Optimal Allocation of Seats in the Presence of Peer Effects: Evidence from a Job Training Program*

Matthew D. Baird
RAND Corporation

John Engberg
RAND Corporation

Isaac M. Opper†
RAND Corporation

November 21, 2021

Abstract

We consider the case in which the number of seats in a program is limited, such as a job training program or a supplemental tutoring program, and explore the implications that peer effects have for which individuals should be assigned to the limited seats. In the frequently-studied case in which all applicants are assigned to a group, the average outcome is not changed by shuffling the group assignments if the peer effect is linear in the average composition of peers. However, when there are fewer seats than applicants, the presence of linear-in-means peer effects can dramatically influence the optimal choice of who gets to participate. We illustrate how peer effects impact optimal seat assignment, both under a general social welfare function and under two commonly used social welfare functions. We next use data from a recent job training RCT to provide evidence of large peer effects in the context of job training for disadvantaged adults. Finally, we combine the two results to show that the program's effectiveness varies greatly depending on whether the assignment choices account for or ignore peer effects.

*This research was generously supported by the City of New Orleans from funding from a U.S. Department of Labor Workforce Innovation Fund grant under SGA/DFA PY-13-06, and we are indebted to Brandi Ebanks-Copes, Tammie Washington, and the NOLA Workforce Investment Board for their help on the project. Italo A. Gutierrez conducted the initial data analysis and provided thoughtful advice at the early stages and was therefore instrumental to the paper. We also thank Michael Dinerstein, Lisa Abraham, and conference participants at the 2020 AEFP and 16th IZA and 3rd IZA/CREST Conference on Labor Market Policy Evaluation for their helpful comments on this paper. Finally, we are also grateful to the participants in the study, without whom this type of research would not be possible.

†Baird: mbaird@rand.org; Engberg: engberg@rand.org; Opper: iopper@rand.org

I Introduction

Understanding how individuals are impacted by their peers has long fascinated academics, policy makers, and the wider public. It is therefore unsurprising that despite the difficulty in identifying the effect of an individual's peers on her own outcomes, a large body of research has aimed to estimate the magnitude of these peer effects in a number of different settings and on a number of different outcomes. For example, Mas and Moretti (2009) uses data from a large supermarket chain to show how the productivity of workers' peers affects their own productivity; Sacerdote (2001) looks at how Dartmouth undergraduates' roommates' and dorm-mates' high school GPA impacts their own college GPA; and Opper (2019) shows that the impact that teachers have on their students spills over to impact their students' future peers.

Despite the different contexts, all three of these papers share a common feature: all use a linear-in-means specification, in which the impact of an individual's peers on her outcome can be summarized by the average of the peers' relevant characteristic. While the linear-in-means specification is the most commonly used specification, an important implication of the linear-in-means specification is that even in contexts where peer effects are large, the average outcome across individuals is not impacted by how individuals are sorted into groups; moving an individual who is a positive influence on her peers from Group A to Group B will increase the average outcome in Group B by the exact same amount as it will decrease the average outcome in Group A. This has led many researchers to focus on more complex forms of peer effects, which are asserted to have more policy relevance.[1]

This conclusion, that the magnitude of linear-in-means peer effects do not impact the average outcome of individuals, no longer holds when only a subset of the individuals in a population get a treatment. For example, suppose that a school has enough resources

[1] This sentiment is stated succinctly in Hoxby and Salyer (2006): "The focus on establishing existence and the linear-in-means model in particular have been problematic because neither educational policy-makers nor economists would care much about peer effects if they merely existed and were linear in means. If peer effects were linear in means, then regardless of how peers were arranged, society would have the same average level of outcomes."

to give five students extra group tutoring after school, that students who have low initial achievement are helped the most by the tutoring, and that there is a linear-in-means peer effect in which the tutoring is more effective for an individual when her peers in the program have high initial achievement. In this set-up, there is now a tension over whom to give the spots to. To see why, suppose that four students are assigned to the group tutoring and a principal who cares only about her students' average outcomes is in charge of assigning the last student. Giving the last spot to the lowest achieving remaining student helps him more than giving it to a highest achieving remaining student helps her; however, the students already assigned to the group would benefit more if the last slot was given to the highest achieving student than if it was given to the lowest achieving student. To whom, then, should the last spot be given? The magnitude of the peer effects now has potentially large implications for the efficiency of the tutoring program and impacts the optimal allocation of individuals to the program.

In this paper, we start by formalizing the question described in the paragraph above. To do so, we consider a social planner who can assign individuals to treatment and explore how the magnitude of linear-in-means peer effects influences the optimal treatment assignment. Regardless of planner's utility function, we show that increasing the magnitude of the peer effects will increase the average ability of individuals assigned to treatment in the optimal allocation.[2] We then focus on two specific social planner utility functions, one in which she only cares about the average outcome of the population and one in which she only cares about the minimum outcome in the population. We characterize the optimal treatment assignment under these utility functions and discuss in more detail how the presence of peer effects influences this allocation.

We then turn our attention to the empirical analysis. For all the research into peer effects in education, the literature largely neglects an important part of the American education system: the adult learning/job training sector.[3] This is despite the fact that the sector is

[2]Here we use "ability" to refer to the characteristics that generates the peer effects.

[3]As we discuss below, there is some work on peer effects in the job training context outside of the United States.

a large component of the United States education system; in 2016, for example, over one-quarter of adults in the United States held a non-degree credential (Cronen et al. (2017)). Furthermore, job training programs are arguably one of the more important settings for peer effects, as these programs often are tasked with not only teaching specific job skills but also with shaping the participants' relationship to work and the labor market. Finally, training programs are an ideal context to study for our purposes, as they differ from many other educational settings in that they are not universal and instead are often a scarce resource, with limited slots given only to a fraction of the individuals who are interested.

In our empirical analysis, we use a randomized control trial (RCT) in which applicants were randomly assigned slots in a job training program in Louisiana funded by the United States Department of Labor Workforce Innovation Fund (WIF), authorized by the Workforce Innovation and Opportunity Act, and its predecessor, the Workforce Investment Act (WIA). We match participants in the study to administrative data on individuals' earnings and employment both before and after the training program. We then use this random assignment to show compelling evidence that the average labor market history of other individuals in the class has a statistically significant and economically meaningful impact on the future job prospects of individuals who attend the training. For example, we find that an individual trained in a cohort with average labor market history one standard deviation above the average is 15 percentage points more likely to be employed in the six quarters post-training than an identical individual trained in a cohort with average labor market history one standard deviation below the average. Similarly, we find that the program's impact on the individual trained in a cohort with peers who have higher than average labor market history is approximately \$900 more per quarter than the program's impact on the individual trained in a cohort with peers who have lower than average labor market history. As a comparison, \$900 is around 25% of the average baseline earnings of participants and roughly equal to the average treatment effect across all cohorts.

In our last section, we conclude by combining the theoretical results with the empirical estimates to determine the optimal allocation of individuals to the program. We show that the average impact of the program would be more than double if the slots were allocated optimally rather than randomly. Even more stark is the comparison between a social planner who optimally assigned individuals to treatment compared to one who attempted to assign individuals to the program optimally, but mistakenly ignored the peer effects when doing so. The impact of the program under the social planner who accounted for the peer effects would be tenfold the impact under the social planner who did not account for them.

I.A Related Literature

The idea that peer effects can impact the optimal allocation of individuals to a limited number of treatment slots is not a new one; however, the papers that do explore this generally assume that the network itself is exogenous to treatment.[4] In these models, treating the well-connected individuals in the network will cause more spillovers than treating the poorly connected individuals (e.g. Cai et al. (2015); Banerjee et al. (2013)).[5] In contrast, we consider the optimal treatment allocation in the opposite extreme, where the treatment assignment completely determines an individual's peers and the effect of the treatment itself depends on who is in the treatment group. While seemingly similar to the network literature, our paper therefore focuses on a separate mechanism and so is only superficially related to it.

As a consequence, in many ways our paper is more related to the literature on how non-linear peer effects impacts the optimal grouping of individuals in a more reduced-form model, e.g. Bhattacharya (2009); Hoxby and Salyer (2006); Duflo et al. (2011); Carrell et al. (2013); Booij et al. (2016). By considering a model in which not all individuals are

[4]There are papers that allow for the possibility that the treatment impacts the network, however, they are more concerned with estimation than with determining the optimal targeting of treatment (e.g. Goldsmith-Pinkham and Imbens (2013); Johnsson and Moon (Forthcoming); Comola and Prina (Forthcoming)).

[5]Defining "connected" is quite subtle and determining the precise optimum in the canonical diffusion models is an NP-hard problem (Kempe et al. (2003)). The economic literature generally focuses on centrality measures in the network and studies, among other things, the game theoretic motivation behind the centrality measures (Galeotti et al. (2020); Demange (2017); Bloch et al. (2020)) and how the the centrality measures themselves can be approximated without full knowledge of the network structure (Banerjee et al. (2019); Breza et al. (2020)). Other common research questions include the importance of treating the central individuals in the network (Akbarpour et al. (2020); Belhaj et al. (2020)) and how measurement error and empirical uncertainty impact the efficacy of and approach to optimal targeting (Viviano (2020)).

assigned a group, we show that even linear-in-mean peer effects can have important efficiency implications in addition to equity implications and often impacts which individuals should be assigned to the group.

Our empirical results also speak to the literature on the effectiveness of job training programs in the United States. Most recent studies of adult job training programs in the United States have found them to be relatively effective, but with qualifications.[6] Fortson et al. (2017), for example, use an RCT design and estimate statistically significant increases of more than \$600 a quarter in earnings across 28 DOL-funded training programs. This is similar to the average treatment effect we estimate and also similar to the estimates in Heinrich et al. (2013) that uses a quasi-experimental design to find positive and statistically significant effects across 12 states.[7]

While there has been a large amount of research both on the question of peer effects in traditional colleges and universities and on the question of how effective job training programs are, the question of whether there are peer effects in the job training programs in the United States has not been explored. Instead, the only work on peer effects in a job training program have been in international settings, where the structure of job training programs are often quite different than those in the United States (Field et al., 2016; Lafortune et al., 2018; van den Berg et al., 2021). We combine the two literatures and show both that there are large peer effects in the United States job training setting and that these peer effects mean that the programs' effectiveness depends crucially on the peer composition of the program.

II Conceptual Framework

[6]Van Horn et al. (2015) provide a useful review of this literature.

[7]Andersson et al. (2013) investigate a different set of WIA programs, but find similar effects as do Heckman (2000) and Hollenbeck et al. (2012) in non-WIA job training programs.

II.A Model Specification

We now formally define our model. In it, there is a finite set of N agents, each endowed with an observable attribute that we denote θ_i. Each individual can be assigned to either a treatment condition or control condition. Their assignment is captured by a variable T_i, which equals one if individual i is assigned to the treatment condition and equals zero if individual i is assigned to the control condition. We focus here for simplicity on the case where there is a single treatment group; however, the conclusions do not change if those who are treated are then assigned to smaller groups.

This treatment assignment is important as it affects each individual's outcome, which we denote as y_i for individual i. We define $y_i(0)$ as individual i's outcome when she is assigned to the control condition. If individual i is instead assigned to the treatment condition her outcome is $y_i(0) + \tau_i$. Given our notation, we can parsimoniously define individual i's outcome as:

$$y_i = y_i(0) + \tau_i \cdot T_i \tag{1}$$

We will assume that the individual treatment effect τ_i depends on both i's own attributes and on the attributes of the other individuals who are assigned to the treatment condition (i.e. peer effects). More formally, we will assume that:

$$\tau_i = \tau(\theta_i) + \gamma \cdot \frac{1}{T} \sum_{\forall j} \theta_j T_j \tag{2}$$

where $\tau(\theta_i)$ is independent of who else is assigned to the treatment and T is the number of individuals assigned to the treatment.[8] Note that the main assumption here is that while we allow for there to be peer effects, we restrict them to be linear-in-means peer effect. In the above specification, we also assume that the individual component of treatment effect heterogeneity is a function of θ_i; however, this is not a critical assumption and the

[8] The fact that the peer effect term includes an individual's own characteristic is innocuous, as we could re-write this as a leave-one-out specification for a slight re-parameterization of $\tau(\theta_i)$ and γ. We opt for this specification in the theory section as it simplifies the algebra.

main conclusions hold even when allowing for additional dimensions of treatment effect heterogeneity.

