

3.3 Reduced-Form Analysis

The descriptive statistics presented above highlight important differences in baseline characteristics and outcomes across training participation groups. We further explore whether training is associated with increased earnings by first estimating the following panel-data model:

$$Y_{itf} = \delta D_{it} + X'_{it}\beta + \alpha_t + \kappa_i + \varepsilon_{itf}, \quad (5)$$

where Y_{itf} represents the log of earnings for worker i , in period t (year since the individual entered the labor market), working at firm f . D_{it} is a dummy that equals 1 if individual i was trained at time t and zero otherwise, α_t is a year fixed effect and κ_i an individual fixed effect. δ recovers the average effect of training on earnings in period t . This equation represents a common approach followed by the literature on training programs (Card et al., 2010).

Table 3 presents the estimated results following from different versions of equation (5). Column 1 shows that concurrent training participation is associated with an earnings gain in the range of 22%. Since background characteristics may affect both training participation and labor market outcomes, in Column 2, we add control variables, including baseline test score performance. Here, the estimated impact of training remains both economically and statistically significant, reaching 18 percent. Given that OLS estimates might not account for potential sorting on unobserved characteristics, in the third column, we add an individual fixed effect. We find that the estimated impact of job training is significantly lower, falling to 0.9 percent, though remaining statistically significant. This result fits in with previous findings by Frazis and Loewenstein (2007), who show

that that controlling for unobserved heterogeneity through an individual fixed effect largely attenuates the estimated impacts of training in an OLS regression. In the last column, we include firm fixed effects to explore whether the returns to training are explained by workers sorting into higher-quality firms. We do not find major changes in the estimated returns to training.

While the fixed-effect estimation in Table 3 may eliminate selection bias, this regression might not identify a pre-determined parameter of interest. First, [Goodman-Bacon \(2018\)](#) shows that in setting where the timing of treatment varies, the usual fixed effect estimator recovers a weighted average of all possible pairs of the underlying differences-in-differences estimator. Moreover, when treatment effects are not constant, some of these weights might be even negative ([de Chaisemartin and D’Haultfœuille, 2018](#)). In the context of our model of choices and counterfactual outcomes, Appendix A defines and tests the assumptions needed for fixed-effect estimators to recover the average treatment effect of training. A first-difference estimator recovers the ATE of training only if: (i) the earnings returns from training are constant across time and training histories and, (ii) the earnings returns from training do not vary with unobserved ability. In Appendix A, using our model estimates, we find evidence against the null hypothesis of lack of differential gains at conventional significance levels.³⁴ Second, the parallel-trends assumption required for identification in reduced-form analyses may not hold if agents sort into training based on unobserved gains from training. Third, even under a parallel trends assumption and assuming a correctly re-weighted fixed-effect estimator ([Callaway and Sant’Anna, 2018](#); [de Chaisemartin and D’Haultfœuille, 2018](#)), we would still miss potentially important questions for understanding the nature of job training programs: To what extent do workers self-select into training based on unobserved characteristics? What is the medium-term effect of training participation? What is the role of continuation values and dynamic complementarity (substitutability) in the returns to training?

³⁴We present reduced-form evidence of heterogeneous effects of training by estimating an OLS regression of different training histories against quarterly earnings in the first period (Table A.2). Confirming the results in Table 2, always trained students outearn both their never-trained and once-trained peers—yet the main difference appears among workers who are trained only once: late-trainees outearn first-year training participants by 8.1 percent. This analysis is similar to the estimation of multi-valued treatment effects, as in [Cattaneo \(2010\)](#).

4 Model Estimates

This section presents the results from the estimation of our discrete choice model. First, we define the variables included in the implementation of the model. We then present the estimated parameters and analyze the extent of sorting on unobserved ability.

4.1 Model Implementation

Table 4 shows the variables we include in the measurement system, training probit and earnings equation. In the measurement system, we include age at the time of PSU, as it may affect test score performance. To identify the distribution of the latent factor, we rely on four different measures, including language and math PSU test scores, high school GPA, and initial wage at the time of entry. We include the initial wage in the measurement system to better capture baseline productivity prior to training choices. In the choice equations, we use age and, as noted above, we include the average training hours at the firm and all firms within a certain geographical location (*comuna*) where the individual is currently working. Lastly, in the earnings equation, we include a gender dummy, age, and a constant.

4.2 Estimated Parameters

Tables B.1-B.3 show the estimated parameters of our econometric model. In Table B.1, we present the estimated parameters of the measurement system. We find that latent ability loads positively on both test score measures, high school GPA, and the initial salary. To understand the relative contribution of observed characteristics and latent ability vector to test score measures, we present a variance decomposition in Figure 2. The unobserved ability measure explains 62% of the variance in the initial salary, but only 5%, 3% and 4% of the math PSU and verbal PSU scores and high school GPA, respectively. Hence, our measure of unobserved ability is more related to initial labor market productivity than workers' academic performance at high school graduation.³⁵ In Figure 3, we depict the distribution of the unobserved ability factor. As the estimated unconditional distribution of θ does not exhibit considerable deviations from normality, we note that assuming

³⁵Including initial salary as part of the measurement system involves a trade-off in estimating the distribution of ability: it reduces the share of the variance in test scores explained by the unobserved factor but allows to better capture unobserved ability at the time of labor market entry.

a standard normal distribution for the distribution of unobserved ability should not significantly change selection patterns and the estimated returns to training.

Tables B.2 and B.3 present the estimated parameters of the training and earnings equations. Across all choice nodes, women are more likely to participate in training and more skilled individuals are more likely to participate in job training. Moreover, workers in firms with a large share of workers participating in FT courses as well as those in geographic areas with more training course availability are more likely to have participated in training in any period.³⁶ The earnings equations indicate that males outearn women by upwards of 0.07 log points. No discernible pattern arises with respect to age-earnings profile. Latent ability has positive effects on earnings, and this result holds across all training nodes. Lastly, we assess the model’s accuracy in predicting observed outcomes and choices. Figure 4 compares observed and simulated training histories. The model matches training decisions well, both in the first and second year. Table 5 contrasts the means and standard deviations of log wages by year and training choices. Overall, simulated earnings show some differences with observed earnings but these gaps are smaller than 0.07 log points in all but one case.

4.3 Selection on Unobserved Characteristics

Figure 5 compares the density of the unobserved ability for workers choosing different training paths. Training choices are denoted by (h, h') , where h and h' denote participation decisions in periods $t = 1$ and $t = 2$, respectively. We find significant differences in the density of ability across one-time participants, depending on the timing of the decision. While the latent ability for those who choose training in the first period but not in the second $(1, 0)$ almost entirely overlaps with the density of the never-trained group, the distribution of the unobserved factor for those in the $(0, 1)$ group clearly surpasses the never-trained group. The always-trained group $(1, 1)$ has the largest ability stock relative to that of workers following other training paths, resembling differences shown in baseline test scores in Table 2. In fact, the never-trained group trails their always-trained counterparts by 0.83 standard deviations in the latent skill distribution. Overall, we find evidence of selection on unobservables, as higher-skilled workers are more likely to have participated in job

³⁶The only statistically insignificant coefficient corresponds to the one on local-level course availability in the training equation of period $t = 2$, conditional on training in the first period.

training.

5 Treatment effects of Job Training on Earnings

This section presents evidence on the impact of job training on earnings. We define and estimate treatment effects both in the static and dynamic sense. We examine the mechanisms driving the dynamic effects of training, by estimating dynamic complementarity (substitutability) in the context of job training participation as well as continuation values. We also examine heterogeneous impacts across the latent ability distribution.

5.1 Static Analysis

We first present evidence on static treatment effects, which capture the effects of training conditional on reaching a particular choice node. Given that we examine earnings in the quarter following the training event, this parameter recovers the short-term effects of training. We first present evidence on the average treatment effect (ATE). Let $E[\cdot]$ denote the expected value taken with respect to the distribution of $(\mathbf{X}, \theta, \varepsilon)$, where ε is the collection of idiosyncratic shocks determining outcomes and choices $\varepsilon \equiv (\varepsilon_j^I, \varepsilon_s^Y)$. We define the average treatment effect as follows:³⁷

$$ATE_1 \equiv E[Y_{i1}(1) - Y_{i1}(0)], \quad (6)$$

$$ATE_2(h_2) \equiv E[Y_{i2}(h_2, 1) - Y_{i1}(h_2, 0) \mid D_{i1} = h_2] \quad \text{with } h_2 \in \mathcal{H}_2 \equiv \{0, 1\}, \quad (7)$$

where $Y_{i1}(1) - Y_{i1}(0)$ is the effect of first-period training on first-period earnings. $Y_{i2}(h_2, 1) - Y_{i1}(h_2, 0)$ denotes the impact of second-period training on second-period earnings, conditional on a training history $h_2 \in \{0, 1\}$. We can additionally estimate the effects of training conditional on having participated in a particular node. That is, the treatment on the treated (TT) parameter is defined as:

$$TT_1 \equiv E[Y_{i1}(1) - Y_{i1}(0) \mid D_{i1} = 1], \quad (8)$$

$$TT_2(h_2) \equiv E[Y_{i2}(h_2, 1) - Y_{i2}(h_2, 0) \mid D_{i2}(h_2) = 1, D_{i1} = h_2] \quad \text{with } h_2 \in \mathcal{H}_2 \equiv \{0, 1\}. \quad (9)$$

³⁷In our main empirical analysis, we use similar ATE parameters that are conditional on θ .

We present the estimated static returns to job training in Table 6. We estimate the average treatment effect across different training periods and histories.³⁸ The average short-term effect of first-period training on earnings (ATE_1) equals 1.7%. The effect on earnings conditional on first-period training (TT_1) does not differ significantly from ATE_1 . For second-period training, conditional on no training in the first year ($ATE_2(0)$), we find bigger short-term effect on earnings, reaching 3.4%. This effect exceeds those of second-year participation for workers who had been trained in the first period ($ATE_2(1)$), which equals 0.4%. Across both second-period returns, we find that the treatment on the treated parameters ($TT_2(0)$ and $TT_2(1)$) are in the same order of magnitude with the average treatment effect parameters ($ATE_2(0)$ and $ATE_2(1)$). Lastly, we note that while the ATE parameters are positive on average, a considerable share of workers would enjoy a negative return from program participation: first-period training ($Pr(ATE_1 < 0)$) lowers earnings for 45% of workers in our sample.³⁹

The estimated short-run effects of training (ATEs and TTs) are similar to the fixed-effects estimates, presented in Table 3. However, ATEs and TTs do not necessarily coincide with the parameters identified by fixed-effects regressions. The coefficient associated with job training in the fixed-effects model is a weighted average—with potentially negative weights—across the three treatment effect parameters (Callaway and Sant’Anna, 2018; de Chaisemartin and D’Haultfœuille, 2018; Goodman-Bacon, 2018). Furthermore, as we show in Appendix A, the conditions for a first-difference estimator to recover ATE are not met in our sample.

Figures 6 and 7 examine heterogeneous effects of training by estimating a non-linear regression of the estimated treatment effects (ATE and TT) onto the latent skill distribution. In the first year of training (Figure 6), we find larger returns to participation for less-skilled workers, though the returns profile is largely flat across the skill distribution (both for ATE_1 and TT_1). Larger differences appear in the returns to second year training (Figure 7). Among workers who had not been initially trained ($ATE_2(0)$), training has larger effects on earnings for high-skill workers,

³⁸We show 95 percent confidence intervals in brackets. Standard errors are obtained from our simulated sample and capture the sample variation of ATE and TT for a given structural parameters as well as parameters’ uncertainty coming from the estimated posteriors.

³⁹We also find negative earnings returns for a non-negligible share of program participants. Since our econometric model is agnostic about the role of expectations in the decision-making process, we cannot directly distinguish whether the negative returns could be explained through financial regret—individuals do not correctly predict the monetary gains following from training—or through psychic costs—the agent is willing to accept a negative monetary return because training yields non pecuniary benefits.

exceeding 4% for those in the top decile of the skill distribution, while remaining close to 3% for individuals in the bottom decile. For first-period participants ($ATE_2(1)$), the impact of second-period training for less-skilled workers stays around 1%, whereas the effects are not different from zero for workers above the latent skill median.⁴⁰

In short, the estimated static treatment effects reveal heterogeneous impacts across different decision margins, job training histories, and latent ability. As such, the constant effect framework required in fixed-effect estimators is rejected in favor of a model of differential short-term gains from job training.

5.2 Dynamic Treatment Effects of Sequential Training Decisions

While we have so far analyzed short-term treatment effects, job training may also affect medium- and long-term labor market outcomes. In this section, we extend our analysis by defining and estimating dynamic treatment effects of sequential training decisions. Furthermore, we analyze heterogeneous effects across the skill distribution and identify the mechanisms through which job training affects labor market outcomes, by estimating the direct effect and continuation value of training.⁴¹

Our goal is to identify the returns to training participation in terms of the present value of earnings.⁴² Formally, let $\tilde{Y}_1(j)$ be the present value of earnings associated with choosing training option j in period $t = 1$ ($D_1 = j$). Since we work with two time periods, we define $\tilde{Y}_i(j)$ as:

$$\tilde{Y}_i(j) \equiv Y_{i1}(j) + \rho(D_{i2}(j)Y_{i2}(j, 1) + (1 - D_{i2}(j))Y_{i2}(j, 0)), \quad j \in \mathcal{H}_2 \equiv \{0, 1\}$$

where ρ is a discount factor. $D_{i2}(j)$ denotes second-period participation and $Y_{i2}(h, j)$ are earnings for training choice j , given history h .⁴³ The individual-level dynamic treatment effect is thus given

⁴⁰In two training nodes, job training has larger effects on less-skilled individuals. This type of reversed-Roy selection has also been found in other contexts. [Kline and Walters \(2016\)](#) find that children who are less likely to attend Head Start gain more from it, which is consistent with evidence from Germany ([Cornelissen et al., 2017](#)). [Mountjoy \(2018\)](#) finds suggestive evidence that the returns to two-year colleges are larger for those who are less likely to attend, a result in line with findings in the context of post-secondary education in Chile ([Rodriguez et al., 2016](#)).

⁴¹[Heckman and Navarro \(2007\)](#) and [Heckman et al. \(2016\)](#) perform a similar decomposition. We compare our results with those of [Heckman et al. \(2016\)](#).

⁴²Even though we focus on the effect of two training events, our framework can be extended to evaluate the impact of training decisions on life-time earnings.

⁴³In what follows, we transform counterfactual earnings from logs to levels to compute treatment effects.

by $\tilde{Y}_i(1) - \tilde{Y}_i(0)$. This effect can be decomposed as:

$$\begin{aligned} \tilde{Y}_i(1) - \tilde{Y}_i(0) = & \underbrace{(Y_{i1}(1) - Y_{i1}(0))}_{\text{Direct effect (short-term)}} + \underbrace{\rho[Y_{i2}(1,0) - Y_{i2}(0,0)]}_{\text{Direct effect (medium-term)}} + \\ & \underbrace{\rho[D_{i2}(1)(Y_{i2}(1,1) - Y_{i2}(1,0)) - D_{i2}(0)(Y_{i2}(0,1) - Y_{i2}(0,0))]}_{\text{Continuation value}}. \end{aligned} \quad (10)$$

The first two terms of the right-hand side capture the direct effect of training at $t = 1$ (properly discounted). The first term and second represent the direct effect of training on earnings immediately following participation and two years after the event, respectively. The third term corresponds to the continuation value of training, which represents the additional gain (if any) of training in the second-period from training in first period for those who take up training in $t = 2$. This term allows us to explore whether early training increases the returns to additional training stints.⁴⁴

In Table 7, we estimate the following dynamic treatment parameters:

$$DATE \equiv E \left[\left(\tilde{Y}_i(1) - \tilde{Y}_i(0) \right) / \tilde{Y}_i(0) \right], \quad (11)$$

$$DTT \equiv E \left[\left(\tilde{Y}_i(1) - \tilde{Y}_i(0) \right) / \tilde{Y}_i(0) \mid D_{i1} = 1 \right], \quad (12)$$

$$DTUT \equiv E \left[\left(\tilde{Y}_i(1) - \tilde{Y}_i(0) \right) / \tilde{Y}_i(0) \mid D_{i1} = 0 \right], \quad (13)$$

and decompose them into short- and medium-term direct effects and continuation values (equation 10).⁴⁵ The dynamic average treatment effect (DATE) indicates that job training in the first period in the labor force increases the present value of earnings by 3.6%. This result is driven largely by the direct effect of job training, as first-period participation increases earnings (in present-value terms) by 1.1 and 2.5% one and two years after training, respectively.⁴⁶ However, early training lowers the returns to training in the second period—the continuation value—by 0.04%, though the effect is not statistically significant. This result is consistent with the estimates of Table 6, which indicates a lower impact on earnings of second-period training, for those who choose training in the first period compared to those who do not. All in all, the direct effects make up for the bulk

⁴⁴Note that a reduced-form approach is unable to decompose treatment effects into the three components of equation (10). Furthermore, a fixed-effects set-up implicitly assumes away continuation values.

⁴⁵For this exercise, we assume a discount factor $\rho = 1/(1.05)$. In addition, we present our results as % increase from the average baseline present value of earnings ($\tilde{Y}_i(0)$).

⁴⁶The short-term direct effect is equivalent to the static treatment effect presented in Table 6.

of the estimated DATE (101%), whereas continuation value accounts for a small part of the total effect (-1%). As with the static treatment effects, we find similar effects in the dynamic TT and TUT parameters.

In Figure 8, we examine heterogeneous dynamic treatment effects across the latent skill distribution, while decomposing the effects into direct effects and the continuation value. Confirming the results shown in Figure 6, the short-term direct effects of training are decreasing across the skill distribution. In contrast, the medium-term direct effects are larger for higher-skilled workers. Lastly, we find that the continuation value of training decreases across the skill distribution. The combination of these three elements leads to a DATE that is largely flat across the ability distribution.

We note that the estimated continuation value of job training stands in contrast with those estimated in the context of formal schooling. For example, Heckman et al. (2016) find that the bulk of the return to high school graduation (around 70%) and to college enrollment (25%) is explained by the continuation value of schooling, with a larger share for higher-skilled students. On the other hand, we find negative continuation values for high-ability workers, yet the contribution of continuation values to the total return to first-period training is small compared to that of the direct effect components.

5.3 Continuation Values and Dynamic Complementarity (Substitutability)

In this subsection, we examine the conceptual and empirical difference between continuation values and dynamic complementarity (substitutability), two commonly estimated parameters in the context of dynamic skill development. Even though these two parameters are closely related phenomena, they represent different components of the production function of human capital. Since dynamic complementarity (substitutability) captures whether first period training raises (reduces) the return to training in future periods, it is informative about the underlying “technology” of job training.⁴⁷ We can formalize this intuition by noting that continuation values can be decomposed

⁴⁷Continuation values are estimated conditional on second-period training participation. Following equation 11 note that for workers who do not take up second-period training, the continuation value equals zero by construction.

as follows:

$$\begin{aligned}
& \underbrace{D_{i2}(1)(Y_{i2}(1, 1) - Y_{i2}(1, 0)) - D_{i2}(0)(Y_{i2}(0, 1) - Y_{i2}(0, 0))}_{\text{Continuation Value}} = \\
& \quad \underbrace{(Y_{i2}(1, 1) - Y_{i2}(1, 0)) - (Y_{i2}(0, 1) - Y_{i2}(0, 0))}_{\text{Dynamic Complementarity/Substitutability}} + \\
& \quad \underbrace{(1 - D_{i2}(0))(Y_{i2}(1, 1) - Y_{i2}(1, 0)) - (1 - D_{i2}(1))(Y_{i2}(0, 1) - Y_{i2}(0, 0))}_{\text{Dynamic Sorting Gains}} \tag{14}
\end{aligned}$$

The continuation value of training equals dynamic complementarity/substitutability plus a sorting term we label “dynamic sorting gains.” As noted above, dynamic complementarity (substitutability) is informative about the production function of human capital of training across multiple periods (Cunha and Heckman, 2006), exhibiting dynamic complementarity (substitutability) if the return from training in a second period is higher (lower) conditional on first-period participation: $Y_{i2}(1, 1) - Y_{i2}(0, 1) > Y_{i2}(1, 0) - Y_{i2}(0, 0)$. The continuation value is partly driven by workers’ second-period participation decisions: when the continuation value is larger than the dynamic complementarity (substitutability), workers are positively sorting into training participation.

In Figure 9, we present evidence from a local polynomial regression of dynamic complementarity (substitutability) against unobserved ability (equation 14).⁴⁸ First-period job training lowers the return from subsequent participation, independent of second-period decisions: on average, we find evidence of dynamic substitutability (-2.7%), which increases with ability. On the other hand, dynamic sorting gains are positive (1.4% on average) and increasing with ability (4% for the top 10% of the distribution of ability), which fits in with the sorting patterns documented in Section 4.3.

In general, any human capital accumulation process is shaped by dynamic substitutability. In our case, it may arise in various contexts. First, job training could comprise multiple courses covering topics in the same area, with workers starting in a baseline course and subsequently taking part in more complex coursework.⁴⁹ In this setting, dynamic substitutability may appear if the first course delivers critical information for improving job performance, with subsequent courses

⁴⁸As before, we estimate these returns as a share relative to baseline earnings, given by the present value of earnings conditional on no training in the baseline period.

⁴⁹For example, workers first need to learn to operate a computer before taking a course on a specific software.

delivering less value-added. As a result, while early trainees would take the second course in their second year in the labor force, non-trainees would participate in the initial course, which delivers larger returns—thus yielding dynamic substitutability. This result may also appear in a context of course heterogeneity, with individuals choosing the most important (or higher-return) courses early on in their labor market careers and subsequently taking courses delivering lower returns.⁵⁰ Workers could rationally follow such a strategy as the returns to the high-payoff courses could be enjoyed over a longer time horizon (Ben-Porath, 1967).

Dynamic substitutability could also emerge if workers face the decision to either accumulate human capital within the firm, or “outside” the firm, through job training courses. Since the returns to training would then recover the gains from formal training relative to on-the-job training, dynamic substitutability would indicate increasing returns to within-firm learning over time, rather than capturing the underlying technology of job training. In this context, the returns to training should be higher for workers switching firms, as the human capital accumulated in job training would transfer to the new employer, though this would not be the case for prior within-firm learning. To test for this possibility, we replicate our empirical analysis for a subsample of workers who do not switch firms in their first two years in the labor force (“stayers”) in Appendix C.⁵¹ As we find similar static and dynamic returns to training in the stayer subsample (Tables C.2 and C.3), we argue that firm-switching behavior induced through job training participation does not drive the estimated impacts of job training. Given these results, we next examine whether increased job training course availability would improve labor market outcomes for young workers and assess the costs and benefits arising from such an exercise.

6 Dynamic Policy-Relevant Treatment Effects

We have so far focused on estimating parameters such as the average treatment effect and the treatment on the treated. However, these may not necessarily be relevant parameters for policy purposes. For instance, workers induced to change their training choices through a particular policy

⁵⁰In this context, workers would first take courses directly related to their industry or occupation and subsequently participate in foreign language courses, for example.

⁵¹This sample consists of 22,247 of the original 37,089 workers—that is, 60% of our sample never switched firms in the first two years of labor force participation. We do not directly model workers’ firm switching behavior, yet remark that the characteristics and training choices of stayers are largely similar to those in the full sample (Figure C.1 and Table C.1).

may have different observed and unobserved characteristics relative to the average worker, and thus their estimated gains from training would not be captured by average population parameters (Mogstad and Torgovitsky, 2018). In this section, we introduce a framework which allows us to examine the returns to various policy alternatives, decompose the effects across response types, and study dynamic responses to these policies.

6.1 Policy Intervention

To examine the impacts of a potential policy affecting training choices and associated outcomes, we follow the literature that defines policy relevant treatment effects in terms of policy shocks that do not affect marginal treatment effects (Heckman and Vytlacil, 2001; Carneiro et al., 2010; Mogstad et al., 2018; Mogstad and Torgovitsky, 2018). In our context, we study the effect of a policy that affects the net cost of first-period training while leaving fixed the net cost of second-period training. Concretely, we examine the effect of an increase in the number of average hours of FT courses available across *comunas* in period $t = 1$. In practice, our simulation may capture a temporary, unexpected shock to the training industry that increases the number of available courses in the market through a policy intervention.⁵² Even though the policy change only affects net costs of job training in the first period directly, it alters second-period decisions by shifting workers' progression through the training tree depicted in Figure 1. Furthermore, the simulated policy change affects observed outcomes exclusively through training choices, not by influencing counterfactual earnings.

To consider how such a policy would impact training choices, we introduce the following notation. Let $D_{i1}^{a_j}$ be the training choice in period $t = 1$ in a given state of the world a_j . We model the policy change as a shift from a_0 to a_1 , which may lead to a change in training decisions in the first period. For instance, workers who are first-period compliers are captured by $(D_{i1}^{a_1} = 1, D_{i1}^{a_0} = 0)$.⁵³ We return to these definitions below, but note that we assume that this policy does not affect

⁵²This policy change could also take place through a subsidy for FT providers to develop additional courses for first-year labor market participants. Our framework can be extended to the case of policies that shift the cost of training across multiple time periods. Here, we study the simplest case to illustrate the benefits of our model in terms of revealing dynamic policy responses.

⁵³In each policy state a_j , we keep fixed draws of all error terms and parameters from posterior distributions from the baseline model. Thus, differences in choices between a_0 and a_1 stem exclusively from changes in the net utility of training participation through a shift in the local availability of course hours.

counterfactual outcomes: $Y_{i1}(j)^{a_1} = Y_{i1}(j)^{a_0} = Y_{i1}(j)$, $j \in \{0, 1\}$.⁵⁴ Finally, the policy may affect second-period choices through changes in first-period decisions: $D_{i2}^{a_j} = D_{i2}(D_{i1}^{a_j})$.

Both the number of workers affected by the policy and the estimated earnings effects might depend on the magnitude of the intervention, which calls into question the external validity of the LATE of a particular policy change. [Mogstad and Torgovitsky \(2018\)](#) discuss the importance of considering the external validity of estimated treatment effect parameters, and they suggest extrapolating local average treatment effects as a solution for the limited external validity of LATEs. To this end, we incorporate this consideration in our policy simulation by estimating the effect of a policy intervention of varying sizes, considering a 10 and 50 percent expansion in the number of FT-hours available in each *comuna*.⁵⁵

6.2 Counterfactual Choices and Outcomes

Observed earnings outcomes in both periods depend on training choices. Let $Y_{it}^{a_j}$ be observed earnings in period t . For period $t = 1$ and policy state a_j , and given our assumption about the nature of the policy, observed earnings can be expressed in a [Quandt \(1958\)](#) switching regression framework:

$$Y_{i1}^{a_j} \equiv D_{i1}^{a_j} Y_{i1}(1) + (1 - D_{i1}^{a_j}) Y_{i1}(0). \quad (15)$$

Following equation 15, we note that the policy change only affects first-period earnings for workers changing their initial training decision. Similarly, second-period earnings can be expressed as:

$$Y_{i2}^{a_j} \equiv D_{i1}^{a_j} \times [D_{i2}(D_{i1}^{a_j}) \times Y_{i2}(D_{i1}^{a_j}, 1) + (1 - D_{i2}(D_{i1}^{a_j})) \times Y_{i2}(D_{i1}^{a_j}, 0)] + (1 - D_{i1}^{a_j}) \times [D_{i2}(1 - D_{i1}^{a_j}) \times Y_{i2}(1 - D_{i1}^{a_j}, 1) + (1 - D_{i2}(1 - D_{i1}^{a_j})) \times Y_{i2}(1 - D_{i1}^{a_j}, 0)]. \quad (16)$$

In this context, we are interested in the effects of the policy on the present value of earnings. Let $\tilde{Y}_i^{a_j} \equiv Y_{i1}^{a_j} + \rho Y_{i2}^{a_j}$, where ρ is the discount factor and $Y_{i1}^{a_j}$ and $Y_{i2}^{a_j}$ are given in equations (15) and

⁵⁴The simulated policy affects the latent utility associated with the first participation node. As such, whether the policy impacts' workers participation decisions depends on the coefficient associated with the instrument being shifted in the policy simulation.

⁵⁵Since the baseline number of hours per worker in each comuna in the first year equals 0.55 hours, the policy is simulated through an increase in 0.055 course hours, while the 50% increase results in an average increase of 0.275 hours.

(16). The effect of a policy that shifts the net benefits of training choices from a_0 to a_1 is given by $\tilde{Y}_i^{a_1} - \tilde{Y}_i^{a_0}$.