For the theoretical results, we do not make any further assumptions about the relationship between θ_i and $y_i(0)$ or about the shape of the $\tau(\theta_i)$ function. In many peer effects models, it is assumed that an individual's effect on their classmates is proportional to their "ability," for example, to their untreated outcome $y_i(0)$. It might also be tempting to assume that a person's contribution to their classmates' learning is related to their own ability to learn in isolation, that is, to their own $\tau(\theta_i)$. For the theoretical analysis, we leave open both possibilities and define the contribution to classmates' learning as a distinct attribute θ_i, which may or may not be related to $y_i(0)$.

Given this model, we consider the perspective of a social planner who must assign exactly $T < N$ individuals to the treatment. For example, this could be a principal who decides which T students attend an after-school tutoring program, a public agency selecting which applicants will receive job training (as in our application), or more generally any case where there is oversubscription to a treatment in which treated individuals affect each others' outcomes and control individuals are unaffected. Formally, denoting the vector of treatment assignments as $\mathbf{T}$ and writing the social planner's utility function[9] as: $U(y_1, y_2, ..., y_N)$, we get that her optimal decision is:

$$\mathbf{T^{opt}} = \arg\max_{\mathbf{T}} U\Big(y_1, y_2, ..., y_N\Big) \tag{3}$$

$$= \arg\max_{\mathbf{T}} U\Big(y_1(0) + \tau_1 T_1, y_2(0) + \tau_2 T_2, ..., y_N(0) + \tau_N T_N\Big) \tag{4}$$

We first consider the impact of peer effects on treatment assignment under a general social welfare function and then consider in more detail how the social planners' decision depends on the presence of peer effects under two extreme circumstances, one in which she cares only about the average outcome of the individuals and the other in which she only cares about the minimum outcome.

[9] We use the term "social welfare function" and "social planners' utility function" interchangeably throughout the paper.

II.B General Result

We start with a general social welfare function, in which our only assumption is that $U\big(y_1, y_2, ..., y_N\big)$ is weakly increasing in each of the y_i's. We then ask how the optimal treatment allocation changes as peer effects become more important. To do so, we need to formally define what is meant by the peer effects becoming "more important." One natural approach is to consider what happens when γ in Equation (2) increases. However, increasing γ affects the optimal treatment allocation in two ways: it both increases the importance of peer effects and directly impacts y_i for those who are treated; this is similar to the idea that changing prices affects the buyers' decision via both an income effect and a substitution effect. To isolate the increasing importance of peer effects, we consider a *compensated change* in γ, in which as γ increases a constant is added to $\tau(\theta_i)$ such that y_i is constant for everyone if there is no change in assignment.

To formally define a compensated change, we add a bit more notation by using $T_i^{opt}(\gamma)$ to indicate whether individual i is treated under the optimal allocation when the peer effect parameter is γ and $\mathbf{T^{opt}}(\gamma)$ to indicate the optimal vector of treatment assignments under γ.[10] We then make the following definition:

Definition. *In a* ***compensated change*** *of size Δ from a baseline γ of $\tilde{\gamma}$, individual i's treatment effect changes from:*

$$\tau_i = \tau(\theta_i) + \tilde{\gamma} \cdot \frac{1}{T} \sum_{\forall j} \theta_j T_j$$

to

$$\tau_i = \tau(\theta_i) + (\tilde{\gamma} + \Delta) \cdot \frac{1}{T} \sum_{\forall j} \theta_j T_j - C$$

where C is a constant equal to $\Delta \cdot \frac{1}{T} \sum_{\forall j} \theta_j T_j^{opt}(\tilde{\gamma})$.

Thus there are two changes to τ_i: first, $\tilde{\gamma}$ increases to $\tilde{\gamma} + \Delta$; second, a constant equal

[10] Throughout we will assume that there is a unique choice of treatment assignment that is optimal. Doing so allows us to write the comparison more easily, without requiring us to define set ordering. We highlight in the proof where we use this assumption and how the proof could be adjusted without this assumption.

to $\Delta \cdot \frac{1}{T}\sum_{\forall j}\theta_j T_j^{opt}(\tilde{\gamma})$ is subtracted from $\tau(\theta_i)$, ensuring that τ_i does not change for any individual if the treatment assignment remains at $\mathbf{T^{opt}}(\tilde{\gamma})$.

In a compensated change of size Δ, the outcomes stay constant for each individual absent a change in treatment assignment; however, the increased importance of peer effects impacts the social planner's trade-offs when considering alternative treatment assignments and so the optimal treatment assignment can change. In the following theorem, which we prove in Appendix A, we formalize the idea that if the treatment assignment does change, it changes in a way that raises the average value of θ_i among individuals assigned to the treatment.

Theorem 1. *For a compensated change from any γ to $\gamma + \Delta$ with $\Delta > 0$, the average θ_i of the treated individuals after the change is no lower than before the change, i.e.:*

$$\frac{1}{T}\sum_{\forall j}\theta_j T_j^{opt}(\gamma+\Delta) \geq \frac{1}{T}\sum_{\forall j}\theta_j T_j^{opt}(\gamma) \tag{5}$$

In other words, if we assume θ_i to be a measure of underlying ability, then the proof states that the average ability of the optimally chosen set of treated individuals will increase in situations where peer effects have a greater impact on treatment effects. While quite intuitive and straightforward to prove, it is worth emphasizing the generality of the result. We placed no meaningful restrictions on the social welfare function, so the result is true regardless of the social planner's preferences. Similarly, while the mechanism that generates the peer effects have many interesting policy implications, the general result that larger peer effects imply that a social planner should assign individuals with higher baseline ability to treatment is true regardless of the mechanisms generating the peer effects.

II.C Maximizing the Average Outcome

We next consider a case when the social planner cares only about maximizing the average outcome of the population, or:

$$U\Big(y_1, y_2, ..., y_N\Big) = \frac{1}{N}\sum_{\forall i} y_i \tag{6}$$

Since the $y_i(0)$ terms are unaffected by the allocation, maximizing the average outcome of the population is equivalent to maximizing the total treatment effect (TTE). Using Equation (2), we can write the TTE of a particular allocation as:

$$\text{TTE}(\mathbf{T}) = \sum_{\forall i} \tau_i \cdot T_i \tag{7}$$

$$= \sum_{\forall i} \left(\tau(\theta_i) + \gamma \cdot \frac{1}{T}\sum_{\forall j} \theta_j T_j\right) T_i \tag{8}$$

Finally, noting that the peer effect term $\gamma \cdot \frac{1}{T}\sum_{\forall j} \theta_j T_j$ is identical for all individuals, we can make one final re-write to show the total treatment effect of a particular allocation is:

$$\text{TTE}(\mathbf{T}) = \sum_{\forall i} \tau(\theta_i) T_i + T \cdot \gamma \cdot \frac{1}{T} \cdot \sum_{\forall j} \theta_j T_j \tag{9}$$

$$= \sum_{\forall i} \Big(\tau(\theta_i) + \gamma \cdot \theta_i\Big) \cdot T_i \tag{10}$$

This simple expression in Equation (10) implies the most efficient allocation is therefore to give the treatment to the T individuals with the highest values of $\tau(\theta_i) + \gamma \cdot \theta_i$ irrespective of their outcome in absence of treatment ($y_i(0)$). An important nuance is that although the allocation decision only depends on $\tau(\theta_i) + \gamma \cdot \theta_i$, the two terms impact different individuals; the size of $\tau(\theta_i)$ affects the degree to which the treatment impacts individual i directly, while the size of $\gamma \cdot \theta_i$ impacts how effective the treatment is on the other treated individuals. Among other things, this means that designing revelation techniques to determine the optimal assignment of individuals to treatment is complex. For example, using an auction to allocate slots will no longer necessarily provide the most efficient allocation. This is beyond the scope of this paper, however, and we will assume the social planner has complete knowledge of $\tau(\theta_i)$ and $\gamma \cdot \theta_i$.

In Appendix A, we explore the optimal allocation decision further by showing how the treatment slots should be allotted under a few simple restrictions on the individual-level treatment effect heterogeneity, i.e. the $\tau(\theta_i)$ function. For example, we discuss there the optimal allocation if $\tau(\theta_i)$ is linear in θ_i or convex in θ_i. Here, we simply use the total treatment effect expression to emphasize the following remark:

Remark 1. *A social planner who cares only about maximizing the average outcome will allocate the limited number of spots to the T individuals with the largest values of $\tau(\theta_i)+\gamma \cdot \theta_i$. The impact of the allocation on the TTE thus depends both on variation in how individuals themselves respond to the treatment and on the linear-in-means peer effects.*

For example, imagine that the social planner is deciding whether to assign individual j or individual k to the treatment. From Equation (10), it is clear that the difference in the ATEs between assigning j to the treatment rather than individual k is: $\Big(\tau(\theta_j) - \tau(\theta_k)\Big) + \Big(\gamma \cdot (\theta_j - \theta_k)\Big)$. Thus, the difference in how the individuals impact the overall level of peer effects, i.e. $\gamma \cdot (\theta_j - \theta_k)$, affects the TTE of the allocation by the same amount as the difference in how they respond to the treatment themselves, i.e. $\tau(\theta_j) - \tau(\theta_k)$. Note that this means that although there is often considerable attention given to the possibility that individual-level heterogeneity can generate differences in program effectiveness, even linear-in-means peer effects can generate substantial differences as well.

II.D Maximizing the Minimum Outcome

Finally, we consider a case where the social planner is interested in more than the average outcome and has equity concerns as well. A common approach is to consider the extreme example where the social planner is a Rawlsian, in that she only cares about the worst-off individual. Under Rawlsian preferences however, when there are no peer effects the social planner may be indifferent between a range of treatment assignments. For example, if $y_1(0) + \tau_1 < y_2(0)$, where the individual's are ordered by the value of $y_i(0)$ and τ_1 does not depend on the other individuals assigned to treatment, a Rawlsian would be indifferent who she assigns the treatment to beyond individual 1. We will therefore instead consider a social planner who has lexicographic preferences, i.e., she cares first about the individual with the lowest y_i, but given two assignments with the same lowest value of y_i then cares about the individual with the second lowest y_i, etc.

With these preferences, it is clear that the substitutability of $\tau(\theta_i)$ and $\gamma \cdot \theta_i$ breaks down given that equity concerns are involved. In addition, the treatment assignment decision depends not only on how individuals respond to the treatment, but on how well they would do without the treatment, i.e. on $y_i(0)$. To illustrate the importance of peer effects under extreme equity concerns, we can make the following remark:

Remark 2. *If there are no peer effects, a social planner with lexicographic preferences assigns the treatment to the T individuals with the lowest values of $y_i(0)$, assuming that the value of the treatment is positive. With peer effects, however, this is no longer necessarily the case and the social planner may assign some of the treatment slots to high $y_i(0)$ individuals if they also have high θ_i.*

As a concrete example, imagine that the social planner is a school principal who is deciding which students to assign to an after-school learning program and, for this decision, only cares about how well her worst student does. If there are no peer effects, she would clearly assign the lowest achieving students to the tutoring program. With peer effects, however, she might instead allocate a few of these scarce seats to the highest achieving students, as a way to make the program more impactful for the lowest achieving students enrolled. This idea that the treatment goes to two types of individuals: those with the lowest values of $y_i(0)$ and those with the highest θ_i is formalized in the following theorem, which is proved in Appendix A.