Since the policy only affects $\tilde{Y}_i^{a_1} - \tilde{Y}_i^{a_0}$ through changes in training choices rather than through a direct impact on earnings, we can decompose its effect by identifying agents' response types. The policy intervention may lead workers to change their first-period participation decision (compliers and defiers) or to maintain their baseline choice under policy state a_0 (always-takers and never-takers). Given our implicit monotonicity assumption in equation (2), an increase in first-period course availability will only affect outcomes through first-period compliers—workers for whom $D_{i1}^{a_0} = 0$ and $D_{i1}^{a_1} = 1$. However, this change would may also affect second-period choices, as workers who modified their first-period decision due to the policy change may make different training choices depending on their training history ($D_{i2}^{a_j} = D_{i2}(D_{i1}^{a_j})$). Therefore, the group of first-period compliers can be further divided by second-period responses in four types: compliers-always takers, compliers-compliers, compliers-never takers, and compliers-defiers.⁵⁶ As the policy change does not have an impact on counterfactual earnings, the effect for workers not changing their training decision in either period will not be different from zero. All in all, the effect of the

⁵⁶We note than an alternative policy change affecting the utility of second-period training participation could also impact $\tilde{Y}_i^{a_1} - \tilde{Y}_i^{a_0}$ through workers who did not change their initial participation decision, but who became second-period compliers. As such, estimating the effects of this policy intervention requires considering the impacts on initial period always-takers and never-takers, who became second-period compliers. We therefore note that evaluating policies in a dynamic context contrasts policy-relevant analysis in an static world, where only contemporary compliers are the relevant group (Heckman and Vytlacil, 2001; Mogstad and Torgovitsky, 2018).

policy change on the present value of earnings is given by:⁵⁷

$$\begin{aligned}
& E[\tilde{Y}_i^{a1} - \tilde{Y}_i^{a0} | D_{i1}^{a0} = 0, D_{i1}^{a1} = 1] \\
& = \\
& \underbrace{E[\tilde{Y}_i^{a1} - \tilde{Y}_i^{a0} | D_{i2}(0) = 1, D_{i2}(1) = 1, D_{i1}^{a0} = 0, D_{i1}^{a1} = 1]}_{\text{Compliers, Always-Takers } (W_{CO,AT})} \times \underbrace{\Pr[D_{i2}(0) = 1, D_{i2}(1) = 1 | D_{i1}^{a0} = 0, D_{i1}^{a1} = 1]}_{\text{Probability of Always-Takers}} \\
& + \underbrace{E[\tilde{Y}_i^{a1} - \tilde{Y}_i^{a0} | D_{i2}(0) = 0, D_{i2}(1) = 1, D_{i1}^{a0} = 0, D_{i1}^{a1} = 1]}_{\text{Compliers, Compliers } (W_{CO,CO})} \times \underbrace{\Pr[D_{i2}(0) = 0, D_{i2}(1) = 1 | D_{i1}^{a0} = 0, D_{i1}^{a1} = 1]}_{\text{Probability of Compliers}} \\
& + \underbrace{E[\tilde{Y}_i^{a1} - \tilde{Y}_i^{a0} | D_{i2}(0) = 0, D_{i2}(1) = 0, D_{i1}^{a0} = 0, D_{i1}^{a1} = 1]}_{\text{Compliers, Never-Takers } (W_{CO,NT})} \times \underbrace{\Pr[D_{i2}(0) = 0, D_{i2}(1) = 0 | D_{i1}^{a0} = 0, D_{i1}^{a1} = 1]}_{\text{Probability of Never-Takers}} \\
& + \underbrace{E[\tilde{Y}_i^{a1} - \tilde{Y}_i^{a0} | D_{i2}(0) = 1, D_{i2}(1) = 0, D_{i1}^{a0} = 0, D_{i1}^{a1} = 1]}_{\text{Compliers, Defiers } (W_{CO,DF})} \times \underbrace{\Pr[D_{i2}(0) = 1, D_{i2}(1) = 0 | D_{i1}^{a0} = 0, D_{i1}^{a1} = 1]}_{\text{Probability of Defiers}}
\end{aligned} \tag{17}$$

In this set-up, we can therefore estimate the aggregate effect of the policy on the net present value of earnings for affected workers and examine the impacts across dynamic response groups. For example, the increase in FT-course availability induces compliers-compliers to move from never taking up job training to participating in the two periods. The complier-defier group may arise if a particular sub-set of workers induced to participate in the first period would take up training in the second period had they not been early trainees. These groups might reveal policy-relevant behavior and we directly test for their presence in our empirical analysis. Note that the weights are given by the prevalence of each response type as a share of all workers who change participation decisions due to the policy.

6.3 Results

The first panel of Table 8 presents the share of workers induced to change their training decisions due to the policy change. We find that a 10% increase in course-hour availability would induce 0.5% of the workers in our sample to participate in job training at some point in their first two

⁵⁷In a similar set-up, Heckman et al. (2016) decompose LATE into the effects of augmenting the availability of colleges on earnings for different subgroups affected by the policy in a dynamic-discrete choice model. The difference with our approach is that Heckman and coauthors look at the effects on earnings at one period for groups that are affected at different choice nodes of past periods. Instead, we look at how both choices and outcomes of different periods are affected by the policy.

years in the labor force. The share of “affected” workers expands linearly across program expansion size, since a 50 percent increase in FT course-hours would induce 2.4% of young workers to change their training decision.⁵⁸ Moreover, we find that almost half of all policy compliers come from the complier-never taker group, which captures workers who take up job training only in the first period in response to the policy. Meanwhile, the complier-complier group, who take-up training in both time periods due to the policy change, accounts for 30% of all affected workers. The weights assigned to dynamic response types are largely constant across the two interventions.

In the second panel of Table 8, we show the estimated effect of the policy simulation on the present value of earnings. We find that a small increase in course availability would increase the earnings of affected workers by 3.1%, reaching 3.5% for a 50 percent increase in course-hour availability, though the difference in the estimated dynamic policy relevant treatment effects is not statistically significant. While the impacts are captured by the DPRTE parameter, presented in equation (6.2), the effect sizes fit largely in line with the dynamic average treatment effect (DATE) presented in Table 7.

We find that the simulated policy would have heterogeneous impacts across dynamic response types. For instance, workers who are induced to take-up training in both periods (compliers-compliers) enjoy an earnings increase in the 4% range. We find similar effects for workers who are induced to participate in *only* one training course, reaching 3.2% for compliers-never-takers. The complier-defier group, which captures workers who move up their training choice due to the policy change, faces heterogeneous earnings gains by program size expansion. For small increases in course-hour availability, their earnings would increase by 1.8%, reaching 2.7% for the larger simulated expansion. In Figure 10, we explore whether differences in the density of the unobserved factor across response type groups may account for the heterogeneous earnings impacts documented above. We find that the latent ability density of always-takers dominates that of the other response types, surpassing the ability of compliers-compliers by 0.11 standard deviations, on average, providing evidence that individuals respond to the policy based on the latent ability.⁵⁹ In this context, we examine whether the 50 percent expansion in first-year course hours would deliver heterogeneous

⁵⁸While the program expansion need not have linear effects on training take-up, the empirical evidence shows no indication of non-linear responses by program expansion size.

⁵⁹There are larger differences with the other response types, as the average latent ability of compliers-always takers exceeds that of compliers-never takers by 0.42 standard deviations.

returns by estimating a local polynomial regression of the dynamic policy relevant treatment effect parameter against the distribution of latent ability (Figure 11).⁶⁰ We find largely homogeneous effects across the ability distribution, in the range of 3%.

All in all, we have found that increased course availability would lead a non-negligible share of workers to change their early-career training decisions, while positively affecting their medium-term earnings. However, the heterogeneous impacts across dynamic response types suggest that the positive impacts will vary by the number and the timing of the courses that workers are induced to take up. Since the static framework necessarily overlooks multi-period response types, these results highlight the importance of considering dynamic effects of policy changes.

7 Conclusions

In this paper, we leverage a large government-subsidized program to present the first estimates of repeated participation in job training for first-time labor market entrants. We document dynamic selection patterns and show that high ability workers are more likely to participate in job training early in their careers, though among one-time participants, second-period trainees are more skilled vis-a-vis those trained in the first year. We find that the static returns to job training are positive and significant, though they vary across the timing of the event, training histories and heterogeneously across the latent ability distribution. The dynamic treatment effects indicate larger medium-term gains from early job training, and this effect is fully explained through the direct effect of training. As such, the continuation value of training is not different from zero, standing in contrast with the positive continuation value from schooling (Heckman et al., 2016). We further document the conceptual and the empirical differences between continuation values and dynamic complementarity (substitutability). We find dynamic substitutability of first-year training: early investments decrease the economic returns to later investments. The result is stronger for high-skilled workers. Dynamic substitutability may be explained by the structure of the job training courses examined in this paper, or more generally by the structure of post-schooling human capital accumulation processes. While we cannot test the potential mechanisms formally, we consider our

⁶⁰The optimal policy design depends on the cost function of course hour expansion. While it would be natural to think that costs follows a convex pattern, the Chilean government has recently allowed e-learning courses to be included as part of *Franquicia Tributaria* courses, suggesting that a larger program expansion need not have a larger per unit cost than a small expansion (SENCE, 2015).

results a first step towards understanding the complex dynamic of the returns to training.

Moreover, while estimating a variety of treatment effects allows us to capture the various margins through which training affects labor market outcomes, these returns may not necessarily be actionable upon for workers ([Mogstad and Torgovitsky, 2018](#)). We have therefore examined the estimated impacts of an expansion in course-hour availability for first-time labor market entrants. In this context, we identify dynamic response types and estimated dynamic policy relevant treatment effects, as early-career policy changes may affect workers' subsequent training decisions. While the increase in course availability would lead to a sizable increase in medium-term earnings, the effects are heterogeneous across dynamic response types. As a result, we remark that any policymaker considering an expansion in training courses should take into account the potential impact on workers' subsequent labor market trajectories, rather than focusing solely on short-term outcomes.

References

- Abbring, J. and G. van den Berg (2003). The non-parametric identification of treatment effects in duration models. *Econometrica* 45, 1491–1517.
- Abbring, J. H. and J. J. Heckman (2007). Econometric evaluation of social programs, part iii: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation. *Handbook of econometrics* 6, 5145–5303.
- Acemoglu, D. and D. Autor (2011). Chapter 12 Skills, Tasks and Technologies: Implications for Employment and Earnings. In D. Card and O. Ashenfelter (Eds.), *Handbook of Labor Economics*, Volume 4, pp. 1043–1171.
- Acemoglu, D. and J.-S. Pischke (1999). Beyond becker: training in imperfect labour markets. *The Economic Journal* 109(453), 112–142.
- Agostinelli, F. and M. Wiswall (2016a). Estimating the Technology of Children’s Skill Formation. NBER Working Paper No. 22442.
- Agostinelli, F. and M. Wiswall (2016b). Identification of Dynamic Latent Factor Models: The Implication of Re-Normalization in a Model of Child Development. NBER Working Paper No. 22441.
- Alzúa, M. L., G. Cruces, D. D. T. Nro, and C. Lopez (2014). Youth Training Programs Beyond Employment. Experimental Evidence from Argentina. Documentos de Trabajo No. 177, CED-LAS.
- Andersson, F., H. Holzer, J. Lane, D. Rosenblum, and J. Smith (2013). Does federally-funded job training work? Nonexperimental estimates of WIA training impacts using longitudinal data on workers and firms. NBER Working Paper No. 19446.
- Arellano, M. and B. Honoré (2001). Panel Data Models: Some Recent Developments. In J. Heckman and E. Lazear (Eds.), *Handbook of Econometrics*.
- Ashenfelter, O. (1978). Estimating the effect of training programs on earnings. *The Review of Economics and Statistics* 60(1), 47–57.

- Ashenfelter, O. and D. Card (1985). Using the longitudinal structure of earnings to estimate the effect of training programs. *The Review of Economics and Statistics* 67(4), 648–660.
- Attanasio, O., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (2015). Estimating the Production Function For Human Capital: Results From a Randomized Control Trail in Colombia. NBER Working Paper No. 20965.
- Attanasio, O., A. Guarín, C. Medina, and C. Meghir (2015). Long Term Impacts of Vouchers for Vocational Training: Experimental Evidence for Colombia. NBER Working Paper 21390.
- Attanasio, O., A. Kugler, and C. Meghir (2011). Subsidizing Vocational Training for Disadvantaged Youth in Colombia: Evidence from a Randomized Trial. *American Economic Journal: Applied Economics* 3(3), 188–220.
- Autor, D. H., F. Levy, and R. J. Murnane (2003, nov). *The Quarterly Journal of Economics* 118(4), 1279–1333.
- Ba, B., J. C. Ham, R. J. Lalonde, and X. Li (2017). Estimating (easily interpreted) dynamic training effects from experimental data. *Journal of Labor Economics* 35(S1).
- Barnow, B. S. and J. Smith (2016). Employment and Training Programs. In R. A. Moffitt (Ed.), *Economics of Means-Tested Transfer Programs in the United States, Volume II*. Chicago: The University of Chicago Press.
- Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. *Journal of Political Economy* 75, 352–365.
- Booth, A. L. (1993). Private sector training and graduate earnings. *The Review of Economics and Statistics* 75, 164–170.
- Buhrmann, J. R. (2017). Good enough jobs: Skill mismatch and two-sided heterogeneity in frictional labor markets. Purdue University Economics Working Papers 1287.
- Callaway, B. and P. H. C. Sant’Anna (2018). Difference-in-Differences With Multiple Time Periods and an Application on the Minimum Wage and Employment. DETU Working Papers 1804, Department of Economics, Temple University.

- Card, D. (1999). Chapter 30 The Causal Effect of Education on Earnings. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 3, pp. 1801–1863.
- Card, D., P. Ibararán, F. Regalia, D. Rosas-Shady, and Y. Soares (2011). The Labor Market Impacts of Youth Training in the Dominican Republic. *Journal of Labor Economics* 29(2), 267–300.
- Card, D., J. Kluve, and A. Weber (2010, nov). Active Labour Market Policy Evaluations: A Meta-Analysis. *The Economic Journal* 120(548), F452–F477.
- Carneiro, P., K. T. Hansen, and J. J. Heckman (2003, may). 2001 Lawrence R. Klein Lecture Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice. *International Economic Review* 44(2), 361–422.
- Carneiro, P., J. J. Heckman, and E. Vytlačil (2010, jan). Evaluating Marginal Policy Changes and the Average Effect of Treatment for Individuals at the Margin. *Econometrica* 78(1), 377–394.
- Carrillo, F., M. E. De Diego, and R. Miranda (2018). Caracterización del Funcionamiento de la Franquicia Tributaria. Unpublished Manuscript, Comisin Nacional de Productividad.
- Cattaneo, M. D. (2010). Efficient semiparametric estimation of multi-valued treatment effects under ignorability. *Journal of Econometrics* 155(2), 138–154.
- Comisión Revisora del Sistema de Capacitación e Intermediación Laboral (2011). Informe Final.
- Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2017). Who benefits from universal childcare? Estimating marginal returns to early childcare attendance. Working Paper.
- Cunha, B. F. and J. Heckman (2006). The Technology of Skill Formation. *The American Economic Review* 97(2), 31–47.
- Cunha, F., J. J. Heckman, L. Lochner, and D. V. Masterov (2006). Interpreting the Evidence on Life Cycle Skill Formation. In E. Hanushek, S. Machin, and L. Woessmann (Eds.), *Handbook of the Economics of Education*, Volume 1 of *Handbook of the Economics of Education*, pp. 697–812. Elsevier.

- de Chaisemartin, C. and X. D’Haultfoeuille (2018). Two-way fixed effects estimators with heterogeneous treatment effects. Working Paper.
- Fitzenberger, B., M. Furdas, and C. Sajons (2016). End-of-year spending and the long-run employment effects of training programs for the unemployed. *IZA DP No. 10441*.
- Fitzenberger, B. and R. Völter (2007, aug). Long-run effects of training programs for the unemployed in East Germany. *Labour Economics 14*(4), 730–755.
- Frazis, H. and M. A. Loewenstein (2007). *On-the-Job Training*. Now Publishers, Inc.
- Fredriksson, P. and P. Johansson (2008). Dynamic treatment assignment: The consequences for evaluations using observational data. *Journal of Business and Economic Statistics* (4).
- Goodman-Bacon, A. (2018). Difference-in-Differences with Variation in Treatment Timing. Working Paper.
- Guvenen, F., B. Kuruscu, S. Tanaka, and D. Wiczer (2015). Multidimensional skill mismatch. *Federal Reserve Bank of Minneapolis. Research Department Working Paper 729*.
- Hansen, K. T., J. J. Heckman, and K. J. Mullen (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics 121*(1), 39–98.
- Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998). Characterizing Selection Bias Using Experimental Data. *Econometrica 66*(5), 1017.
- Heckman, J., R. Pinto, and P. Savelyev (2013, oct). Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes. *American Economic Review 103*(6), 2052–2086.
- Heckman, J. J. (1981). Statistical models for discrete panel data. In C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Panel Data with Econometric Applications*, pp. 114–78. MIT Press.
- Heckman, J. J., J. E. Humphries, and G. Veramendi (2016). Dynamic treatment effects. *Journal of Econometrics 191*(2), 276–292.

- Heckman, J. J., R. J. LaLonde, and J. A. Smith (1999). The Economics and Econometrics of Active Labor Market Programs. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labour Economics: Volume 4B*, pp. 1865–2097. Elsevier Masson SAS.
- Heckman, J. J. and S. Navarro (2007). Dynamic discrete choice and dynamic treatment effects. *Journal of Econometrics* 136(2), 341–396.
- Heckman, J. J., J. Stixrud, and S. Urzua (2006, jul). The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *Journal of Labor Economics* 24(3), 411–482.
- Heckman, J. J. and E. Vytlacil (2001). Policy-Relevant Treatment Effects. *The American Economic Review* 91(2), 107–111.
- Ibarrarán, P., J. Kluve, L. Ripani, and D. R. Shady (2015). Experimental Evidence on the Long-Term Impacts of a Youth Training Program. IZA Discussion Paper No. 9136.
- Ingram, B. F. and G. R. Neumann (2006). The returns to skill. *Labour Economics* 13(1), 35–59.
- Keane, M. P., P. E. Todd, and K. I. Wolpin (2011). Chapter 4 The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications. In O. Ashenfelter and D. Card (Eds.), *Handbook of Labor Economics*, Volume 4, pp. 331–461.
- Kline, P. and C. R. Walters (2016, nov). Evaluating Public Programs with Close Substitutes: The Case of Head Start. *The Quarterly Journal of Economics* 131(4), 1795–1848.
- Lechner, M. (2009). Sequential causal models for the evaluation of labor market programs. *Journal of Business & Economic Statistics* 27(1), 71–83.
- Lengermann, P. (1999). How long do the benefits of training last? evidence of long term effects across current and previous employers. *Research in Labor Economics* 19.
- Lise, J., C. Meguir, and J.-M. Robin (2016)). Mismatch, sorting, and wages. *Review of Economic Dynamics* 53(19), 63–87.
- Lynch, L. M. (1992). Private-sector training and the earnings of young workers. *American Economic Review* 82, 299–312.

- Mincer, J. (1974). Schooling, experience, and earnings. *human behavior & social institutions* no. 2.
- Mogstad, M., A. Santos, and A. Torgovitsky (2018). Using Instrumental Variables for Inference about Policy Relevant Treatment Effects. Forthcoming in *Econometrica*.
- Mogstad, M. and A. Torgovitsky (2018). Identification and Extrapolation with Instrumental Variables. Forthcoming in *Annual Review of Economics*.
- Mountjoy, J. (2018). Community Colleges and Upward Mobility. Unpublished manuscript.
- Postel-Vinay, F. and J. Lise (2015). Multidimensional skills, sorting, and human capital accumulation. *Meeting Papers* 386.
- Quandt, R. E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. *Journal of the American Statistical Association* 53(284), 873–880.
- Rodriguez, J., S. Urzua, and L. Reyes (2016, apr). Heterogeneous Economic Returns to Post-Secondary Degrees: Evidence from Chile. *Journal of Human Resources* 51(2), 416–460.
- Saltiel, F., M. Sarzosa, and S. Urzua (2018). Cognitive and Socio-emotional Ability. In G. Jones, J. Johnes, T. Agasisti, and L. Lopez-Torres (Eds.), *Handbook of Contemporary Education Economics*. Edwar Elgar Publishing.
- Sanders, C. and C. Taber (2012, sep). Life-Cycle Wage Growth and Heterogeneous Human Capital. *Annual Review of Economics* 4(1), 399–425.
- SENCE (2015). Resolucion exenta numero 1664.
- SpitzOener, A. (2006, apr). Technical Change, Job Tasks, and Rising Educational Demands: Looking outside the Wage Structure. *Journal of Labor Economics* 24(2), 235–270.
- Veum, J. (1997). Training and job mobility among young workers in the united states. *Journal of Population Economics* 10, 219–233.

Table 1: Summary Statistics

Variable	Mean	(Std. Dev.)
Male	0.46	(0.5)
Age at Graduation	17.81	(0.64)
Math PSU (Standardized)	-0.04	(0.98)
Verbal PSU (Standardized)	-0.03	(0.99)
High School GPA (Standardized)	0.01	(0.99)
Monthly Salary after First Year (USD)	551.04	(333.4)
Monthly Salary after Second Year (USD)	623.96	(388.31)
Observations	37,089	

Notes: Table 1 presents summary statistics of our estimation sample (see Section 3). The dependent variable is the monthly average of earnings in the first quarter following each training stint. For simplicity, we refer to this variable as concurrent with the training decision. Tests scores (Math and Verbal) and high school GPA are standardized across the general population of test-takers to be of mean zero and variance 1.

Table 2: Summary Statistics by Training Node

	Tree = 0,0	Tree = 0,1	Tree = 1,0	Tree = 1,1
Male	0.46	0.46	0.47	0.43
Age at Graduation	17.82	17.79	17.81	17.78
Math PSU (Standardized)	-0.10	0.04	-0.04	0.19
Verbal PSU (Standardized)	-0.07	0.02	-0.03	0.17
High School GPA (Standardized)	-0.01	0.04	0.00	0.14
Monthly Salary after First Year (USD)	506.00	611.15	566.32	732.64
Monthly Salary after Second Year (USD)	572.32	702.36	640.29	820.12
Observations	23675	5306	4360	3748

Notes: Table 2 presents summary statistics of the estimation sample (see Section 3) across different choice nodes. A choice node is denoted as “Tree = h, h' ”, where h, h' denotes training choices for period $t = 1$ and $t = 2$ ($h, h' \in \{0, 1\}$).

Table 3: Reduced-Form Estimates: Returns to Job Training

	(1)	(2)	(3)	(4)
Effect of training	0.223*** (0.005)	0.179*** (0.004)	0.009*** (0.003)	0.013*** (0.002)
OLS	X			
OLS + controls		X		
Individual FE			X	
Individual and firm FE				X
Observations ($N \times T$)	74178	74178	74178	66352

Notes: Table 3 presents regressions of log-earnings against a dummy variable capturing job training participation, college entrance exam performance, high school GPA and age. The dependent variable is the monthly average of earnings in the first quarter following the training period. Column (1) presents OLS regressions without control variables. Column (2) includes PSU test scores, highschool GPA, a gender dummy, age, and age squared. Column (3) computes the first differences estimator. Column (4) uses an individual and firm fixed-effect. p -values are in parenthesis, where * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$.

Table 4: Variables Used in Implementation of the Model

Variables	Earnings Equation	Training Probit	Measurement System
Constant	Yes	Yes	Yes
Gender	Yes	Yes	Yes
Age at Test			Yes
Age in Year t	Yes	Yes	
Cognitive Factor	Yes	Yes	Yes
Average Training Hours at Firm		Yes	
Average <i>Comuna</i> Hours		Yes	

Notes: Table 4 shows the variables used in our empirical model. In the measurement system, we use math and language college entrance test scores, high school GPA and the initial salary and include as the observed measures. Training decisions include gender and age as control variables as well as training-course availability across training decision nodes.

Table 5: Goodness of Fit: Labor Market Outcomes by Training History

Estimate	$Y_1(1)$	$Y_1(0)$	$Y_2(1, 1)$	$Y_2(1, 0)$	$Y_2(0, 1)$	$Y_2(0, 0)$
A. Means						
Actual	6.30 [6.29,6.32]	6.12 [6.12,6.13]	6.55 [6.53,6.57]	6.29 [6.28,6.31]	6.40 [6.39,6.42]	6.20 [6.19,6.20]
Model	6.23	6.14	6.42	6.27	6.35	6.23
B. Standard Deviations						
Actual	0.57	0.52	0.58	0.57	0.55	0.54
Model	0.51	0.50	0.53	0.52	0.53	0.52

Notes: Table 5 shows the means of log earnings by year and training choice from the observed data and the simulated sample. $Y_i(h, h')$ denote earnings in period t , where h and h' represent training choices in the first and second period ($h, h' \in 0, 1$). In brackets, we show a 95% confidence interval on the mean of observed earnings. The second panel shows the observed and estimated standard deviation of earnings across training histories.

Table 6: Static Returns to Job Training (in %)

Treatment effect	$t = 1$	$t = 2 (D_1 = 0)$	$t = 2 (D_1 = 1)$
<i>ATE</i> (percentage points)	1.69 [1.65,1.72]	3.39 [3.31,3.47]	0.43 [0.27,0.58]
<i>TT</i> (percentage points)	1.79 [1.72,1.86]	3.51 [3.31,3.70]	0.35 [0.11,0.60]
$Pr(ATE < 0) \times 100$	45.33 [45.21,45.44]	45.89 [45.76,46.02]	49.46 [49.22,49.71]
$Pr(TT < 0) \times 100$	45.14 [44.90,45.39]	45.66 [45.37,45.96]	49.55 [49.17,49.93]

Notes: Table 6 presented the estimated Average Treatment Effects (ATE) and Treatment on the Treated (TT) parameters along with the share of workers who enjoy negative returns to training across training histories and over time. The first column shows the returns to first-period training ($t = 1$). The second and third columns present the estimated returns to second-period job training, conditional on not participating in first-period training (Column 2) and conditional on first-period participation (Column 3). We show 95% confidence intervals in brackets.

Table 7: Dynamic Returns to First-Period Job Training (in %)

	DATE	DTT	DTUT
Direct Effect (short-term)	1.11 (0.01) [31%]	1.16 (0.02) [32%]	1.10 (0.01) [31%]
Direct Effect (medium-term)	2.51 (0.02) [70%]	2.49 (0.05) [69%]	2.52 (0.02) [70%]
Continuation Value	-0.04 (0.02) [-1%]	-0.05 (0.04) [-1%]	-0.04 (0.02) [-1%]
Total	3.58 (0.02)	3.59 (0.05)	3.58 (0.03)

Notes: Table 7 presents the estimated Dynamic Average Treatment Effects (DATE), Dynamic Treatment on the Treated (DTT) and Dynamic Treatment on the Untreated (DTUT) of first-period training on the present value of earnings. Let $\tilde{Y}_i(j) \equiv Y_{i1}(j) + \rho(D_{i2}(j)Y_{i2}(1, j) + (1 - D_{i2}(j))Y_{i2}(0, j))$ be the present value of earnings for training choice $j \in \{0, 1\}$. The individual-level dynamic treatment effect equals $\tilde{Y}_i(1) - \tilde{Y}_i(0)$, where

$$\tilde{Y}_i(1) - \tilde{Y}_i(0) = \underbrace{(Y_{i1}(1) - Y_{i1}(0))}_{\text{Direct effect (short-term)}} + \underbrace{\rho[Y_{i2}(0, 1) - Y_{i2}(0, 0)]}_{\text{Direct effect (medium-term)}} + \underbrace{\rho[D_{i2}(1)(Y_{i2}(1, 1) - Y_{i2}(0, 1)) - D_{i2}(0)(Y_{i2}(1, 0) - Y_{i2}(0, 0))]}_{\text{Continuation value}}.$$

We show the above decomposition for:

$$\begin{aligned} DATE &\equiv E \left[\left(\tilde{Y}_i(1) - \tilde{Y}_i(0) \right) / \tilde{Y}_i(0) \right], \\ DTT &\equiv E \left[\left(\tilde{Y}_i(1) - \tilde{Y}_i(0) \right) / \tilde{Y}_i(0) \mid D_{i1} = 1 \right], \\ DTUT &\equiv E \left[\left(\tilde{Y}_i(1) - \tilde{Y}_i(0) \right) / \tilde{Y}_i(0) \mid D_{i1} = 0 \right]. \end{aligned}$$

We present DATE, DTT, and DTUT as percentage of mean baseline of the present value of earnings ($E[\tilde{Y}_i(0)]$). We present standard errors in parenthesis and the percentage contribution of each term in brackets.