Theorem 2. *Suppose that a social planner has lexicographic preferences, in that she chooses*

$\boldsymbol{T}$ to maximize the minimum outcome, breaking ties by then maximizing the second lowest outcome, etc. The optimal allocation of treatment slots then involves assigning the treatment to the K individuals with the lowest values of $y_i(0)$ and the $T-K$ individuals with the highest values of θ_i.

III Context and Data

III.A Workforce Innovation Fund Randomized Controlled Trial

We test these models on a set of adult learning job training programs for disadvantaged workers. Specifically, we investigate peer effects and their implication for optimal allotment of training using the Career Pathways training program run by the New Orleans Office of Workforce Development (OWD), and funded by a US Department of Labor (DOL) Workforce Innovation Fund (WIF) grant starting in 2014. This setting provides an instructive application of the theory for two reasons: first, the intervention was implemented as a randomized controlled trial (RCT), which allows for clean identification both of the treatment effect and the peer effect; second, the setting is one in which peer effects were reasonably likely to happen, with small group sizes (in our sample, average cohort size was below 12 people). In addition, there is a paucity of research on peer effects in the adult learning settings, and so beyond being an application of the theory the peer effect estimates are themselves interesting and informative.

The training targeted lower-income workers in New Orleans interested in job training in one of three pathways: advanced manufacturing, health care, and information technology. For each pathway, there was around 20 hours of training time per week for two to four months. The program was constituted of four distinct stages: recruitment, screening, randomization, and training. We describe the four stages of the program here.

Stage 1: Recruitment OWD recruited potential trainees using several methods: (1) local One-Stop Centers provided program information to individuals using government assistance programs (i.e., SNAP, UI, etc.) who are required to partake in career readiness and reporting activities under the federal funding mandate; (2) fixed tablet stations were placed in targeted local hotspots throughout the city (i.e., community centers, public assistance offices, etc.) to provide program information and collect information on interested applicants; (3) a communications firm hired by OWD provided general program outreach as well as target particular populations via paid advertising and online marketing campaigns, via Facebook, Craigslist, Twitter, and radio ads; and (4) OWD used their existing community partnerships to distribute information about the training program by co-hosting a series of workshops, open to both referrals and members of the general public.

Stage 2: Screening One objective of the Career Pathways program was to access a population of disadvantaged jobseekers (unemployed, underemployed, discouraged, and other interested workers) who had potential to succeed in both the training and subsequent work situations. OWD sought to screen the applicant pool to select individuals who were well-informed about the available career pathways, the rigor of the training program and the employment possibilities post-completion.

In order to do so, OWD screened interested candidates before placing them into the randomization pool. The screening process typically included four components: (1) attendance at a mandatory orientation, (2) drug testing, (3) completion of a relevant assignment or test to determine basic literacy and numeracy (such as the Test of Adult Basic Education, or TABE) and (4) completion of a structured and scored 45-minute interview with OWD about their capacity. Those deemed likely to succeed were placed into the randomization pool, while those deemed unlikely to succeed were directed to other government programs and benefits. While selection of trainees is always a consideration of any training program, this rigorous level of determination and selection is uncommon. In Fortson et al. (2017) review of 28 WIF programs, they found that sites used assessment tools, including the TABE, but did so for information on how to support the trainee and not as a strict selection mechanism.

Stage 3: Randomization Individuals who pass the screening and consented to participate in the study were then placed into the randomization pool for each cohort. Within that group, half of the persons were randomly selected to be invited to the training program, and half were not. Randomization was done using random numbers within strata defined by gender, baseline income, current employment status, and age. The intention of the control group, i.e., the individuals who were not assigned to the treatment, was to serve as a business-as-usual benchmark. While this meant that they were not allowed to enroll in the cohort, it also meant that they were able to enter subsequent randomization cohorts as well as receive the other services provided by the one-stop job centers, which were available to everyone.

Stage 4: Training The training program represented a coordinated effort between the OWD and the training providers, which included a local community college, established professional training providers, and firms. Training was typically offered through twenty hours across five days per week. The first round of training was around two months long. At the completion of the first round of training, candidates in most cohorts were invited to enter an additional subsequent "stackable" credit program, which would comprise an additional two months of training within their career pathway. Training would involve the opportunity to obtain various industry-based credentials along the way, conditional on attendance and success on a set of tasks or tests.

III.B Data

We define a cohort as a group of individuals entering the same randomization pool to enter a specific training program together at the same time. There were 25 total cohorts in the Career Pathways program. The first cohort was not randomized, and significant programmatic changes to the stages were made after the third cohort. Thus, following Baird et al. (2019), we do not include in our analysis these three cohorts. There are an additional two cohorts at the end of the program for which no data was available, and so we do not include them in our analysis either. Our final sample thus uses 20 cohorts for this paper.

Our data for this paper comes from two sources. First, in order to enter a randomization pool, individuals had to fill out a brief survey (the baseline survey) with questions about their demographics and current employment and earnings. They also provided their social security numbers and consented to researchers obtaining data from the state of Louisiana. This enabled the second data source: the history of employment, earnings, and industry records were obtained from the Louisiana Workforce Commission (LWC) for each individual for each job they held that was reported to the state. This data is at the quarterly level, between and including the first quarter of 2014 and the first quarter of 2019, for 21 total consecutive quarters of data for each individual. Most of these quarters were before the start of training.

Table 1 presents information regarding the training program enrollments. After restricting to the 20 cohorts as described above, there were nearly 500 individuals in the study, with over 2,000 post-training quarterly observations. Advanced manufacturing had the largest number of participants, driven by having larger cohorts on average. Health care was the smallest pathway, both in terms of number of cohorts and participants.

Table 2 presents the baseline characteristics of the sample we use for this analysis, as well as the outcomes from the LWC data for a period of two years prior to training. The sample is relatively low-income, with average earnings of a little more than $3,500 per quarter (or around $14,000 per year). 53% of the sample had a job at time of randomization. Approximately 90% of the sample are Black individuals and half of the sample are over age 35. We find that there is no statistical difference between the treatment and control group on any of these dimensions and the differences are quite small relative to the pooled standard deviations. Thus, there is good balance between the two groups.[11]

III.C Labor Market History

We now construct a measure that captures individuals' labor market history, which serves as our measure of θ_i as defined in the above conceptual framework for our empirical analysis. We start by standardizing four measures, so all are mean zero and have a standard deviation of one. These measures are: a) the number of quarters employed in the two years before

[11] We discuss balance on peer characteristics in Section IV and Appendix D.1.

randomization; b) the total earnings in the two years before randomization; c) total earnings in the year before the screening occurs; d) an indicator for whether individuals are working at the time of screening. The choice of these measures reflects our assumption that the individuals who will be most helpful classmates are those who have had positive labor market experiences, and therefore can share experiences of job search and job success. It is also plausibly related to underlying ability, which means individuals with positive labor market experiences may also be better able to help other classmates learn difficult material.

We then average these standardized measures to calculate a measure of each individual's labor market history. To construct our measure of an individual's peers' average labor market history, we calculate a leave-one-out average for each individual in which we calculate the average labor market history for all individuals who apply, are accepted, and are assigned to the same arm (i.e., treatment or control) in the same cohort as individual i. Importantly, this highlights the fact that the average value of θ_i of an individuals' peers is largely random, as it depends on whether the individuals who were well-connected to the labor market are the ones that were randomly assigned to the treatment or to the control group within a cohort.

Figure 1 shows the distribution of both individual labor market history and the peer average labor market history. As can be seen, individual values of θ_i range from around -2.75 to about 1.75, and has a standard deviation of around 1.38. As would be expected, variation in the peer average labor market history is smaller than variation in individual labor market history. However, Figure 1 shows that there is still a reasonable variation in the peer average labor market history, with values ranging from approximately -1.5 to 1, and has a standard deviation of 0.32. Thus, as measured by the standard deviation, there is approximate 25% as much variation in peer average labor market history as there is in individual labor market history.

IV Empirical Strategy

IV.A Empirical Specification

We now discuss the empirical specification that we use to estimate the average treatment effect, the degree of individual-level treatment heterogeneity, and the size of the peer effect. After detailing the exact specification, we discuss the identification intuition in the next subsection.

Given the relatively small sample size and large number of potential explanatory covariates, we first use a random forest regression to residualize the outcome and reduce the variance of the error term. More specifically, we use as our main outcome:

$$r_{i,k,t} = y_{i,k,t} - \hat{g}(Z_i) \tag{11}$$

where $y_{i,k,t}$ is the outcome of interest, Z_i are covariates described below, and

$$\hat{g}(Z_i) = \frac{1}{2}\hat{\mathbb{E}}[y_{i,k,t}|Z_i, T_i = 1] + \frac{1}{2}\hat{\mathbb{E}}[y_{i,k,t}|Z_i, T_i = 0] \tag{12}$$

As the covariates Z_i, we use: information on individual i's gender, race, and age at randomization; employment status and total wages received in each of the twelve quarters prior to randomization; information used to determine eligibility at screening (discussed in Section III); and time period post-training. Since all of the covariates are not impacted by the treatment, residualizing the outcome by any function $g(Z_i)$ would not change the estimated coefficients. See Appendix B for a short proof of this statement and more discussion about the motivation both behind this residualization approach as well as the form of $g(Z_i)$ we use. Finally, we estimate $\hat{\mathbb{E}}[y_{i,k,t}|Z_i, T_i = 1]$ and $\hat{\mathbb{E}}[y_{i,k,t}|Z_i, T_i = 0]$ using all the cohorts except for k, which ensures that the estimation of $\hat{g}(Z_i)$ is independent of individual i's treatment status.

To estimate the size of the peer effect and degree of individual treatment heterogeneity, we then run the following regression:

$$r_{i,k,t} = \alpha\theta_i + \Big[\tau + \beta\theta_i + \gamma\overline{\theta}_{i,k}\Big]T_i + \delta_k + \nu_i + \eta_t + \epsilon_{i,k,t} \tag{13}$$

where:

- $r_{i,k,t}$ is the residualized outcome for individual i who applied to cohort k in period t;
- θ_i is individual i's labor market history, described in the section above;
- $\overline{\theta}_{i,k}$ is the average labor market history for individuals in the same cohort k and treatment status as individual i;[12]
- T_i is an indicator for whether individual i was randomly assigned treatment;
- δ_k is a fixed effect indicating which cohort individual i applied for;
- ν_i is a vector of strata fixed effects, indicating which strata individual i is in; the strata are defined by gender, income, employment status, and age all measured at the time of randomization;
- η_t are period fixed effects;
- $\epsilon_{i,k,t}$ is an individual error term.

In the main specification, the empirical model restricts the theoretical model developed above by assuming that $\tau(\theta_i)$ is linear in θ_i, i.e. $\tau(\theta_i) = \tau + \beta \cdot \theta_i$. Thus, τ is the average treatment effect, $\hat{\beta}$ is the estimate of the amount of individual-level treatment heterogeneity and $\hat{\gamma}$ is the estimate of the size of the peer effects.[13]

In the specification above, we control for non-random selection into cohorts by including cohort fixed effects in the specification to account for potentially endogenous selection into

[12] Formally, we define $\overline{\theta}_{i,k}$ as $\frac{T_i}{T^k-1}\sum_{\forall j \neq i}\theta_j T_j A_j^k + \frac{1-T_i}{C^k-1}\sum_{\forall j\neq i}\theta_j(1-T_j)A_j^k$, where T^k is the number of individuals assigned to treatment in cohort k, C^k is the number of individuals assigned to the control in cohort k, and A_i^k is an indicator for whether individual i applied for cohort k.

[13] The fact that τ identifies the average treatment effect follows from the fact that both θ_i and $\overline{\theta}_{i,k}$ are defined to be mean-zero.

cohorts. We could augment this by including the average labor market history of the control individuals' "peers'", i.e., the individuals that applied to the same cohort and were not assigned to the treatment. In doing so, the specification would be nearly identical to the one suggested by Caeyers and Fafchamps (2020), for example, with the only difference that we allow for the impact of ones' peers and own labor market history to matter more or less in the treatment groups than the control groups. While we leave this specification to Appendix A4, we find nearly identical results in this specification as in our main specification and, reassuringly, we find no evidence than controls individuals' "peers" labor market history are related to their own outcomes.