Table 8: Share of Compliers and Dynamic-Policy Treatment Effects (in %)**Panel A.** Share of Compliers by Intervention Size and Dynamic Response Type Group Weights

Share Compliers		
	Policy: +10%	Policy: +50%
	(1)	(2)
Compliers (<i>CO</i>)	0.47%	2.38%
Weights by type		
Compliers-Always Takers (<i>CO, AT</i>)	0.106	0.105
Compliers-Compliers (<i>CO, CO</i>)	0.297	0.294
Compliers-Never Takers (<i>CO, NT</i>)	0.489	0.487
Compliers-Defiers (<i>CO, DFT</i>)	0.106	0.109

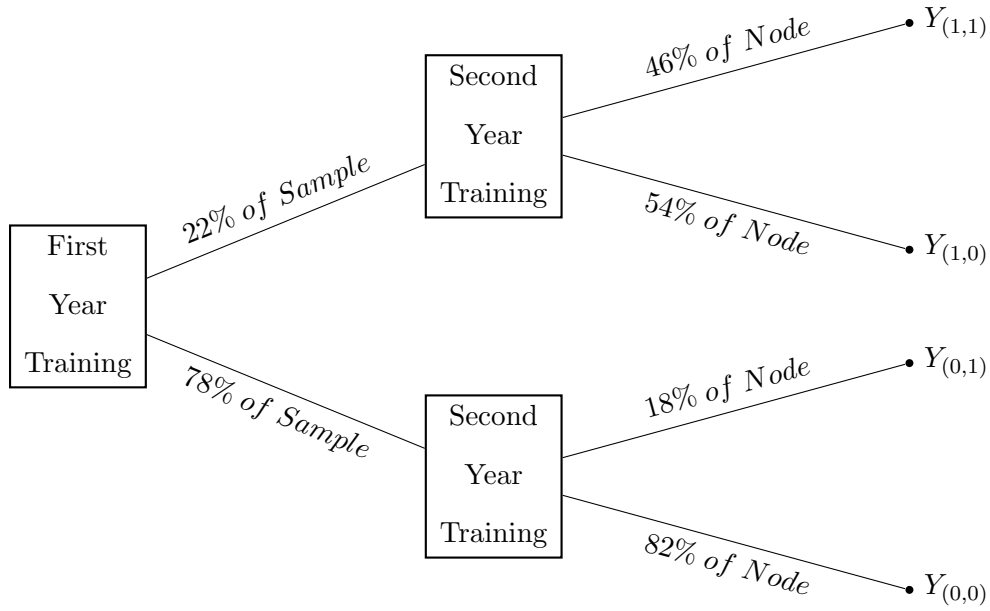
Panel B. Dynamic Policy Relevant Treatment Effect by Response Type

Dynamic Policy Relevant Treatment Effects		
	Policy: +10%	Policy: +50%
	(1)	(2)
Compliers	3.09%	3.51%
	[2.44,3.74]	[3.23,3.79]
Compliers-Always Takers (<i>CO, AT</i>)	1.54%	2.05%
	[-0.39,3.46]	[1.20,2.91]
Compliers-Compliers (<i>CO, CO</i>)	3.97%	4.02%
	[2.81,5.14]	[3.51,4.53]
Compliers-Never Takers (<i>CO, NT</i>)	3.19%	3.71%
	[2.24,4.14]	[3.30,4.12]
Compliers-Defiers (<i>CO, DF</i>)	1.84%	2.65%
	[-0.18,3.87]	[1.78,3.51]

Notes: Table 8 shows policy-relevant treatment effects of two policy shocks: a temporary increase in FT-hours by 10 and 50%. We present the effect of the policy on the present value of earnings conditional on being a W_j type of complier (as a percentage of baseline earnings), a 95% confidence interval of these returns (in brackets), and the proportion of compliers (in % terms). Formally, let $\tilde{Y}_i^{a_j} \equiv Y_{i1}^{a_j} + \rho Y_{i2}^{a_j}$ be the present value of earnings for a policy regime a_j . Following the notation used in the main text, this table presents the probabilities associated with the different groups (weights) and $E[(\tilde{Y}_i^{a_1} - \tilde{Y}_i^{a_0})/\tilde{Y}_i^{a_0} | W_j]$ for $j \in \{(CO, AT), (CO, CO), (CO, NT), (CO, DF), (CO)\}$, where

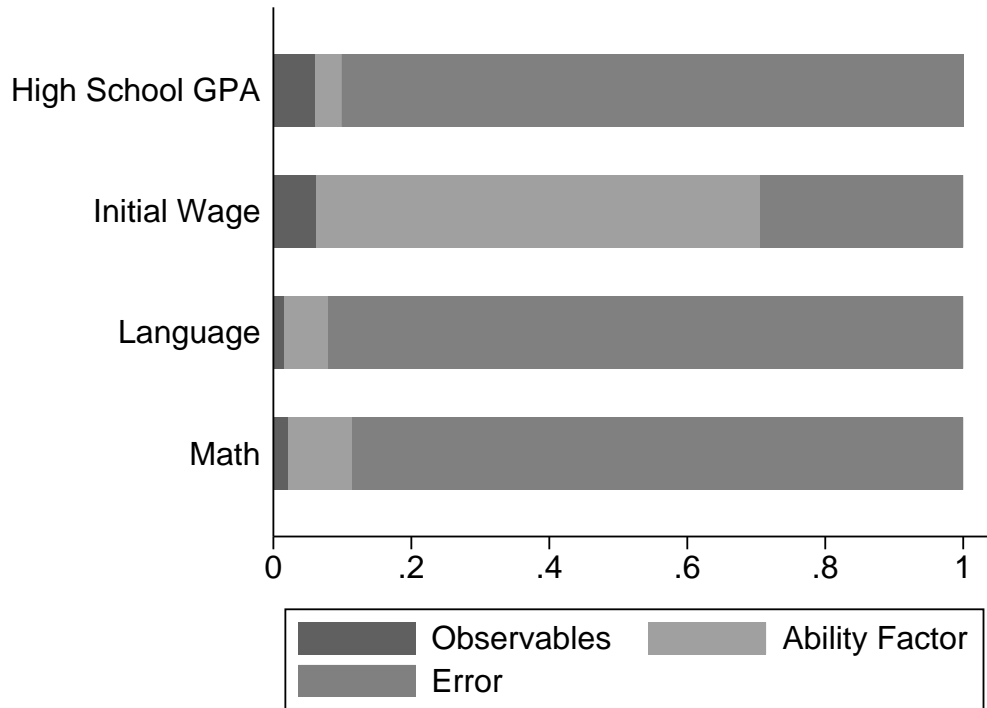
$$\begin{aligned}
CO, AT &\equiv \{D_{i1}^{a_0} = 0, D_{i1}^{a_1} = 1, D_{i2}(0) = 1, D_{i2}(1) = 1\}, \\
CO, CO &\equiv \{D_{i1}^{a_0} = 0, D_{i1}^{a_1} = 1, D_{i2}(0) = 0, D_{i2}(1) = 1\}, \\
CO, NT &\equiv \{D_{i1}^{a_0} = 0, D_{i1}^{a_1} = 1, D_{i2}(0) = 0, D_{i2}(1) = 0\}, \\
CO, DF &\equiv \{D_{i1}^{a_0} = 0, D_{i1}^{a_1} = 1, D_{i2}(0) = 1, D_{i2}(1) = 0\}. \\
CO &\equiv \{D_{i1}^{a_0} = 0, D_{i1}^{a_1} = 1\}.
\end{aligned}$$

Figure 1: Observed Training Choices



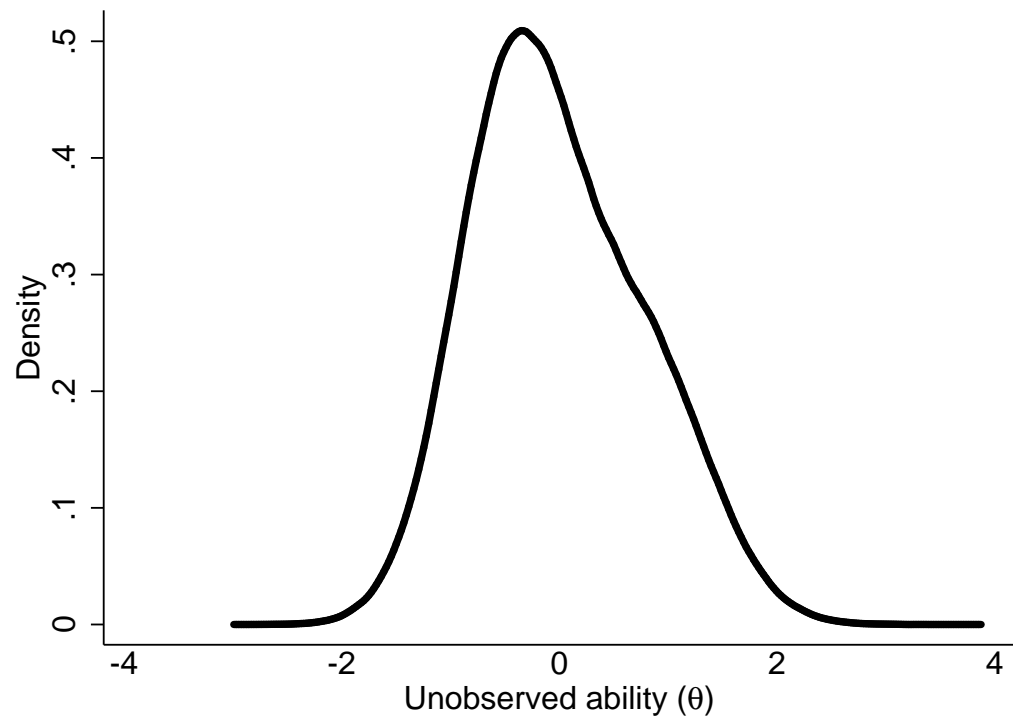
Note: In Figure 1, we present the decision tree through which workers decide whether to participate in training in each of their first two years in the labor force. In each node, we include the observed share of workers in our sample who decide to participate in training.

Figure 2: Measurement System: Variance Decomposition



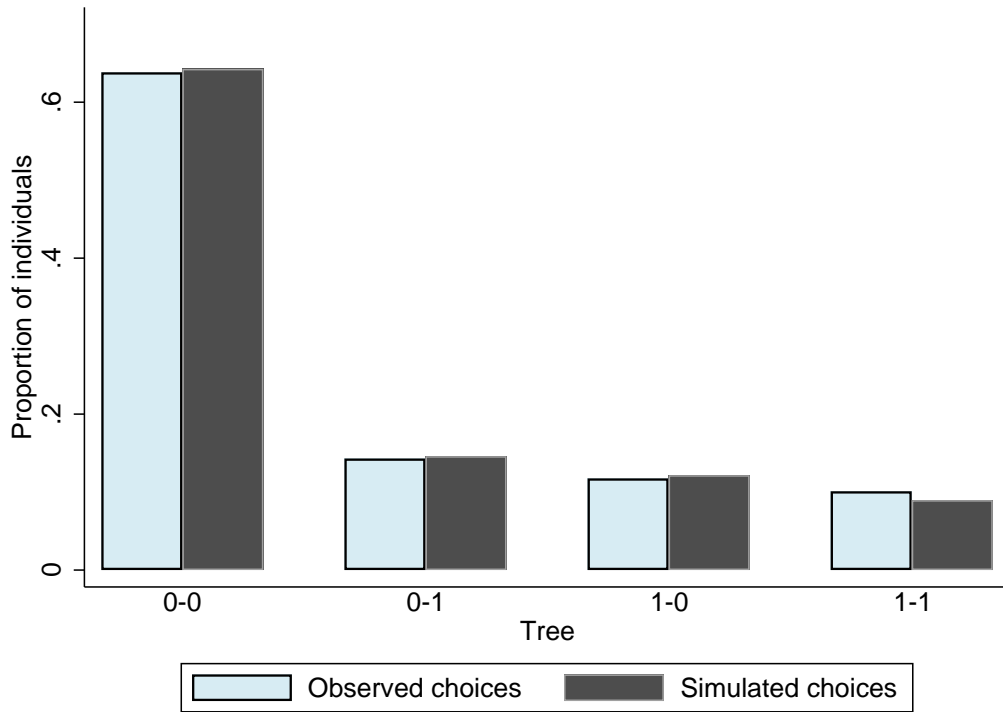
Note: In Figure 2, we show the contribution of each variable to the variance of observed measures using the simulated sample from our model. The “Observables” row indicates the share of the variance of the measure explained by the observed variables: age at the time of test score, gender, parental employment dummy variables, mother’s and father’s education, as well as household size. The “Ability Factor” component shows the proportion of the test score variance explained by unobserved ability. Finally, the “Error term” represents the share of the variance in each observed measure explained by the unobserved idiosyncratic error of the measurement equation.

Figure 3: Distribution of Unobserved Ability



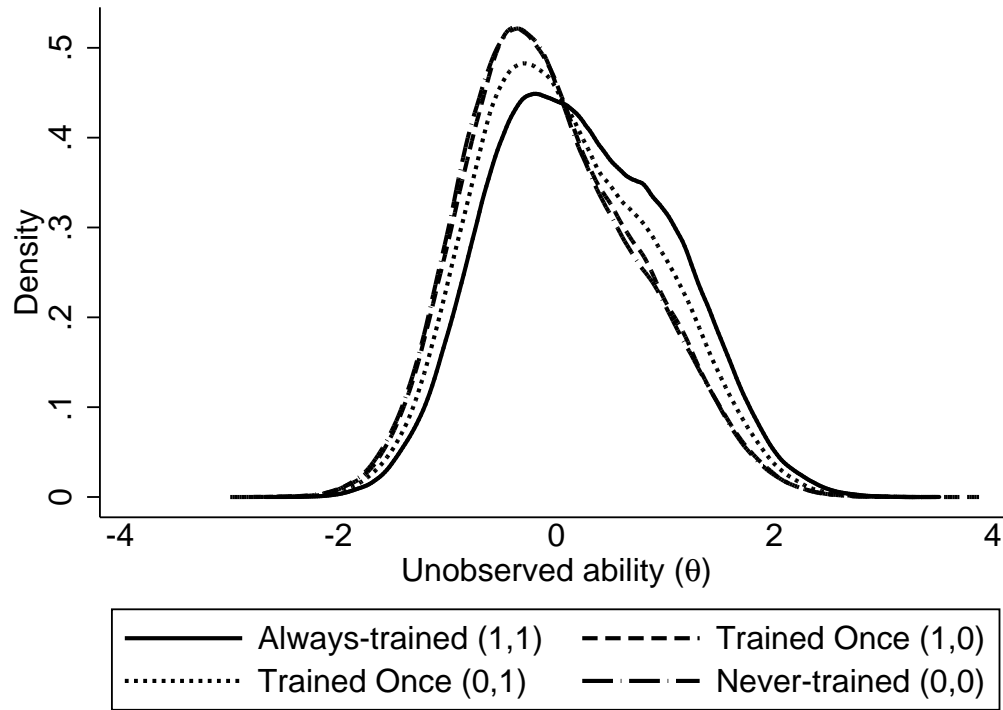
Note: In Figure 3, we show the estimated density of the unobserved ability factor. We obtain this density using the simulated sample from our estimated model. We approximate the distribution of the individual's unobserved ability factor by a mixture of normal distributions.