When estimating we allow for correlation of the $\epsilon_{i,k,t}$ error terms both within-individuals and within-treatment groups by clustering at the individual-level for individuals not randomly assigned to the treatment and at the cohort-by-treatment-level for those randomly assigned to the treatment.[14]

IV.B Identification Intuition

We now discuss more explicitly how the random assignment of individuals to treatment within-cohorts allows us to credibly estimate the magnitude of the peer effects, despite the many difficulties that peer effect estimation entails (e.g., Manski (1993); Angrist (2014); Feld and Zölitz (2017)). To do so, we start by highlighting a relevant empirical result. As we show in Figure 2a), when looking across cohorts the estimated cohort-specific treatment effect is positively correlated with the cohort average labor market history. Note that the within-cohort randomization of individuals to treatment implies that the cohort-specific treatment effect estimates are unbiased estimates of the true cohort-specific treatment effect.

This finding is consistent with the hypothesis that there are peer effects in the program, but is not necessarily indicative of their presence. Notably, the positive correlation between cohort average labor market histories and cohort average treatment effect may be due to individual treatment effect heterogeneity rather than peer effects. Specifically, this positive

[14]Results are similar when we cluster at the cohort-by-treatment level for all individuals.

correlation could also be the result of individuals with higher labor market histories having larger individual treatment effects. We can test whether this is the case by looking within-cohorts, however, and as shown in Panel 2b), within-cohorts the estimated treatment effect is negatively correlated with an individual's labor market history.[15]

These twin empirical findings - that the relationship between average treatment effect and average labor market history across-cohorts is positive and that the relationship between individual treatment effects and individual labor market history within-cohorts is negative - necessarily implies that the effect of the treatment on an individual is positively correlated with their peers' average labor market history. This appears to suggest the presence of peer effects, i.e., that attending a session where your peers have high average labor market history *causes* the session to have larger effects. There is, however, one alternative explanation for these empirical results, that within cohorts individuals with high labor market histories have lower treatment effects than those with worse labor market histories, but that individuals with high labor market histories generally select into cohorts that have larger treatment effects. While we cannot rule this out directly, there are three reasons to think that this is not the case.

First, informally, estimating the causal effect of an individuals' peers on her outcome requires variation in the ability of the individuals' peers that is uncorrelated with her own ability.[16] Normally, this is quite challenging as similar individuals endogenously group together, making it difficult to tease apart the impact of an individual's own ability on her outcome as opposed to the impact of the individual's peers' average ability. In our case, however, the random assignment of individuals to treatment is what generates a significant amount of the variation in an individual's peers' labor market history.

[15]We estimate this relationship by regressing each individuals' outcomes on cohort dummies, cohort dummies interacted with treatment status, their own labor market history, and their own labor market history interacted with treatment status. The line in Figure Panel 2b) reflects the coefficient on individuals' own labor market history interacted with treatment status and the dots reflect averages outcomes for individuals in the ten deciles of labor market history after netting out cohort averages. Again, it is the the random assignment of individuals to treatments within-cohorts that allow us to interpret these estimates as causal.

[16]This definition is informal in part because it ignores important differences between "observed" ability and "unobserved" ability at both the individual and group level.

Of course, an individuals' peers depends both on who else applied for the specific cohort and how the random assignment of individuals to treatment within-cohorts occurs; the random assignment solves the latter potential source of endogenous peer group formation, but not the former. However, we find no evidence that individuals selected into cohorts based on their labor market history. Specifically, we note that if individuals select into cohorts based on their labor market history, we would find larger across cohort variation in average labor market history than if individuals are randomly assigned to cohorts. We can then test for evidence of endogenous choice into cohorts by testing how the observed across-cohort variance of average labor market history compares to the distribution of across-cohort variance of average labor market history that would occur under random assignment. We do so via a simulation, in which we randomly assign individuals to cohorts and calculate the across-cohort variance in average labor market history under these hypothetical assignments. After doing so, we find that approximately one-quarter of the hypothetical assignments have larger across-cohort variation in average labor market history than the observed assignment. Formally, this means that we cannot reject the null hypothesis of random assignment of individuals to cohorts at any reasonable significance level; our specific p-value of a two-sided hypothesis test of random assignment to cohorts is 0.472. Furthermore, we do not find any evidence of peer effects on the individuals' *expected* labor market outcomes, where the expectation is estimated using their pre-randomization outcomes. See Appendix D.1 for more details on the simulations and regressions.

Finally, although our identification approach means that we cannot allow for cohort-specific treatment effects in the specification, we can allow for pathway or area-specific treatment effects.[17] That is, we can allow for health care cohorts to have different average effects than IT cohorts, for example. If the results are due to non-random selection into these different types of training programs in a way that correlates with the treatment effect, allowing for different average effects would cause the peer effect estimates to disappear. In contrast, allowing for pathway- or area-specific treatment effects makes almost no impact

[17] See Table 1 for the list of pathways and areas.

on the peer effect estimates, as we show in Table A6.

To summarize, our identification assumption is supported by several points: first, much of the variation in peer characteristics are generated by the within-cohort random assignment of individuals to treatment groups; second, the remaining variation in peer characteristics is consistent with the random assignment of individuals to groups; and third that the peer effect estimates do not meaningfully change when allowing for additional treatment effect heterogeneity. Together, these three empirical findings collectively give credence to the empirical specification outlined in the previous section.

V Peer Effect Estimates

We now present evidence that the effectiveness of the job training programs on an individual depended in part on the labor market history of the individuals they took the training with. The estimated peer effects are shown in Table 3, which also shows the average impact of the training and the how the impact of the training depends on the individuals' own labor market history. The four columns pool all quarters post-training and include fixed effects for the training cohort, for the strata used during randomization, and the time period. The four columns differ only in the outcome they show: the first two shows the impact on total wages (in logs and levels respectively), the third on whether an individual was employed or not, and the fourth on conditional wages, i.e. it restricts the sample to those employed after the training period and shows the impact on their total wages.

As seen in the third row of the first three columns of Table 3, the average treatment effect of the program on both wages and employment was positive. Since the fourth column conditions on an endogenous variable, the coefficient should be interpreted with caution. The fact that average treatment effect is positive here too, however, does provide some suggestive evidence that not all of the wage effect shown in columns (1) and (2) are due to the employment effect shown in column (3).[18] While the third row suggests that the

[18]For much more information about the average effect of the program, see the RAND Report written by two of the authors of this paper (and others): Baird et al. (2019). In that report, the average treatment

average effect is positive, the second row of Table 3 shows that the program has a larger effect on individuals who themselves have a poor labor market history; that is, those who were less connected to the labor market before the randomization saw a larger benefit of attending the program than individuals with a better labor market history. Finally, the top row shows that the effectiveness of the program on a particular individual is larger when the other individuals attending the job training program have a good labor market history, i.e. there are peer effects in the program. This result is statistically significant at the 1% level if the outcome of interest is total wages earned (when measured in logs) or employment probability and statistically significant at the 10% level if the outcome of interest is total wages earned (when measured in levels).

To interpret the size of the coefficients, consider an individual who has an average labor market history herself and who attends a training session where the peer average labor market history is one standard deviation below the mean. Given that the standard deviation of peer average labor market history is 0.32, the coefficients from Table 3 suggest that she benefits from attending the program. By attending the program she sees an increase in her total wages of approximately \$490 per quarter, or \$935 - \$1390*0.32. Contrast this with another individual who also had an average labor market history herself, but who attended a training session where the peer average labor market history is one standard deviation *above* the mean. The benefit from attending the training session for this individual is around \$1,380 more per quarter, or \$935 + \$1390*0.32. Thus, while both individuals benefit from attending the training program, the benefit for the individual who attended the training session with peers who had above average labor market history is about \$900 more than the individual who attended the training session with peers who had below average labor market history.

It is worth dwelling on this magnitude and putting it in context. Perhaps the most natural comparison is to compare the differential effect the program had on these two

effect (ATE) is estimated by comparing the treatment and control means, but the estimates are quite similar to the ones reported in Table 3. For example, the ATE on total wage in the report is \$805.

hypothetical individuals and to the average effect of the program. We find that the average effect of the program is \$935, so the impact of a one-standard deviation change in the group average labor market history on the program's effect is around half of the average treatment effect.[19] Put differently, imagine a two-step randomization procedure in which the first coin flip determined whether an individual was treated or not and the second, if she was assigned to treatment, determined whether she was assigned to a group with average labor market history one standard deviation above the average or one standard deviation below; our results suggest that both coin flips have approximately the same impact on her eventual earnings and employment probability.

Finally, we note that while hypothetical, these calculations do not involve extrapolating outside of the sample variation. In other words, the calculations do not consider a hypothetical group where each individual has a labor market history one-standard deviation above the mean, with the standard deviation measured using the distribution of individuals' labor market history. Instead, we consider a group where the group average labor market history is one-standard deviation above the mean, with the standard deviation measured using the observed group average labor market history.

The above analysis pools all quarters post-training session. We next explore how the importance of peers evolves over time. To do so, we separately estimate Equation (13), first restricting the sample to the quarter in which the training takes place and then restricting the sample to each of the six quarters after the training ended. These results are shown in Figure 3, with Panel 3(a) showing the results for log wages, Panel 3(b) showing the results for total wage, and Panel 3(c) showing the result for employment status. The results are similar for all three outcomes: for the period in which the training takes place, the labor market history of the peer individuals taking the training with an individual do not affect her total earnings or employment status. This result serves as a placebo test of sorts, adding further credence to the identification approach. In the quarters after the training program ends,

[19] Specifically, $\frac{1{,}390 \times 0.32}{935} = 0.48$ so the effect on the treatment effect of a one-standard deviation change in the group average labor market history is 47% as large as the average treatment effect.

however, the effect of the program becomes increasingly dependent on the individuals' peers within the program.[20] This suggests that the results in Table 3 understate the importance of an individuals' peers on the long-run effectiveness of the training program.

In short, we find evidence of peer effects in this program and that these peer effects are large relatively to the average treatment effect. We next consider how these results affect the optimal assignment of individuals to the program; however, it is worth emphasizing again that these results are important in-and-of themselves as this result fills a gap in the literature by illustrating the existence of large peer effects in an adult learning context.

VI Optimal Treatment Allocation

In Section V, we considered as fixed the random assignment of individuals to treatment. We now use the empirical results to once again consider the decision of the social planner, as discussed in Section II.

VI.A Average Treatment Effect Calculations

To start, we will focus on a social planner who cares only about maximizing the total wages earned by the study participants. As we showed above, we can write the total treatment effect (TTE) of a particular allocation as:

$$\mathrm{TTE}(\mathbf{T}) = \sum_{\forall i} \Big(\tau(\theta_i) + \gamma \cdot \theta_i\Big) \cdot T_i \tag{14}$$

which also can be thought of as the aggregate benefit of the program on the specific outcome examined.

We can now go further than in Section II and explicitly calculate the aggregate benefit of the training program for any given $\mathbf{T}$, i.e. for any assignment of individuals to treatment.

[20]For the cohorts that started late, we do not have data for all cohorts for a full six periods post-training. Thus, there is some sample attrition in the later periods, which explains the larger standard errors in the later periods. Note, however, that there is no selective attrition within cohorts and instead all the attrition is due to our data ending in the fall of 2019.

We do so using the fact that Section III discusses how we estimate θ_i for each individual and Table 3 reports that we find: $\tau(\theta_i) \approx 935 - 431\theta_i$ and $\gamma \approx 1390$. We can then use these values to calculate four values: a) the aggregate benefit of the program under the optimal allocation of individuals to treatment; b) the aggregate benefit of the program under the allocation that a social planner who ignores the presence of peer effects would do; c) the expected aggregate benefit of the program under random assignment of individuals to treatment; and d) the actual aggregate benefit of the program under the realized assignment of individuals to treatment.

For our calculations, we restricted the social planner by: a) only allowing the social planner to assign individuals to treatment within the cohort they applied and b) keeping the number of individuals per program fixed. Given Equation (14), calculating the most efficient and least efficient allocation is straightforward. The first step is to calculate $\tau(\theta_i) + \gamma \cdot \theta_i$ for each individual. For each cohort k which has T_k treatment slots, the most efficient allocation then assigns the treatment to those individuals with the T_k largest values of $\tau(\theta_i) + \gamma \cdot \theta_i$. Similarly, a naive social planner who ignores the presence of peer effects would instead assign the treatment to those individuals with the T_k largest values of $\tau(\theta_i)$.

The results are shown in the first column of Table 4. It shows that the overall benefit of the program can vary widely, depending on who is assigned to the treatment. Under the optimal allocation the program increases the wages of participating individuals by over \$1,823 per quarter. While attempting to assign the slots optimally, a social planner who ignores the presence of peer effects, on the other hand, would lead to a program that is ineffective and only increases participants' per-quarter earnings by \$124 per quarter. This difference demonstrates the discussion under the assumption of linear treatment effect heterogeneity described in Appendix A.B. That is, disregarding peer effects (Naive Optimal) leads the social planner to select the T individuals with the lowest θ_i in each cohort, given the negative estimated relationship between own labor market history (θ_i) and outcomes (Table 3). However, the true optimal allocation is to assign treatment to the highest θ_i individuals, given the stronger and positive impact of peer effects compared to own labor market history. Thus, the "Naive Optimal" assignment is actually the least efficient allocation of individuals to treatment. The results shown here are not dependent on the linearity assumption, however; we obtain very similar estimates when we allow $\tau(\theta_i)$ to be a higher order polynomial. We also note that since the peer effect is much larger than the own labor history impact in this context, the social planner could incorrectly assume smaller peer effects and still make the optimal assignment.

Naturally, this analysis relies strongly on the assumption of linear-in-means peer effects and involves extrapolating that linear relationship outside the range that it was estimated on. Furthermore, as Carrell et al. (2013) show, exploiting peer effects is fraught with challenges. We therefore do not view these estimates as proof positive that the job training program we study would have had a negligible impact on participants had all the individuals with the lowest labor market history in a cohort been the ones assigned to the program nor that the effect of the program would be twice as large had the individuals been assigned optimally, rather than randomly.[21] Instead, we view this result as indicative of our main conclusion, that the presence of peer effects - even linear-in-means peer effects - can have a large impact of the efficacy of a program.

VI.B Equity Calculations

Like in Section II, we next consider a Rawlsian social planner who wants to choose the treatment allocation to maximize the worst-off individual. An additional challenge here is that the optimal allocation decision depends in part on $y_i(0)$, which is unobservable. We have, however, generated predictions of $y_i(0)$ when we conducted the residualization. We therefore use these predictions to approximate $y_i(0)$.[22]

Furthermore, the complex nature of the Rawlsian social planner's utility function makes

[21] These other important sources of uncertainty are why we do not provide 95% confidence intervals on these measures, as they would only reflect statistical uncertainty in the peer effects estimates and not the other important sources of uncertainty.

[22] Another approach is to use as our approximation of $y_i(0)$ the baseline wages that individuals received, which we observe. We get similar results when we use their baseline wages as our approximation of $y_i(0)$, rather than the machine learning predictions.

maximizing the utility function more computationally challenging than maximizing Equation (14). However, we can leverage the unique form of the optimal allocation as discussed Theorem 2 to make this maximization possible. Specifically, the results means we can consider only $T_k - 1$ possible optimal treatment allocations for each cohort rather than consider each of the $\frac{N_k!}{(N_k - T_k)! T_k!}$ total number of possible treatment allocations; here T_k is the number of treated individuals in each cohort and N_k is the total number of individuals in the cohort.

Like before, we require the planner to exactly fill the number of slots that were available and only allow them to allocate an individual to a training program if they applied to the program, i.e. if the individual was in the cohort. We then consider the utility of the social planner under the optimal assignment, the assignment they she would choose if she ignored peer effects, the expected utility under random assignment, and the actual utility under the realized assignment.

When calculating the social planner's utility, we assume a form of narrow-framing by the social planner (Barberis et al. (2006)). More specifically, we assume that while she cares only about the individual with the minimum outcome *within each cohort*, her overall utility is the average of this utility *across the cohorts*. We make this assumption to ensure that the treatment assignment of each cohort has impact to the overall utility. An alternative interpretation is that a different Rawlsian social planner makes the treatment allocation decision for each cohort.

These results are shown in the second column of Table 4. Again, a social planner who naively ignores the presence of peer effects does much worse than the social planner who accounts for peer effects. In the job training program estimation here, the social planner can incorrectly assume very small peer effects and still arrive at a minimum realized earnings that is very close to the optimal level, albeit with a different allocation decision. This is because of Theorem 2, wherein under peer effects, the social planner may begin switching out medium $y_i(0)$ individuals, who will still fare better in absence of treatment than the lowest outcome, and switching in the highest θ_i individuals, whose peer effects will pull up the outcome for the lowest trainee. Given the smaller size of many of these cohorts, switching only a few such individuals can have dramatic effects on the average labor history of the cohort, and thus yield significant peer effects.

VII Conclusion

In job training programs, as in many other contexts, there are a limited number of seats into which a policy maker must assign individuals. This leads to the question of who should be admitted to training programs if peer effects are large. Is it optimal to only train those most in need or should some less needy individuals be admitted in order to capture the positive impact they have on other participants? Does the answer depend on the policy makers' utility function?

Our theoretical analysis first shows that the presence of peer effects generally leads an optimizing policy maker to assign seats to relatively more able individuals, regardless of her utility function. For example, we show that even a policy maker who cares only about equity, while ensuring that some seats go to the most needy, will also potentially reserve some seats for the relatively more able individuals.

Our empirical analysis demonstrates that these observations are not merely of academic interest. The presence of a linear-in-means peer effect has a large impact on the efficiency of job training programs and changes the optimal allocation of the limited training seats. We use the random assignment of individuals to treatment groups to skirt many of the issues prevalent in estimating peer effects and provide the first evidence of large peer effects in the context of job training programs. Our estimates indicate that the heterogeneous treatment effect due to peer effects is equivalent to roughly half of the average treatment effect; formally, the relative effectiveness of two programs which differ in peer average labor market history by a one-standard deviation is roughly half of the average treatment effect. We then consider two types of social welfare functions: one that aims to maximize the aggregate benefit of the program and another that emphasizes equity. In either case, we show that the social planner's utility could be greatly increased if she allocated slots optimally and that accounting for peer effects is quite important when determining the optimal assignment.

The results from this study provide guidance to practitioners in the allocation of scarce resources in many education and training settings. For example, who should be granted access to extra-curricular programs that aim to supplement classroom learning? If seats in these programs are scarce, whether the goal is to raise the outcomes of the lowest performing students or to raise the outcome averaged over all students, our findings suggest that the presence of peer effects may well make it optimal to implement an admission procedure that ensures some participants are from each end of the academic spectrum. While assigning scare resources to more affluent children should be done with care, our results suggest that in some limited contexts that may well be the best way to improve outcomes for the most disadvantaged children.

Finally, training and learning occurs in many settings, not all of which are formal. One can imagine a manager who is tasked with deciding which of her employees are trusted to work on solo projects and which should be grouped into teams. Our results suggest that there may be some value in assigning the high ability workers to the team, as a way to capture their peer effects and improve the ability of lower-performing members of the team. Of course, whether this is optimal in the workplace is likely to depend on whether peer effects function through explicit teaching or through example. If the peer effects from the more productive members of a team require that they reduce their own output, then our results might not generalize to the workplace. If, on the other hand, peer effects operate through example, then peer effects can be realized at little cost.

Better understanding why we find such large peer effects in the job training program is therefore of crucial importance if one is to consider how what our results imply about similar decisions in alternative contexts. A more full understanding of the mechanisms may also inform ways in which job training programs could be improved; for example, if the peer effects are due to example, having participants in a job training program shadow high ability individuals who previously participated in the training may be a way to improve the effectiveness of the program. While not providing a clear answer on the mechanisms, by showing both that there are large peer effects in job training programs and that these can greatly impact the effectiveness of such programs, we highlight the importance of more research into how individuals interact in these programs. More generally, our paper also highlights that peer effects play a crucial and under-appreciated role in determining the optimal assignment of individuals to programs and in explaining cross-program treatment effect heterogeneity.

References

Akbarpour, Mohammad, Suraj Malladi, and Amin Saberi, "Just a Few Seeds More: Value of Network Information for Diffusion," *Working Paper*, 2020.

Andersson, Fredrik, Harry J Holzer, Julia I Lane, David Rosenblum, and Jeffrey Smith, "Does federally-funded job training work? Nonexperimental estimates of WIA training impacts using longitudinal data on workers and firms," Technical Report, National Bureau of Economic Research 2013.

Angrist, Joshua D., "The Perils of Peer Effects," *Labour Economics*, 2014.

Baird, Matthew D., John Engberg, Gabriella C. Gonzalez, Thomas Goughnour, Italo A. Gutierrez, and Rita T. Karam, *Effectiveness of Screened, Demand-Driven Job Training Programs for Disadvantaged Workers: An Evaluation of the New Orleans Career Pathway Training*, Santa Monica, CA: RAND Corporation, 2019.

Banerjee, Abhijit, Arun G. Chandrasekhar, Esther Duflo, and Matthew O. Jackson, "The Diffusion of Microfinance," *Science*, 2013, *341* (6144).

_, Arun G Chandrasekhar, Esther Duflo, and Matthew O Jackson, "Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials," *The Review of Economic Studies*, 02 2019, *86* (6), 2453–2490.

Barberis, Nicholas, Ming Huang, and Richard H. Thaler, "Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing," *American Economic Review*, September 2006, *96* (4), 1069–1090.

Belhaj, Mohamed, Frederic Deroian, and Shahir Safi, "Targeting in Networks under costly agreements," 2020.

Bhattacharya, Debopam, "Inferring Optimal Peer Assignment From Experimental Data," *Journal of the American Statistical Association*, 2009, *104* (486), 486–500.

Bloch, Francis, Matthew O. Jackson, and Pietro Tebaldi, "Centrality Measures in Networks," 2020.

Booij, Adam S, Edwin Leuven, and Hessel Oosterbeek, "Ability Peer Effects in University: Evidence from a Randomized Experiment," *The Review of Economic Studies*, 09 2016, *84* (2), 547–578.

Breza, Emily, Arun G. Chandrasekhar, Tyler H. McCormick, and Mengjie Pan, "Using Aggregated Relational Data to Feasibly Identify Network Structure without Network Data," *American Economic Review*, 2020, *110* (8), 2454–2484.

Caeyers, Bet and Marcel Fafchamps, "Exclusion Bias in the Estimation of Peer Effects," *NBER Working Paper 22565*, 2020.

Cai, Jing, Alain De Janvry, and Elisabeth Sadoulet, "Social Networks and the Decision to Insure," *American Economic Journal: Applied Economics*, 2015, *7* (2), 82–108.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West, "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation," *Econometrica*, 2013, *81* (3), 855–882.

Comola, Margherita and Silvia Prina, "Treatment Effect Accounting for Network Changes," *The Review of Economics and Statistics*, Forthcoming.

Cronen, Stephanie, Meghan McQuiggan, and Emilty Isenberg, "Adult Training and Education: Results from the National Household Education Surveys Program of 2016," Technical Report NCES 2017103REV, National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. 2017.

Demange, Gabrielle, "Optimal targeting strategies in a network under complementarities," *Games and Economic Behavior*, 2017, *105*, 84–103.