Figure 4: Goodness of Fit: Training Decisions



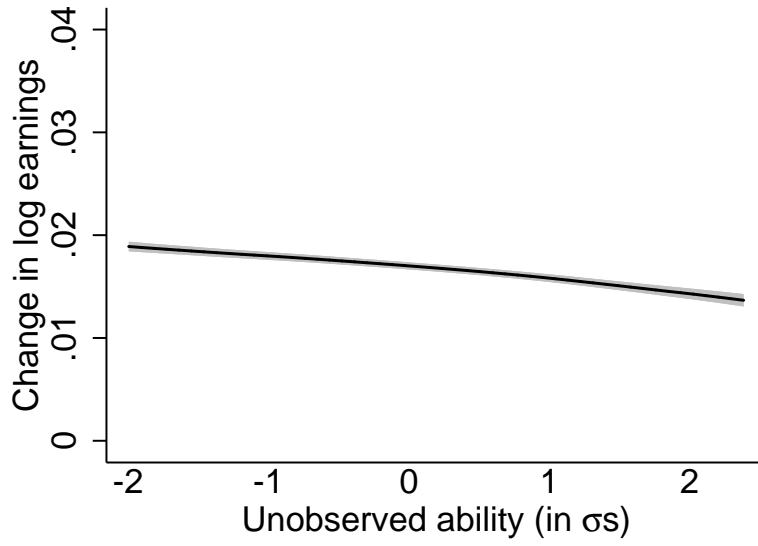
Note: In Figure 4, we compare the share of workers who followed each of the four possible training histories in their first two years in the labor force. A training history is $h - h'$ for $h, h' \in \mathcal{H}$, where the first and second entry indicate training decisions in the first and second period, respectively

Figure 5: Distribution of Unobserved Ability by Training History

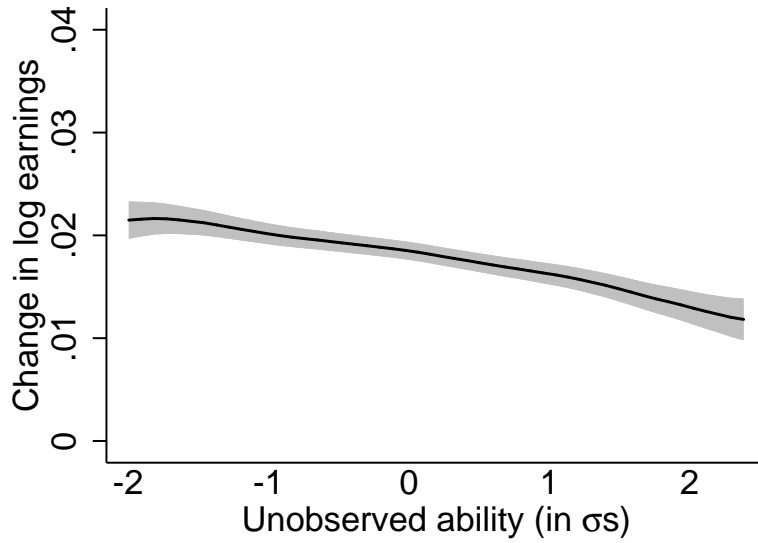


Note: Figure 5 shows the estimated density of unobserved ability for different training paths for workers in their first two years in the labor force. A training path is defined by (h, h') , where h and h' denotes training choices ($h, h' \in \{0, 1\}$) for periods 1 and 2, respectively

Figure 6: Heterogeneous Returns to First-Period Training Participation



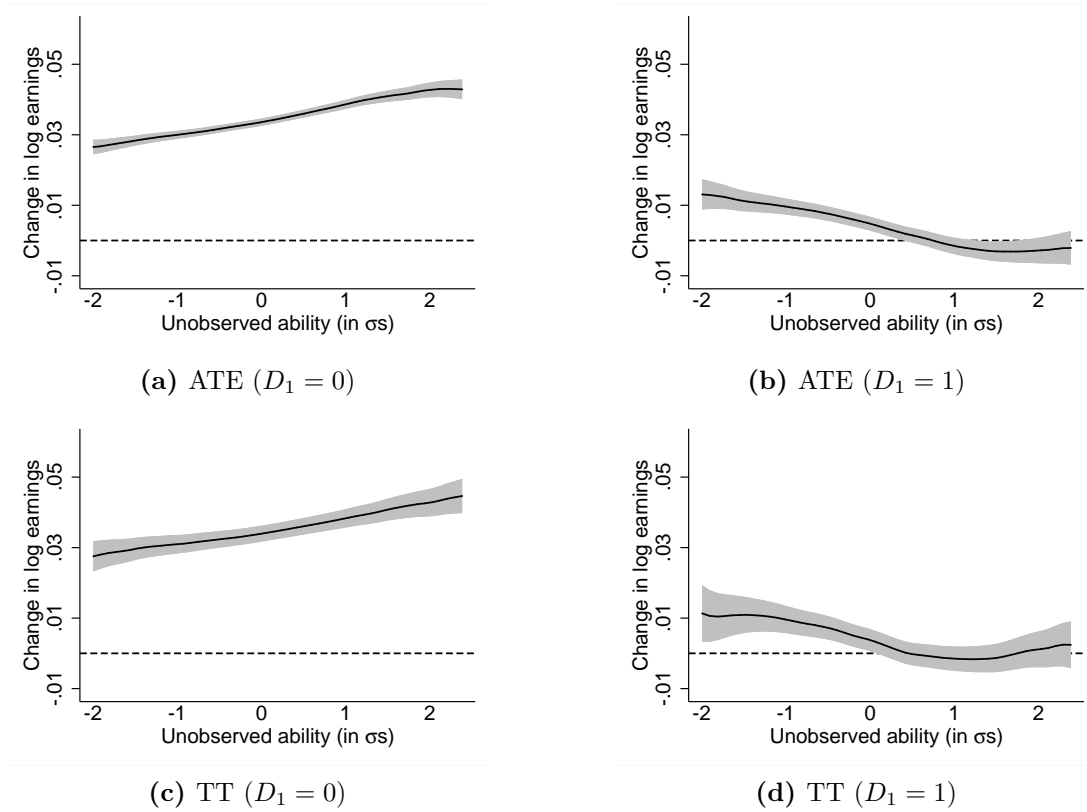
(a) Average Treatment Effect



(b) Treatment on the Treated

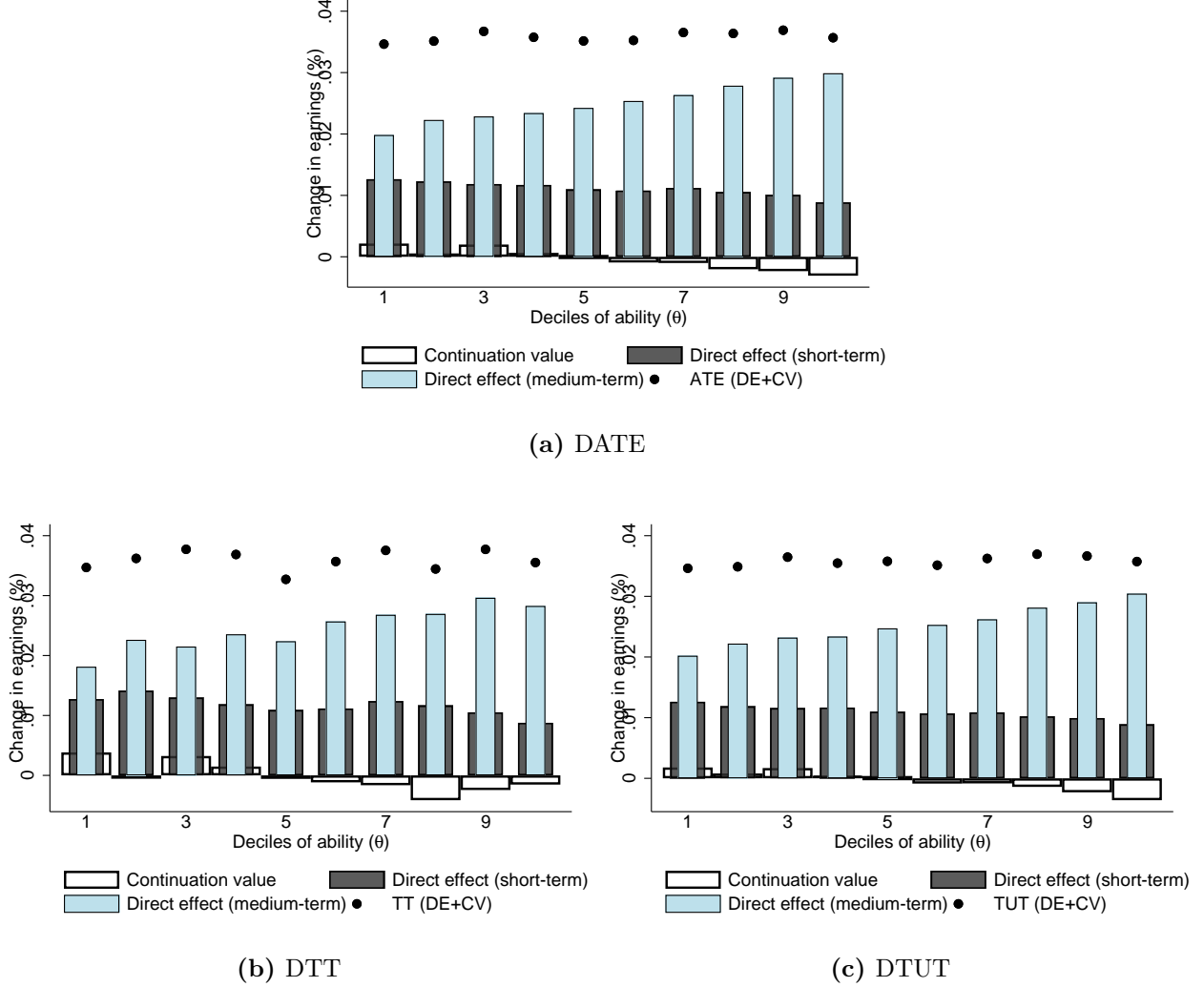
Note: In Figure 6, we estimate local polynomial regressions of the estimated ATE and TT parameters for the first training event (at $t = 1$) against the distribution of unobserved ability.

Figure 7: Heterogeneous Returns to Second-Period Training Participation



Note: In Figure 7, we estimate local polynomial regressions of the estimated average treatment effect and treatment on the treated parameters of second-period job training participation against the distribution of latent ability, conditional on first-period choices ($D_1 \in \{0, 1\}$). The upper-left panel shows, for instance, the average treatment effect of second-period participation for workers who had not taken up training in the first year.

Figure 8: Heterogeneous Dynamic Returns to First-Period Training Participation



Note: In Figure 8, we estimate Dynamic Average Treatment Effects (DATE), Dynamic Treatment on the Treated (DTT) and Dynamic Treatment on the Untreated (DTUT) of training in $t = 1$ on the present value of earnings across deciles of unobserved ability. Let $\tilde{Y}_i(j) \equiv Y_{i1}(j) + \rho(D_{i2}(j)Y_{i2}(1, j) + (1 - D_{i2}(j))Y_{i2}(0, j))$ be the present value of log earnings for training choice $j \in \{0, 1\}$. The individual-level dynamic treatment effect equals $\tilde{Y}_i(1) - \tilde{Y}_i(0)$, where

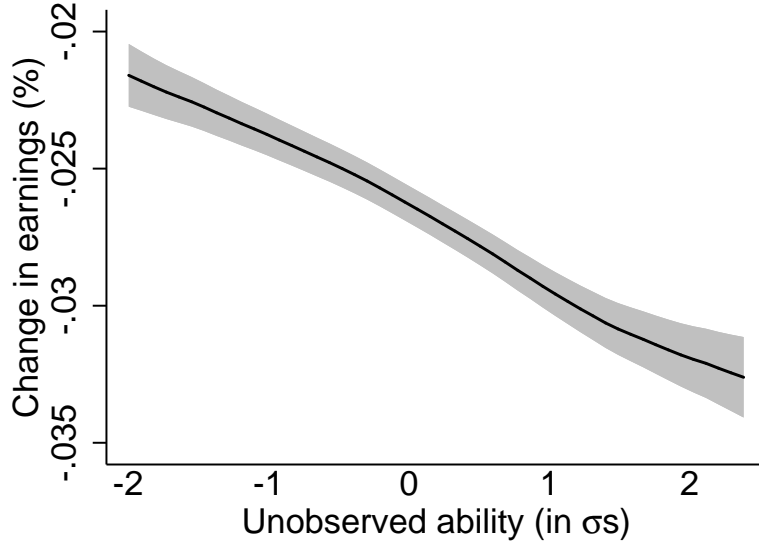
$$\tilde{Y}_i(1) - \tilde{Y}_i(0) = \underbrace{(Y_{i1}(1) - Y_{i1}(0))}_{\text{Direct effect (short-term)}} + \underbrace{\rho[Y_{i2}(0, 1) - Y_{i2}(0, 0)]}_{\text{Direct effect (medium-term)}} + \underbrace{\rho[D_{i2}(1)(Y_{i2}(1, 1) - Y_{i2}(0, 1)) - D_{i2}(0)(Y_{i2}(1, 0) - Y_{i2}(0, 0))]}_{\text{Continuation value}}.$$

We show the above decomposition for:

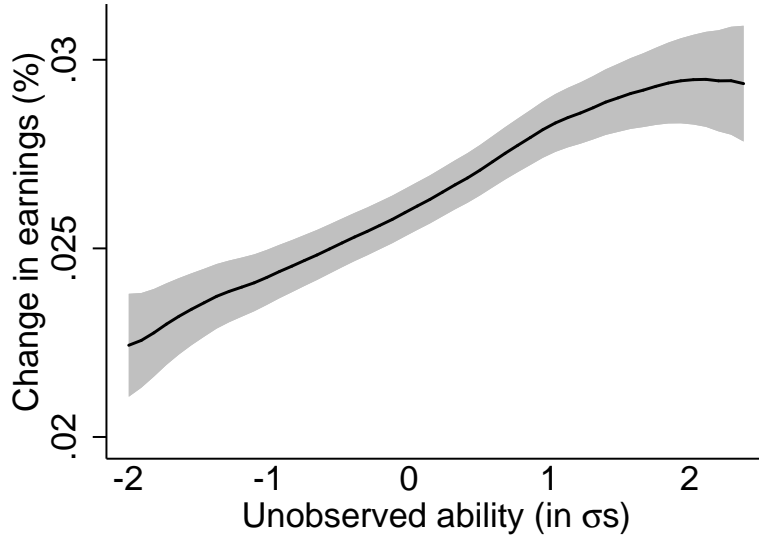
$$\begin{aligned} DATE(\theta) &\equiv E \left[\left(\tilde{Y}_i(1) - \tilde{Y}_i(0) \right) / \tilde{Y}_i(0) \mid \theta \right], \\ DTT(\theta) &\equiv E \left[\left(\tilde{Y}_i(1) - \tilde{Y}_i(0) \right) / \tilde{Y}_i(0) \mid D_{i1} = 1, \theta \right], \\ DTUT(\theta) &\equiv E \left[\left(\tilde{Y}_i(1) - \tilde{Y}_i(0) \right) / \tilde{Y}_i(0) \mid D_{i1} = 0, \theta \right], \end{aligned}$$

where θ captures unobserved ability.