Duflo, Esther, Pascaline Dupas, and Michael Kremer, "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya," *American Economic Review*, August 2011, *101* (5), 1739–1774.

Feld, Jan and Ulf Zölitz, "Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects," *Journal of Labor Economics*, 2017, *35* (2), 387–428.

Field, Erica, Seema Jayachandran, Rohini Pande, and Natalia Rigol, "Friendship at Work: Can Peer Effects Catalyze Female Entrepreneurship?," *American Economic Journal: Economic Policy*, May 2016, *8* (2), 125–53.

Fortson, Kenneth, Dana Rotz, Paul Burkander, Annalisa Mastri, Peter Schochet, Linda Rosenberg, Sheena McConnell, and Ronald D'Amico, "Providing Public Workforce Services to Job Seekers: 30-month Impact Findings on the WIA Adult and Dislocated Worker Programs," Technical Report, Mathematica Policy Research, Washington, D.C. 2017.

Galeotti, Andrea, Benjamin Golub, and Sanjeev Goyal, "Targeting Interventions in Networks," *Econometrica*, 2020, *88* (6), 2445–2471.

Goldsmith-Pinkham, Paul and Guido W. Imbens, "Social Networks and the Identification of Peer Effects," *Journal of Business & Economic Statistics*, 2013, *31* (3), 253–264.

Guryan, Jonathan, Kory Kroft, and Matthew J. Notowidigdo, "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments," *American Economic Journal: Applied Economics*, October 2009, *1* (4), 34–68.

Heckman, James J, "Policies to foster human capital," *Research in economics*, 2000, *54* (1), 3–56.

Heinrich, Carolyn J, Peter R Mueser, Kenneth R Troske, Kyung-Seong Jeon, and Daver C Kahvecioglu, "Do public employment and training programs work?," *IZA Journal of Labor economics*, 2013, *2* (1), 1–23.

Hollenbeck, Kevin et al., "Study of Washington's Unemployment Training Benefits Program," 2012.

Horn, Carl Van, Tammy Edwards, and Todd Greene, "Transforming US workforce development policies for the 21st century," *Federal Reserve Bank of Atlanta and WE Upjohn Institute for Employment Research, Kalamazoo*, 2015.

Hoxby, Caroline and Gretchen Weingarth Salyer, "Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects," 2006.

Jochmans, Koen, "TESTING RANDOM ASSIGNMENT TO PEER GROUPS," 2021.

Johnsson, Ida and Hyungsik Roger Moon, "Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach," *The Review of Economics and Statistics*, Forthcoming.

Kempe, David, Jon Kleinberg, and Éva Tardos, "Maximizing the Spread of Influence through a Social Network," in "Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining" KDD '03 Association for Computing Machinery New York, NY, USA 2003, pp. 137–146.

Lafortune, Jeanne, Marcela Perticara, and Jose Tessada, "The benefits of diversity: Peer effects in an adult training program in Chile," 2018.

Lavy, Victor and Analia Schlosser, "Mechanisms and Impacts of Gender Peer Effects at School," *American Economic Journal: Applied Economics*, 2011, *3* (2), 1–33.

Manski, Charles F., "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, July 1993, *60* (3), 531–542.

Mas, Alexandre and Enrico Moretti, "Peers at Work," *American Economic Review*, 2009, *99* (1), 112–145.

Opper, Isaac M., "Does Helping John Help Sue? Evidence of Spillovers in Education," *American Economic Review*, March 2019, *109* (3), 1080–1115.

Sacerdote, Bruce, “Peer Effects With Random Assigment: Results for Dartmouth Roomates,” *Quarterly Journal of Economics*, 2001, *116*, 681–704.

van den Berg, Gerard J., Sylvie Blasco, Bruno Crepon, Daphne Skandalis, and Arne Uhlendorff, “Peer Effects of Job Search Assistance Group Treatments. Evidence from a Randomized Field Experiment among Disadvantaged Youths,” 2021.

Viviano, Davide, “Policy Targeting under Network Interference,” *Working Paper*, 2020.

VIII Tables and Figures

VIII.A Tables

Table 1: Description of training cohorts

Pathway	Area	Cohorts	Number Treatment	Number Control	Observations
Advanced manufacturing	Electrical	5	87	79	1011
	Pipefitting	1	6	6	24
	Welding	1	11	8	57
Health care	Medical billing and coding	4	42	37	208
	Patient access representative	1	19	20	195
Information technology	Information technology	8	72	68	757
Total		20	237	218	2252

*Note: The number of observations are post-training quarterly observations of outcomes, while the number treatment and number control are the unique number of individuals.

Table 2: Baseline characteristics of sample

	Control	Treatment	P-Value	Difference 95% CI	Effect Size
Male	0.556	0.566	0.949	[-0.267,0.286]	0.020
Over 35 years old	0.553	0.596	0.310	[-0.045,0.131]	0.088
Working	0.527	0.500	0.517	[-0.123,0.066]	0.054
Prior year's income > $5,000	0.610	0.641	0.610	[-0.093,0.145]	0.064
Black	0.887	0.915	0.323	[-0.025,0.084]	0.088
Proportion of prior quarters employed	0.611	0.600	0.802	[-0.096,0.074]	0.023
Average prior quarterly earnings	3788.713	3522.633	0.531	[-1100.0,562.3]	0.064

*Note: The counts of the individuals who are included in the averages are shown in Table 1 above. The "P-Value" column reports the p-value of a null hypothesis that there is no difference in the means of the treatment and control observations. The "Effect Size" column reports the mean difference divided by the control group's standard deviation.

Table 3: Average Treatment Effect, Peer Effect, Individual-Level Treatment Heterogeneity

	(1) Log Wage	(2) Total Wage	(3) Employed	(4) Total Wage Conditional
Impact of Peers' Avg. Labor Market History on Treatment Effect	2.026***	1390.8*	0.219***	1142.9
	(0.756)	(777.3)	(0.0811)	(846.4)
Impact of Own Labor Market History on Treatment Effect	-0.238	-430.9**	-0.0254	-773.0***
	(0.211)	(217.7)	(0.0231)	(223.9)
Average Treatment Effect	0.866***	935.9***	0.0784**	1330.1***
	(0.259)	(265.3)	(0.0323)	(297.6)
Cohort Fixed Effects	X	X	X	X
Time Period Fixed Effects	X	X	X	X
Strata Fixed Effects	X	X	X	X
Number of Clusters	174	174	174	139
Observations	1632	1632	1632	1095

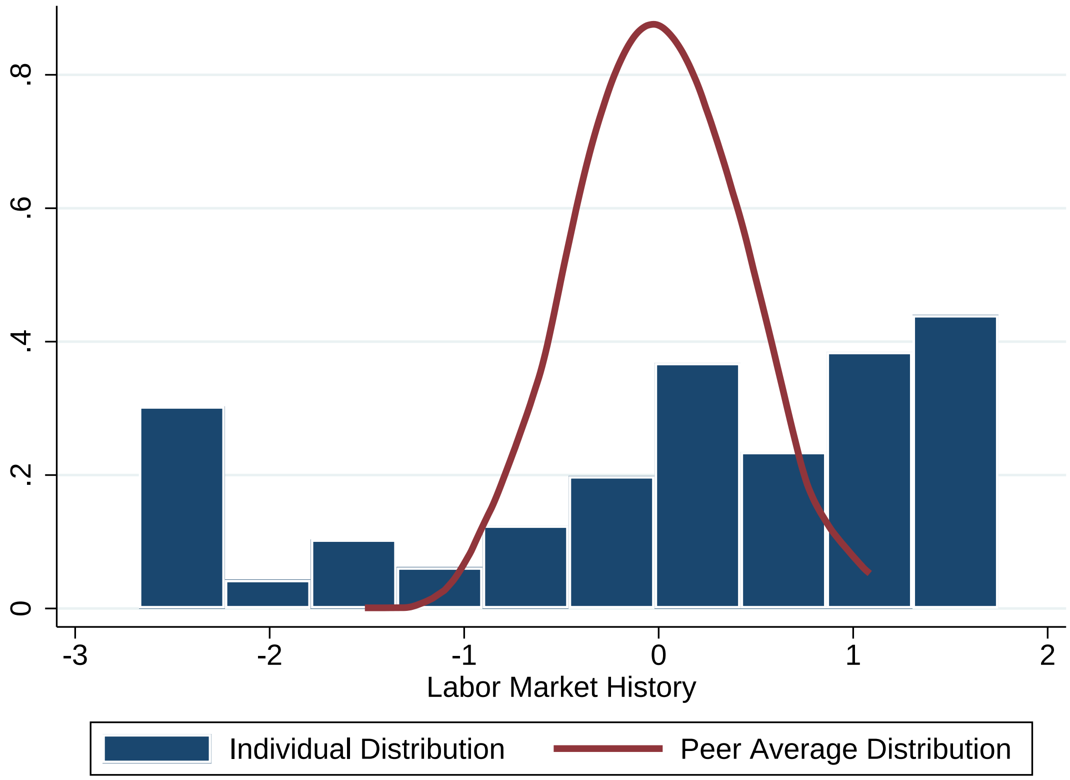
$***p < 1\%, **p < 5\%, *p < 10\%$. An observation is an individual-time period. The standard errors are clustered at the individual-level for the individuals not assigned to the treatment and the training cohort for the individuals assigned to the treatment. While labeled "log wage", we use an inverse hyperbolic sine function to allow for zero value in the outcome. "Total Wage Conditional" uses the total wage as the outcome, but only includes individuals with a positive wage in the quarter.

Table 4: Impact of Treatment Assignment Mechanism on Social Planner's Utility

	Social Planner is Focused On:	
Treatment Assignment	Average Outcomes	Minimum Outcomes
Optimal	1,823	1,329
Naive Optimal	124	388
Random	938	285
Realized	893	308

VIII.B Figures

Figure 1: Labor Market History Distributions



Note: This figure shows the distributions of individual labor market history and peer average labor market history. See Section III.C for information on how this measure is constructed.

Figure 2: Identification Intuition

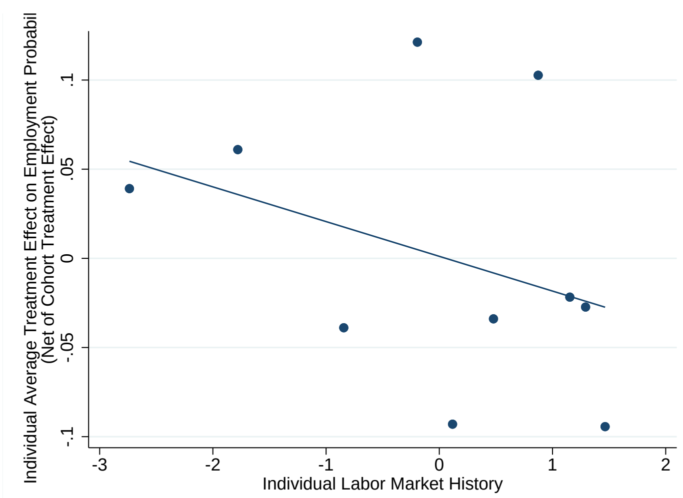
(a) Across Treatment-Cohort Variation

Relationship between Average Treatment Effect and Average Labor Market History

Note: This figure illustrates the across-cohort relationship between a cohort's average treatment effect and the average labor market history of individuals within the treatment. Each dot represents the estimated treatment effect for a single cohort and the size of the dots reflect the number of individuals in the cohort. The line represents the results of a linear regression, weighted by the size of the cohort. For more details about how we measure individuals' labor market history see Section III; for more information on the empirical strategy see Section IV.

(b) Within Treatment-Cohort Variation

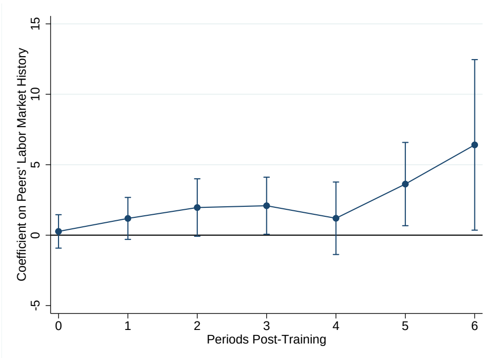
Relationship between Individual Treatment Effect and Individual Labor Market History



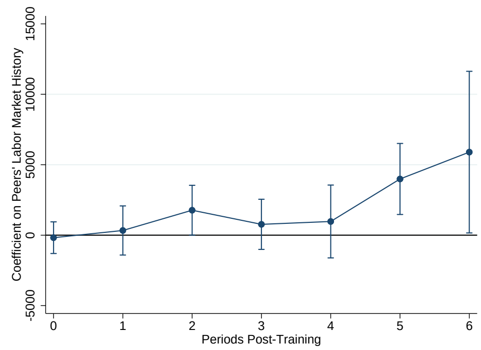
Note: This figure illustrates the within-cohort relationship between an individual's treatment effect and the individual's own labor market history. Each dot represents the difference in average outcomes between the treated and control individuals for each decile of labor market history. Before calculating this difference we adjust the treated individuals' outcomes by subtracting the average impact of the cohort they attended. The blue line reflects the results of an OLS regression. For more details about how we measure individuals' labor market history see Section III; for more information on the the empirical strategy see Section IV.

Figure 3: Impact of Peers Over Time

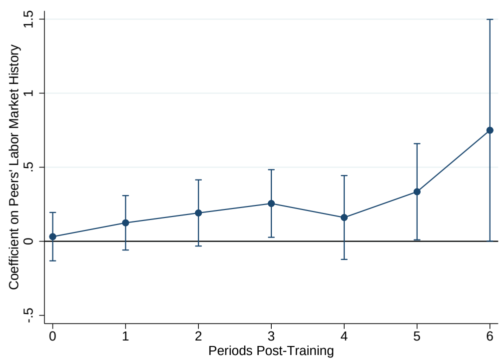
(a) Effect on Log Wages



(b) Effect on Total Wages



(c) Effect on Employment Status



Note: These figures show the how the importance of an individuals' peers evolves in the quarter during and quarters after completing a job training program. The coefficient is estimated using Equation 13, when restricting the sample to the specific quarter; the coefficient being shown here is γ in that expression. The vertical lines highlight the 95% confidence interval. See Section IV for more details on the specification.

A Further Exploration of the Model and Proofs

A.1 General Result

Theorem. 1 *For a compensated change from any γ to $\gamma + \Delta$ with $\Delta > 0$, the average θ_i of the treated individuals after the change is no lower than before the change, i.e.:*

$$\frac{1}{T}\sum_{\forall j}\theta_j T_j^{opt}(\gamma+\Delta) \geq \frac{1}{T}\sum_{\forall j}\theta_j T_j^{opt}(\gamma) \tag{15}$$

Proof. Our approach to the proof is to consider any assignment where the average θ_i of the treated individuals is lower than the optimal assignment before the compensated change, which we refer to as the "initial optimal assignment." We then show that the social planner gains more from switching from this assignment to the initial optimal assignment after the compensated change than before the compensated change. Since the initial optimal assignment is optimal before the compensated change, it follows that none of the treatment assignments where the average θ_i of the treated individuals is lower than in the initial optimal assignment can be optimal after the compensated change.

To show this formally, denote the social planner's utility under treatment assignment $\mathbf{T}$ before the compensated change as $U(\mathbf{T}, \gamma)$ and after the compensated change as $U(\mathbf{T}, \gamma+\Delta)$. Next, consider any $\mathbf{T}'$ such that:[23]

$$\frac{1}{T}\sum_{\forall j}\theta_j T_j^{opt}(\gamma) > \frac{1}{T}\sum_{\forall j}\theta_j T_j' \tag{16}$$

[23]Without the assumption that $\mathbf{T}^{opt}$ is unique, we would consider a $\mathbf{T}'$ that satisfies the constraint below for every choice in $\mathbf{T}^{opt}$. The logic below shows that the minimum value of $\frac{1}{T}\sum_{\forall j}\theta_j T_j^{opt}(\gamma+\Delta)$ is weakly larger than the minimum value of $\frac{1}{T}\sum_{\forall j}\theta_j T_j^{opt}(\gamma)$. To show that $\frac{1}{T}\sum_{\forall j}\theta_j T_j^{opt}(\gamma)$ is increasing in the strong set order, we would also need to show that the maximum value of $\frac{1}{T}\sum_{\forall j}\theta_j T_j^{opt}(\gamma)$ is weakly increasing. Considering a $\mathbf{T}'$ that satisfies the constraint below for at least one choice in $\mathbf{T}^{opt}$ and following the same logic below is sufficient to conclude that.

We then get that:

$$\Big[U(\mathbf{T'},\gamma+\Delta)-U(\mathbf{T^{opt}}(\gamma),\gamma+\Delta)\Big]-\Big[U(\mathbf{T'},\gamma)-U(\mathbf{T^{opt}}(\gamma),\gamma)\Big]= \tag{17}$$

$$\Big[U(\mathbf{T'},\gamma+\Delta)-U(\mathbf{T'},\gamma)\Big]-\Big[U(\mathbf{T^{opt}}(\gamma),\gamma+\Delta)-U(\mathbf{T^{opt}}(\gamma),\gamma)\Big]= \tag{18}$$

$$U(\mathbf{T'},\gamma+\Delta)-U(\mathbf{T'},\gamma)\leq 0 \tag{19}$$

The first equality is just a re-arrangement and the second comes from the fact that under the compensated change utility is unaffected if the treatment choice does not change from the initial optimal choice, i.e. $U(\mathbf{T^{opt}}(\gamma),\gamma+\Delta)=U(\mathbf{T^{opt}}(\gamma),\gamma)$. Finally, the last inequality comes from the fact that the τ_i post-compensation change under treatment assignment $\mathbf{T'}$ minus τ_i pre-compensation change and treatment assignment $\mathbf{T'}$ is equal to: $\Delta\cdot\Big[\frac{1}{T}\sum_{\forall j}\theta_j T'_j-\frac{1}{T}\sum_{\forall j}\theta_j T^{opt}_j(\gamma)\Big]<0$. Since everyone's τ_i is lower, and hence all the treated individuals' y_i is lower, we get that $U(\mathbf{T'},\gamma+\Delta)-U(\mathbf{T'},\gamma)\leq 0$.

Thus, we have that $\Big[U(\mathbf{T'},\gamma+\Delta)\leq U(\mathbf{T^{opt}}(\gamma),\gamma+\Delta)\Big]<\Big[U(\mathbf{T'},\gamma)-U(\mathbf{T^{opt}}(\gamma),\gamma)\Big]$. However, we further know that $U(\mathbf{T'},\gamma)<U(\mathbf{T^{opt}}(\gamma),\gamma)$, since $\mathbf{T^{opt}}$ is the optimal treatment assignment and we assume that the optimal treatment assignment is unique. Thus, it must be that $U(\mathbf{T'},\gamma+\Delta)<U(\mathbf{T^{opt}}(\gamma),\gamma+\Delta)$, and so $\mathbf{T'}$ cannot be the optimal treatment assignment post-compensated change. Since this is true for any $\mathbf{T'}$ that lowers the average θ_i of the treated individuals, the optimal treatment choice post-compensated change must weakly increase the average θ_i of the treated individuals.

Finally, we note that without the assumption that $\mathbf{T}^{opt}$ is unique, we could consider a $\mathbf{T'}$ that satisfies the constraint in Equation (16) for every choice in $\mathbf{T}^{opt}$. The logic above shows that the minimum value of $\frac{1}{T}\sum_{\forall j}\theta_j T^{opt}_j(\gamma+\Delta)$ is weakly larger than the minimum value of $\frac{1}{T}\sum_{\forall j}\theta_j T^{opt}_j(\gamma)$. To show that $\frac{1}{T}\sum_{\forall j}\theta_j T^{opt}_j(\gamma)$ is increasing in the strong set order, we also need to show that the maximum value of $\frac{1}{T}\sum_{\forall j}\theta_j T^{opt}_j(\gamma)$ is weakly increasing. Considering a $\mathbf{T'}$ that satisfies the constraint in Equation (16) for at least one choice in $\mathbf{T}^{opt}$ and following the same logic below is sufficient to conclude that. $\square$

A.2 Maximizing the Average Outcome

In Remark 1, we show that a social planner who cares only about maximizing the average outcome will give assign the treatment slots to the T individuals with the largest values of $\tau(\theta_i) + \gamma\theta_i$. We now add more structure to $\tau(\theta_i)$, which allows us to more explicitly define which individuals receive the treatment.

Homogenous Treatment Effects. *Suppose that all individuals are affected equally by the treatment, i.e. that $\tau(\theta_i) = \tau$ for some τ. Then a social planner who the maximizes average outcome will assign the T individuals with the largest values of θ_i to the treatment.*

Linear Treatment Effect Heterogeneity. *Suppose that $\tau(\theta_i) = \alpha\theta_i + \tau$ for some τ and α. Then a social planner who maximizes the average outcome will assign either the T individuals with the largest values of θ_i or the T individuals with the smallest values of θ_i to the treatment.*

Convex Treatment Effect Heterogeneity. *Suppose that $\tau(\theta_i)$ is such that the treatment effect is most beneficial for those with low values of θ_i, i.e. $\frac{\partial\tau(\theta_i)}{\partial\theta_i} < 0$, and that this differential impact lessens as θ_i increases, i.e. $\frac{\partial^2\tau(\theta_i)}{\partial\theta_i^2} > 0$. Then there exists a K such that the social planner who maximizes the average outcome assigns the treatment to the K individuals with the lowest values of θ_i and the $T - K$ individuals with the largest values of θ_i.*

Proof. The lemmas all follow directly from the fact that the social planner will assign the treatment slots to the T individuals with the largest values of $\tau(\theta_i) + \gamma\theta_i$. □

A.3 Maximizing the Minimum Outcome

Theorem 2 in Section II states that there are two types of individuals who receive treatment: those with the lowest values of $y_i(0)$ and those with the highest θ_i. Below, we re-state the formal theorem and prove the result.

Theorem. 2 *Suppose that a social planner chooses $\boldsymbol{T}$ to maximize the minimum outcome. Then the optimal allocation of treatment slots involves assigning the treatment to the K*

individuals with the lowest values of $y_i(0)$ and the $T-K$ individuals with the highest values of θ_i.

Proof. Let individual L be any individual that is untreated under the optimal allocation. We will then show that if individual K is treated in the optimal allocation and $\theta_L > \theta_K$, it must be the case that $y_K(0) \leq y_L(0)$, which is sufficient to prove the theorem.

To do so, we will add a bit of notation. First, we will denote $y_i(\mathbf{T})$ as individual i's allocation under treated assignments $\mathbf{T}$. Similarly, we will denote the Rawlsian social planner's utility under an allocation $\mathbf{T}$ as $U(\mathbf{T}) \equiv min\Big\{y_1(\mathbf{T}), y_2(\mathbf{T}), ..., y_N(\mathbf{T})\Big\}$. We will then compare the optimal allocation, denoted as $\mathbf{T^{opt}}$, to the the allocation that is identical to $\mathbf{T^{opt}}$ except for the fact that K is not treated and L is; we denote the second allocation as $\mathbf{T'}$.