Figure 9: Heterogeneous Dynamic Complementarity (Substitutability) and Dynamic Sorting



(a) Dynamic Complementarity (Substitutability)

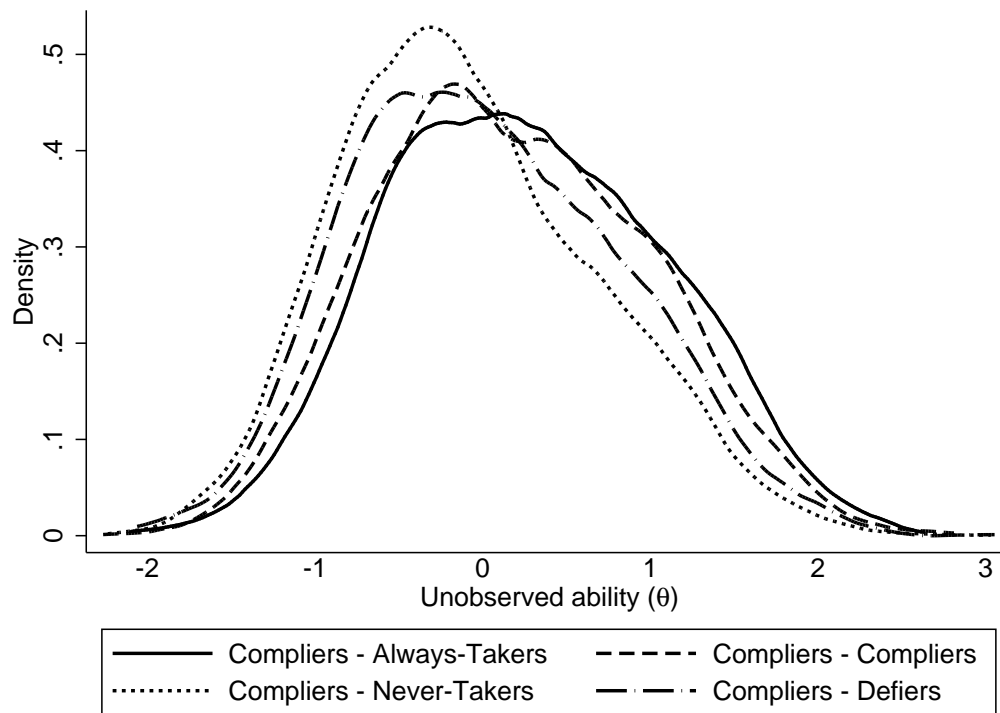


(b) Dynamic Sorting Gains

Note: In Figure 9, we show dynamic complementarity (substitutability) and dynamic sorting gains (as a percentage of mean baseline of the present value of earnings) as a function of unobserved ability, where

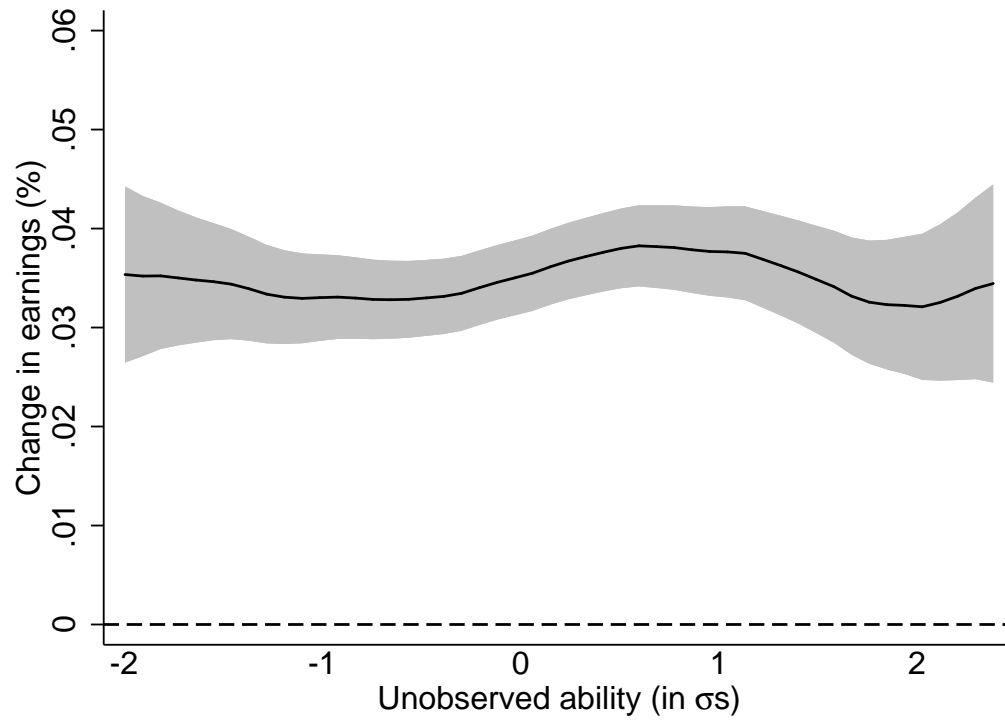
$$\begin{aligned}
 & \underbrace{D_{i2}(1)(Y_{i2}(1,1) - Y_{i2}(0,1)) - D_{i2}(0)(Y_{i2}(1,0) - Y_{i2}(0,0))}_{\text{Continuation value}} = \\
 & \underbrace{(Y_{i2}(1,1) - Y_{i2}(0,1)) - (Y_{i2}(1,0) - Y_{i2}(0,0))}_{\text{Dynamic Complementarity/Substitutability}} + \\
 & \underbrace{(1 - D_{i2}(0))(Y_{i2}(1,1) - Y_{i2}(0,1)) - (1 - D_{i2}(1))(Y_{i2}(1,0) - Y_{i2}(0,0))}_{\text{Dynamic Sorting Gains}}
 \end{aligned}$$

Figure 10: Density of Unobserved Ability by Dynamic Response Types



Note: Figure 10 shows the estimated density of unobserved ability across dynamic response types. The simulated policy reflects a 50% expansion in program course-hours availability in both periods. The density of the latent factor across response types is similar across different program expansion levels.

Figure 11: Heterogeneous Dynamic Policy Relevant Treatment Effects



Note: Figure 11 shows the estimated impact of a 50% expansion in course availability on the present-value of earnings across levels of unobserved ability, captured by the dynamic policy-relevant treatment effect parameter.

Appendices

A FD Estimator and Treatment Effects

Here, we analyze if fixed-effects estimators recover the ATE of training choices. We show that, in the context of our model, the fixed-effect estimator can solve the inconsistency problem of OLS by controlling for unobserved heterogeneity. However, it fails to identify the average treatment effect.

By using the structure of our model, we formally define the ATE. The impact of training for an individual i at period t for a given history h equals $Y_{it}(h_t, 1) - Y_{it}(h_t, 0)$. Let $H_{it}(h_t)$ be an indicator variable that equals 1 if individual i in period t followed training history h and 0 otherwise. The overall average of these individual treatment effects is defined as:

$$\begin{aligned} ATE &\equiv E \left[\sum_{h \in \mathcal{H}_t} H_{it}(h_t) (Y_{it}(h_t, 1) - Y_{it}(h_t, 0)) \right] \\ &= E \left[\sum_{h \in \mathcal{H}_t} H_{it}(h_t) (\mu^Y(h_t, 1) - \mu^Y(h_t, 0) + (\lambda^Y(h_t, 1) - \lambda^Y(h_t, 0))\theta_i) \right], \end{aligned} \quad (\text{A.1})$$

where the expected value operator integrates with respect to i and t . Therefore, ATE is a weighted average of individual treatment effects across periods and different potential training histories.

In a longitudinal data set-up, the analyst's goal is to identify (A.1) using observed data (Y_{it}, D_{it}) , where D_{it} and Y_{it} represent the observed training indicator and outcome variable. As a starting point, consider the following linear regression:

$$Y_{it} = \pi_0 + \pi_1 D_{it} + \xi_{it} \quad \text{for } i = 1, \dots, N \text{ and } t = 1, \dots, T \quad (\text{A.2})$$

where ξ_{it} is an error term. OLS identifies:

$$\delta^{OLS} \equiv \frac{\text{Cov}(Y_{it}, D_{it})}{\text{Var}(D_{it})} = E[Y_{it}|D_{it} = 1] - E[Y_{it}|D_{it} = 0]$$

If the data generating process follows our dynamic model, then potential self-selection into training results in a correlation between ξ_{it} and D_{it} (Ashenfelter and Card, 1985). To see how self-selection affects the reduced-form estimate, first, let us define the following:

$$\mu^Y(j) \equiv \sum_{h \in \mathcal{H}_t} H_{it}(h_t) \mu^Y(h_t, j), \quad \lambda^Y(j) \equiv \sum_{h_t \in \mathcal{H}_t} H_{it}(h_t) \lambda^Y(h_t, j), \quad \epsilon_{it}^Y(j) \equiv \sum_{h_t \in \mathcal{H}_t} H_{it}(h_t) \epsilon_{it}^Y(h_t, j)$$

for $j \in \{0, 1\}$. Second, following the standard switching regression model, we can express observed variables (Y_{it}, D_{it}) as functions of underlying potential outcomes and choices. Observed variables are given by:

$$D_{it} \equiv \sum_{h_t \in \mathcal{H}_t} H_{it}(h_t) D_{it}(h_t), \quad (\text{A.3})$$

$$Y_{it} \equiv \sum_{h_t \in \mathcal{H}_t} H_{it} [D_{it}(h_t) Y_{it}(h_t, 1) + (1 - D_{it}(h_t)) Y_{it}(h_t, 0)]. \quad (\text{A.4})$$

Using the above definitions of D_{it} and Y_{it} , and summing over observed and unobserved parameters across training histories, we have:

$$Y_{it} = \mu^Y(0) + D_{it}(\mu^Y(1) - \mu^Y(0)) + \xi_{it} \quad (\text{A.5})$$

where the unobserved part of the equation is:

$$\xi_{it} \equiv (\lambda^Y(0)\theta_i + \epsilon_{it}^Y(0)) + D_{it}(\epsilon_{it}^Y(1) - \epsilon_{it}^Y(0)) + (\lambda^Y(1) - \lambda^Y(0))\theta_i$$

Thus, the consistency of OLS depends on whether the individuals know their unobserved latent ability endowment (θ_i) and act on it. In this case, net benefits of training ($I_{it}(h_t)$ in equation 1) depend on the unobserved latent ability endowment and the OLS estimator of π_1 (equation A.2) is inconsistent.

Since the inconsistency is originated because the analyst does not observe θ_i —and, thus, it cannot control for it—, one commonly-used approach is to assume that an individual fixed-effect factor drives selection bias. Even though the analyst does not observe θ_i , she can take advantage of the longitudinal nature of the data to eliminate this fixed effect. To see how, reorganize terms in equation (A.5) in the following way:

$$Y_{it} = \overbrace{\mu^Y(0)}^{\pi_0} + D_{it} \overbrace{(\mu^Y(1) - \mu^Y(0))}^{\pi_1} + \overbrace{[(\lambda^Y(1) - \lambda^Y(0))\theta_i + \lambda^Y(0)\theta_i]}^{u_i} + v_{it} \quad (\text{A.6})$$

where $v_{it} \equiv \epsilon_{it}^Y(0) + D_{it}(\epsilon_{it}^Y(1) - \epsilon_{it}^Y(0))$, and note that the equation above is the standard fixed-effect regression. Here, the fixed effect u_i is a function of the unobserved productivity θ_i .

One way of estimating (A.6) is by taking First Differences (FD). Since we observe (Y_{it}, D_{it}) for various periods, we could run OLS on:

$$\Delta Y_{it} = \pi_1 \Delta D_{it} + \Delta v_{it},$$

where the fixed effect has been eliminated and the resulting error term is independent of D_{it} .⁶¹ Therefore, by controlling for u_i , we can recover consistent estimates of π_1 .

Which treatment parameter is the FD estimator recovering? Next, we show that the FD estimator identifies the average treatment effect (that is, $\pi = ATE$ as defined in equation A.1) only if the underlying model of counterfactual outcomes is independent of training histories—which means ignoring the dynamics we laid out in the previous section.

Consider the following assumptions:

Assumption 1. In equation (3), $\mu^Y(h_t, 1) - \mu^Y(h_t, 0) = \pi_1$ for all $h_t \in \mathcal{H}_t$ and $t \in \mathcal{T}$.

Assumption 2. In equation (3), $\lambda^Y(h_t, 1) - \lambda^Y(h_t, 0) = 0$ for all $h_t \in \mathcal{H}_t$ and $t \in \mathcal{T}$.

Assumptions 1 and 2 restrict the gains from training to be constant for all training histories. Assumption 2 rules out any potential gains to treatment for individuals with different levels of unobserved heterogeneity, thereby disregarding the possibility that individuals with higher levels of unobserved ability may enjoy larger returns to training. As a result, these assumptions not only impose strong restrictions within periods, but also across labor market and training histories; Assumptions 1 and 2 imply that the returns to training are equivalent for workers trained at time t

⁶¹When $T = 2$, the fixed-effect estimator is equivalent to the first-differences estimator. In this paper, we focus on the first-differences estimator, but the results are equivalent in the fixed-effect framework.

with training histories $h \in \mathcal{H}_t$ and $h'' \in \mathcal{H}_t$ as well as for workers trained at time $t-1$ with histories $h' \in \mathcal{H}_{t-1}$. Furthermore, these assumptions imply absence of complementarities in the human capital accumulation process—a particularly strong restriction in the context of skills development in the labor market (Mincer, 1974).

Under these assumptions, we can show the following.

Proposition 1. *Suppose outcomes are determined by equation (3) and Assumptions 1 and 2. Then the FD estimator from equation (A.6) follows:*

$$\delta^{FD} = \pi_1 = \mu^Y(h_t, 1) - \mu^Y(h_t, 0) \quad \text{for } h_t \in \mathcal{H}_t, t \in T$$

Proof. Let h and h' denote elements of \mathcal{H}_t and \mathcal{H}_{t-1} . We can express the FD estimator as

$$\begin{aligned} \delta^{FD} = & 1/2 \times E \left[\sum_{h \in \mathcal{H}_t} H_{it}(h) Y_{it}(h, 1) - \sum_{h' \in \mathcal{H}_{t-1}} H_{it-1}(h') Y_{it-1}(h', 0) \right] \\ & - 1/2 \times E \left[\sum_{h \in \mathcal{H}_t} H_{it}(h) Y_{it}(h, 0) - \sum_{h' \in \mathcal{H}_{t-1}} H_{it-1}(h') Y_{it-1}(h', 1) \right] \end{aligned}$$

Given our assumption about counterfactual outcomes (equation 3), the equation above reduces to:

$$\begin{aligned} \delta^{FD} = & 1/2 \times E \left[\sum_{h \in \mathcal{H}_t} H_{it}(h) (\mu^Y(h, 1) + \lambda^Y(h, 1)\theta_i) - \sum_{h' \in \mathcal{H}_{t-1}} H_{it-1}(h') (\mu^Y(h', 0) + \lambda^Y(h', 0)\theta_i) \right] \\ & - 1/2 \times E \left[\sum_{h \in \mathcal{H}_t} H_{it}(h) (\mu^Y(h, 0) + \lambda^Y(h, 0)\theta_i) - \sum_{h' \in \mathcal{H}_{t-1}} H_{it-1}(h') (\mu^Y(h', 1) + \lambda^Y(h', 1)\theta_i) \right], \end{aligned}$$

and collecting terms, we have

$$\begin{aligned} \delta^{FD} = & 1/2 \times E \left[\sum_{h \in \mathcal{H}_t} H_{it}(h) (\mu^Y(h, 1) - \mu^Y(h, 0)) + \sum_{h \in \mathcal{H}_t} H_{it}(h) (\lambda^Y(h, 1) - \lambda^Y(h, 0))\theta_i \right] \\ & + 1/2 \times E \left[\sum_{h' \in \mathcal{H}_{t-1}} H_{it-1}(h') (\mu^Y(h', 1) - \mu^Y(h', 0)) + \sum_{h' \in \mathcal{H}_{t-1}} H_{it-1}(h') (\lambda^Y(h', 1) - \lambda^Y(h', 0))\theta_i \right]. \end{aligned}$$

Reducing the expression above by applying the expected value operator cannot yield ATE, because of two fundamental reasons. First, $H_t(h)$ is, in general, not independent of θ_i , since agents may sort into training at different periods based on their knowledge of θ_i . Second, even if $H_t(h)$ and θ_i were independent, the resulting weighted averages of treatment effects of t and $t-1$ may not necessarily have to be the same. Under assumptions 1 and 2, the second term in each square bracket collapses to 0 and the first term to a constant π_1 . We have then

$$\delta^{FD} = 1/2 \times \pi_1 + 1/2 \times \pi_1 = \pi_1$$

As a result, assumptions 1 and 2 imply that the ATE equals π_1 across all training nodes and training histories (see equation A.1).

□

Under Assumptions 1 and 2, Proposition 1 shows that the FD estimator recovers an average treatment effect which is constant in time and across histories. Hence, the FD recovers our parameter of interest only under the assumption of constant returns to training.

Another potential set of parameters of interest—specially relevant in the context of a dynamic setting—are treatment effects in a dynamic sense. Dynamic treatment effects can be of interest as they allow capturing potential complementarities in the returns to training. For instance, we may be interested in estimating the effect of training for a worker who has received training at time t and $t - 1$ relative to a counter-factual history with no training in either period. Formally, this parameter can be defined as:

$$E[Y_{it}((h', 1), 1) - Y_{it}((h', 0), 0)], \quad h' \in \mathcal{H}_{t-1} \quad (\text{A.7})$$

Is the FD able to identify dynamic treatment effects as defined in equation (A.7)? One can show the FD estimator equals $E[\Delta Y_{it} | \Delta D_{it} = 1] - 1/2 \times E[\Delta Y_{it} | \Delta D_{it} = -1]$. Then, since the FD estimator requires using the sample of workers who have changed their participation decision in periods t and $t - 1$, we cannot use the FD estimator to recover a dynamic treatment effect.

Table A.1 performs formal tests of Assumptions (1) and (2). Panel A presents the parameters associated with assumption (1). Assumption (1) requires that $\mu^Y(1) - \mu^Y(0) = \mu^Y(1, 1) - \mu^Y(1, 0) = \mu^Y(0, 1) - \mu^Y(0, 0) = \pi_1$ for $h \in \mathcal{H}_t$. In the implementation of the dynamic model, $\mu^Y(j, h_t)$ equals $X_i^Y \beta^Y(j, h_t)$ for $j \in \{0, 1\}$ and for all histories h_t . Using a F test, we test the null hypothesis that the three parameters are equal to each other. Our results indicate a strong rejection the null hypothesis (p-value < 0.01). Panel B presents the parameters associated with Assumption (2). In our context, this assumption requires $\lambda(1) - \lambda(0) = \lambda(1, 1) - \lambda(1, 0) = \lambda(0, 1) - \lambda(0, 0) = 0$ for $h \in \mathcal{H}_t$. This assumption implies that higher ability workers cannot enjoy additional returns to training across different time periods and training histories. We conduct the same F test and find the three parameters are statistically different from each other (p-value < 0.01). Therefore, we find evidence against the null hypothesis that fixed-effect estimators recover the ATE.

Table A.1: Testing Assumptions 1 and 2 for Validity of FD Estimators

(a) Assumption 1: $\mu^Y(h_t, 1) - \mu^Y(h_t, 0)$			
	$\mu^Y(1) - \mu^Y(0)$	$\mu^Y(0, 1) - \mu^Y(0, 0)$	$\mu^Y(1, 1) - \mu^Y(1, 0)$
Coefficient	0.017	0.034	0.006
p-value of test	0.000		

(b) Assumption 2: $\lambda^Y(h_t, 1) - \lambda^Y(h_t, 0)$			
	$\mu^Y(1) - \mu^Y(0)$	$\mu^Y(0, 1) - \mu^Y(0, 0)$	$\mu^Y(1, 1) - \mu^Y(1, 0)$
Coefficient	-0.003	0.007	-0.006
p-value of test	0.000		

Notes: We test for assumptions 1 and 2 using a simulated sample drawn from the estimated dynamic model. We show the p-value of the joint hypothesis of equality of parameters.

Table A.2: Estimated Returns to Different Job Training Histories

	Monthly Earnings After $t = 2$	
	(1)	(2)
Second-Period Trainees	0.209*** (0.008)	0.177*** (0.007)
First-Period Trainees	0.0964*** (0.009)	0.0966*** (0.007)
Trained in Both Periods	0.354*** (0.009)	0.266*** (0.008)
Control Variables		✓
<i>Observations</i>	37089	37089
R^2	0.045	0.322

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Standard errors in parentheses. The omitted category is workers who do not participate in training in either year. Table A.2 presents the estimated returns to different training histories on average monthly earnings in the first quarter following second-year training. Column 1 does not include control variables. Column 2 includes baseline college entrance test scores, high school GPA, parental education and workers' gender as control variables.

B Estimated Parameters

Table B.1: Measurement system estimates

	Math PSU	Language PSU	High School GPA	Initial Salary
Constant	3.71 (0.14)	3.28 (0.14)	4.14 (0.14)	-2.58 (0.06)
Male	0.13 (0.01)	-0.09 (0.01)	-0.38 (0.01)	0.11 (0.01)
Age	-0.21 (0.01)	-0.18 (0.01)	-0.22 (0.01)	0.10 (0.00)
Cognitive Factor	0.38 (0.01)	0.31 (0.01)	0.24 (0.01)	1.00 (0.00)
Precision	1.20 (0.01)	1.12 (0.01)	1.14 (0.01)	3.58 (0.03)
Observations	1000	1000	1000	1000

Note: The table displays the estimation results from the measurement system of test scores (equation 4). We obtain these estimates by simulating 1,000 values of parameters using our estimated posterior. The dependent variables are the standardized test score and the initial log earnings at the time of labor market entry. The earnings equation includes year-of-entry dummies. Standard errors are in parentheses. The loading on cognitive factor in the initial earnings modification equation is normalized to 1.

Table B.2: Training Probits: Estimated Parameters

	I	$I(0)$	$I(1)$
Constant	-1.32 (0.04)	-1.40 (0.05)	-0.43 (0.08)
Gender	-0.03 (0.02)	-0.04 (0.02)	-0.09 (0.03)
Firm average training hours	0.60 (0.01)	0.72 (0.01)	0.35 (0.02)
Comuna average training hours	0.28 (0.07)	0.16 (0.08)	-0.03 (0.13)
Cognitive Factor	0.11 (0.01)	0.15 (0.01)	0.27 (0.02)
Precision	1.00 (0.00)	1.00 (0.00)	1.00 (0.00)
Observations	1000	1000	1000

Note: We show the estimated parameters of the training probits (equation 2). We obtain these estimates by simulating 1,000 values of parameters using our estimated posterior. The dependent variable corresponds to the training dummy $I(h)$, for lagged training choice $h_t \in \{0, 1\}$. Standard errors are in parentheses.

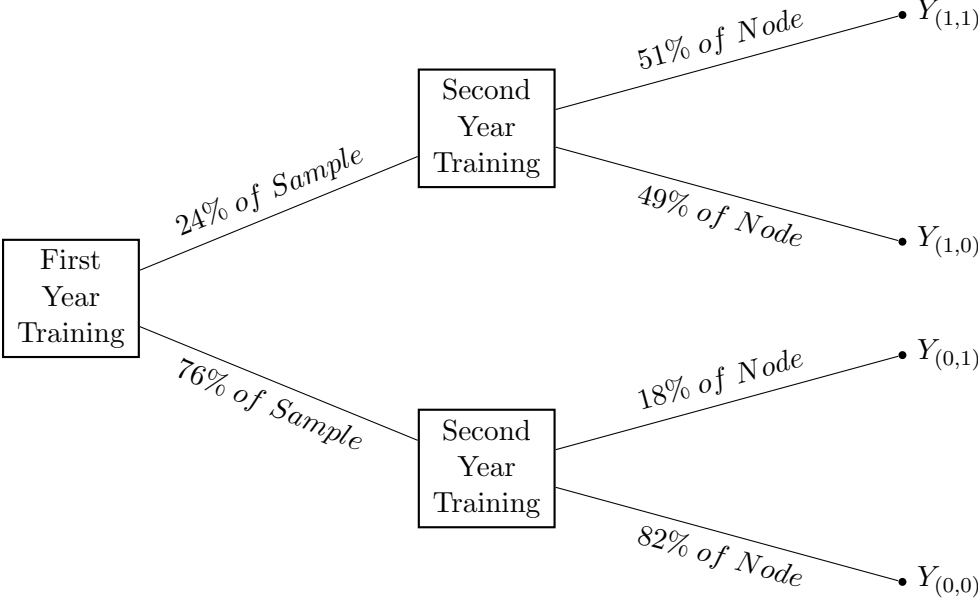
Table B.3: Earnings Equations: Estimated Parameters

	$Y(0)$	$Y(1)$	$Y(0, 0)$	$Y(0, 1)$	$Y(1, 0)$	$Y(1, 1)$
Constant	5.91 (0.17)	4.69 (0.33)	5.82 (0.23)	5.13 (0.38)	4.60 (0.50)	4.88 (0.49)
Gender	0.07 (0.01)	0.07 (0.01)	0.10 (0.01)	0.10 (0.01)	0.10 (0.01)	0.11 (0.01)
Age	-0.03 (0.02)	0.08 (0.03)	-0.02 (0.02)	0.05 (0.04)	0.09 (0.05)	0.06 (0.04)
Age^2	0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	-0.00 (0.00)	0.00 (0.00)
Cognitive Factor	0.59 (0.00)	0.58 (0.00)	0.55 (0.00)	0.56 (0.01)	0.56 (0.01)	0.55 (0.01)
Precision	114.82 (4.95)	87.79 (4.78)	17.68 (0.20)	19.60 (0.44)	18.36 (0.49)	20.85 (0.60)
Observations	1000	1000	1000	1000	1000	1000

Note: We show the estimated parameters of the earnings process (equation 3). We obtain these estimates by simulating 1,000 values of parameters using our estimated posterior. The dependent variable corresponds to average monthly earnings $Y(j, h)$, for training choice $j \in \{0, 1\}$ and lagged training choice $h_t \in \{0, 1\}$. All earnings equations include year-of-entry dummies. Standard errors are in parentheses.

C Returns to Job Training: Stayer Sample

Figure C.1: Observed Training Choices for Stayers



Notes: We present the decision tree through which workers decide whether to participate in training in each of their first two years in the labor force. In each node, we include the observed share of workers in our sample who decide to participate in training.

Table C.1: Summary Statistics: Stayer Sample

Variable	Mean	(Std. Dev.)
Male	0.44	(0.5)
Age at Graduation	17.81	(0.63)
Math PSU (Normalized)	-0.03	(0.98)
Verbal PSU (Normalized)	-0.01	(0.99)
High School GPA (Normalized)	0.04	(0.99)
Monthly Salary after First Year (USD)	567.01	(344.8)
Monthly Salary after Second Year (USD)	623.53	(392.41)
N		22247

Note: The table presents summary statistics of the sample of stayers. The dependent variable is the monthly average of earnings in the first quarter following the training period. For simplicity, we refer to this variable as concurrent with the training decision. Tests scores (Math and Verbal) and high school GPA are standardized across the general population of test-takers to be of mean zero and variance 1.

Table C.2: Static Returns to Training: Stayer Sample

Treatment effect	$t = 1$	$t = 2 (D_1 = 0)$	$t = 2 (D_1 = 1)$
ATE (percentage points)	2.54 [2.51,2.58]	2.95 [2.87,3.04]	0.35 [0.18,0.52]
TT (percentage points)	2.60 [2.53,2.68]	3.02 [2.82,3.22]	0.07 [-0.17,0.31]
$Pr(ATE < 0) \times 100$	41.77 [41.63,41.92]	45.42 [45.25,45.59]	49.51 [49.21,49.82]
$Pr(TT < 0) \times 100$	41.65 [41.34,41.95]	45.14 [44.75,45.53]	49.96 [49.51,50.41]

Notes: We report estimates of the Average Treatment Effects (ATE), Treatment on the Treated (TT), and likelihood of negative treatment effects across three training nodes. The first column show treatment effects after individuals make their first choice ($t = 1$). The second and third columns show estimates for after individuals make a second choice ($t = 2$), conditional on two possible choices in the first period ($D_1 \in \{0, 1\}$). We show 95% confidence intervals in brackets.

Table C.3: Dynamic Returns to Training: Stayer Sample

	DATE	DTT	DTUT
Direct effect (short-term)	1.49 (0.01) [37%]	1.52 (0.02) [38%]	1.48 (0.01) [37%]
Direct effect (medium-term)	2.61 (0.02) [65%]	2.63 (0.04) [66%]	2.60 (0.02) [64%]
Continuation value	-0.07 (0.02) [-2%]	-0.17 (0.04) [-4%]	-0.04 (0.02) [-1%]
Total	4.03 (0.02)	3.97 (0.05)	4.05 (0.03)

Note: We estimate Dynamic Average Treatment Effects (DATE), Dynamic Treatment on the Treated (DTT) and Dynamic Treatment on the Untreated (DTUT) of training in $t = 1$ on the present value of earnings. Let $\tilde{Y}_i(j) \equiv Y_{i1}(j) + \rho(D_{i2}(j)Y_{i2}(1, j) + (1 - D_{i2}(j))Y_{i2}(0, j))$ be the present value of log earnings for training choice $j \in \{0, 1\}$. The individual-level dynamic treatment effect equals $\tilde{Y}_i(1) - \tilde{Y}_i(0)$, where

$$\tilde{Y}_i(1) - \tilde{Y}_i(0) = \underbrace{(Y_{i1}(1) - Y_{i1}(0))}_{\text{Direct effect (short-term)}} + \underbrace{\rho[Y_{i2}(0, 1) - Y_{i2}(0, 0)]}_{\text{Direct effect (medium-term)}} + \underbrace{\rho[D_{i2}(1)(Y_{i2}(1, 1) - Y_{i2}(0, 1)) - D_{i2}(0)(Y_{i2}(1, 0) - Y_{i2}(0, 0))]}_{\text{Continuation value}}.$$

We show the above decomposition for:

$$\begin{aligned} DATE &\equiv E[\tilde{Y}_i(1) - \tilde{Y}_i(0)], \\ DTT &\equiv E[\tilde{Y}_i(1) - \tilde{Y}_i(0) \mid D_{i1} = 1], \\ DTUT &\equiv E[\tilde{Y}_i(1) - \tilde{Y}_i(0) \mid D_{i1} = 0], \end{aligned}$$

We present standard errors in parenthesis and the percentage contribution of each term in brackets.