To do so, we start by showing that if $\theta_L > \theta_K$, then K must be the individual with the minimum value of y_i under the treatment assignment $\mathbf{T'}$. To do so, we first note that:

$$y_i(\mathbf{T^{opt}}) - y_i(\mathbf{T'}) = \begin{cases} \gamma \frac{1}{T-1}\Big(\theta_K - \theta_L\Big) \cdot T_i \text{ if } i \neq K, L \\ \tau(\theta_K) + \gamma \frac{1}{T-1} \sum_{\forall j \neq K} \theta_j T_j \text{ if } i = K \\ -\tau(\theta_L) - \gamma \frac{1}{T-1} \sum_{\forall j \neq J} \theta_j T_j \text{ if } i = L \end{cases} \tag{20}$$

This comes from the fact that the treatment assignment for all individuals other than K and L is the same between $\mathbf{T^{opt}}$ and $\mathbf{T'}$, which allows us to write this without clarifying whether T_i refers to the i's assignment under $\mathbf{T^{opt}}$ or $\mathbf{T'}$, unless $i = K$ or L.

Denoting M as the individual with the minimum value of y_i in $\mathbf{T'}$, we next note that $y_M(\mathbf{T^{opt}}) - y_M(\mathbf{T'}) \geq U(\mathbf{T^{opt}}) - U(\mathbf{T'})$. This follows from the fact that $U(\mathbf{T'}) = y_M(\mathbf{T'})$ and $y_M(\mathbf{T^{opt}}) \geq U(\mathbf{T^{opt}})$. We can then conclude that $y_M(\mathbf{T^{opt}}) - y_M(\mathbf{T'}) \geq 0$, since by definition of optimal it must be the case that $U(\mathbf{T^{opt}}) \geq U(\mathbf{T'})$. Finally, Equation (20) shows that individual K is the only individual whose y_i increases when moving from $\mathbf{T'}$ to $\mathbf{T^{opt}}$ and so individual K must be the one with the minimum value of y_i in $\mathbf{T'}$.[24]

[24]The fact that individual L's outcome decreases relies on the assumption that $\tau_i > 0$ for all i.

Importantly, the fact that individual K is the individual minimum value of y_i in $\mathbf{T}'$ and the fact that by assumption individual K is not assigned to the treatment in $\mathbf{T}'$, we know that $U(\mathbf{T}') = y_K(0)$. We also know that $U(\mathbf{T^{opt}}) \leq y_L(0)$, since individual L is untreated in $\mathbf{T^{opt}}$. Thus, $y_L(0) - y_K(0) \geq U(\mathbf{T^{opt}}) - U(\mathbf{T}')$. Since $\mathbf{T^{opt}}$ is the optimal assignment, it follows that $U(\mathbf{T^{opt}}) - U(\mathbf{T}') \geq 0$ and so $y_L(0) \geq y_K(0)$. □

B Residualization Motivation

Here we discuss the motivation behind our residualization approach, which stems in part from the following theorem.

Theorem 3. *Let T_i be an indicator that denotes whether individual i was assigned to treatment. Further, suppose that all Z_i and X_i are not affected by the treatment and that treatment is randomly assigned. Consider two OLS regressions:*

$$y_i = \hat{\alpha_0} + \hat{\beta}_0 X_i + \hat{\tau_0} T_i + \hat{\gamma_0} X_i \cdot T_i \tag{21}$$

$$y_i - g(Z_i) = \hat{\alpha_1} + \hat{\beta}_1 X_i + \hat{\tau_1} T_i + \hat{\gamma_1} X_i \cdot T_i \tag{22}$$

Then under standard assumptions, both $\hat{\tau_0}$ and $\hat{\tau_1}$ converge to the same τ and, similarly, both $\hat{\gamma_0}$ and $\hat{\gamma_1}$ coverage to the same γ regardless of the function $g(Z_i)$.

Proof. Given the linearity, it is sufficient to show that in a regression of the form:

$$g(Z_i) = \hat{\alpha_2} + \hat{\beta}_2 X_i + \hat{\tau_2} T_i + \hat{\gamma_2} X_i \cdot T_i \tag{23}$$

we get that $\hat{\tau_2} \to 0$ and $\hat{\gamma_2} \to 0$. Next we define a and b as the linear projection parameters of $g(Z_i)$ onto X_i and a dummy, i.e. $g(Z_i) = a + bX_i + e_i$ where $\mathbb{E}[e_i] = 0$ and $\mathbb{E}[e_i X_i] = 0$. Note that the assumption that X_i and Z_i are unaffected by treatment and T_i is randomly assignment, ensures that $\mathbb{E}[f(X_i, Z_i)|T_i = 1] = \mathbb{E}[f(X_i, Z_i)|T_i = 0]$ for any function of X_i and Z_i.[25] Since e_i is a function of Z_i and X_i and $\mathbb{E}[e_i] = 0$ and $\mathbb{E}[e_i X_i] = 0$, it therefore

[25]Note that this omits some technical details required to ensure that the expectations exist.

follows that $\mathbb{E}[e_i T_i] = 0$ and $\mathbb{E}[e_i X_i T_i] = 0$. Thus, under the standard assumptions: $\hat{\alpha}_2 \to a$, $\hat{\beta}_2 \to b$, $\hat{\tau}_2 \to 0$, and $\hat{\gamma}_2 \to 0$.

$\square$

We next turn our attention to the question of what function $g(Z_i)$ is optimal. Given that we want to minimize the variance of our estimated peer effect we want to choose $g(Z_i)$ to minimize $\mathbf{X}'\Sigma\mathbf{X}$ where Σ is the variance-covariance matrix of the error terms $\epsilon_{i,k,t}$ in Equation (13) and $\mathbf{X}$ is the matrix of covariates from Equation (13). However, we cannot do this directly without knowing the parameters we aim to estimate or the correlation structure between outcomes. One could push harder on this approach and potentially develop a feasible generalized least squares, in which one iteratively: a) estimates Equation (13) inefficiently, b) uses these coefficients to approximate Σ; c) use this to determine the optimal $g(Z_i)$; d) estimate $\hat{g}(Z_i)$; and then re-estimate Equation (13) using the resulting residuals. Instead, we opt for a simpler approach and note that setting $g(Z_i) = \frac{1}{2}\mathbb{E}[y_{i,k,t}|Z_i, T_i = 1] + \frac{1}{2}\mathbb{E}[y_{i,k,t}|Z_i, T_i = 0]$ minimizes $\mathbb{V}\Big(y_i - g(Z_i)\Big)$. We therefore estimate $\frac{1}{2}\hat{\mathbb{E}}[y_{i,k,t}|Z_i, T_i = 1] + \frac{1}{2}\hat{\mathbb{E}}[y_{i,k,t}|Z_i, T_i = 0]$ and use this to residualize the outcome which, while not optimal, significantly reduces the variance of the estimated peer effects.

Note finally that while not strictly necessary, doing the residualization in a separate step than the regression allows us to easily estimate a non-linear function $\hat{g}(Z_i)$, which we do using a random forest regression, while sticking with a linear specification to estimate the main effects.

C Robustness Checks

C.1 IV Regressions

As in most experiments, compliance was not perfect in this study. For the main results, we ignore this complication and use "assigned to treatment" in place of "attended treatment." Since we observe whether an individual attended treatment, we can use the average labor market history of those assigned to treatment as an instrument for the average labor market history of those who actually attended the training. The results, shown below in Table A1, show that doing so increases the magnitude of the estimated peer effects. These changes are all relatively small in magnitude, however, and the IV method does not change the statistical significance or economic interpretation of the empirical results. This is due to the fact that compliance was relatively high (~85% of those assigned treatment attended the training and none assigned control attended) and is the reason why we ignore imperfect compliance in the main specification.

C.2 Residualization Alternatives

The motivation for our residualization approach is discussed above; here we show that alternative approaches provide a similar story. First, instead of residualizing the outcome here we include the prediction from the random forest regression as a covariate in the final regression. This is shown in the first four columns of Table A2. The results illustrate that including the prediction as a residual leads to identical conclusions as using it to residualize the outcome. In addition, we show the results when no additional covariates are added. The results are next four columns of Table A2. The standard errors of both approaches are larger than in our main specification, and the point estimate of the peer effect on total wages is lower when no covariates are included than when the outcome is residualized or when the prediction is used as a covariate. However, in all specifications we find evidence of peer effects on the probability that an individual is employed and on her log wage.

C.3 Interacting the Fixed Effects

In our main specifications, we include cohort, strata, and time fixed effects. In Table A3, we show that interacting the cohort and strata fixed effects gives similar results. In fact, the estimated magnitude of the peer effects increases in this specification across all four outcomes. We also get broadly similar results with interacting all three sets of fixed effects, i.e., including cohort-by-strata-by-time period fixed effects.

C.4 Alternative Specification

As discussed in Section IV, we also estimate a specification that is more similar to the traditional specification that estimates exogenous peer effects, e.g., see discussion in Section 4.1 in Caeyers and Fafchamps (2020) on exogenous peer effect specifications. However, we need to account for the fact that, unlike most peer effect contexts, not all groups within the cohort are identical, as some were assigned to the treatment group and some were assigned to the control group. We account for this by: a) allowing for the mean to be different in the treatment group and control group, i.e., allowing for an average treatment effect; b) interacting the individuals' own labor market history with a treatment dummy to allow for individually heterogeneous treatment effects; c) interacting the individuals' peers' labor market history with a treatment dummy to allow for the impact of the peers to differ depending on whether the individual is assigned to the treatment group or control group. More specifically, in this alternative specification we run the following regression:

$$r_{i,k,t} = \tau T_i + \Big(\beta_0 \theta_i + \beta_1 \theta_i T_i\Big) + \Big(\gamma_0 \overline{\theta}_{i,k} + \gamma_1 \overline{\theta}_{i,k} T_i\Big) + \nu_i + \delta_k + \eta_t + \epsilon_{i,k,t} \tag{24}$$

where again:

- $r_{i,k,t}$ is the residualized outcome for individual i who applied to cohort k in period t;
- T_i is an indicator for whether individual i was randomly assigned treatment;
- θ_i is individual i's labor market history, described Section III;
- $\overline{\theta}_{i,k}$ is the average labor market history for individuals in the same cohort k and treatment status as individual i, defined in more detail in Section IV
- δ_k is a fixed effects indicating which cohort individual i applied for;
- ν_i is a vector of strata fixed effects, indicating which strata individual i is in; the strata are defined by gender, income, employment status, and age all measured at the time of randomization;

- η_t are period fixed effects.
- $\epsilon_{i,k,t}$ is an individual error term.

Note that while motivated differently, the only difference between this regression and our main specification is that this includes $\overline{\theta}_{i,k}$ rather than only including the interaction term. The logic for only including the interaction term is that the control groups' "peers" are not peers in the sense that there is in fact no interaction between them. We should therefore expect the effect of $\overline{\theta}_{i,k}$ on the control group to be zero. In fact, this can serve as a nice placebo effect. The results can be seen in Table A4. The top three rows show that the main coefficients do not change much in this specification than in the main specification, with the estimated peer effect actually increasing slightly across all four outcomes. Furthermore, the fourth row shows that the effect of $\overline{\theta}_{i,k}$ on the control group is not statistically significant in any of the outcomes. This provides additional evidence that the peer effect estimates reported in this paper are indeed the impact of the peer group composition on the effectiveness of the job training program.

C.5 Labor Market History Components

In our main measure of an individuals' labor market history we average four components into one summary measure. As mentioned in Section III, these measures are: a) the number of quarters employed in the two years before randomization; b) the total earnings in the two years before randomization; c) total earnings in the year before the screening occurs; d) an indicator for whether individuals are working at the time of screening. Here, we estimate the main model separately for each component of our summary measure as well as varying whether wages are measured in levels or logs. Specifically, we re-estimate Equation (13), this time using each of the five potential measures of labor market history as our measure of θ_i, rather than their average. The results are shown in Table A5, which also includes in the last column the estimate when using the summary measure as our proxy for θ_i. As can be seen, the point estimates are broadly similar across all measures, although the peer effect estimate is often not statistically significant due in part to larger standard errors when θ_i is approximated using a single measure instead of the multiple measures.

C.6 Pathway and Area-Specific Treatment Effects

While the program consisted of 20 cohorts, some of these cohorts differed in what pathway and area they were focused on. As shown in Table 1, there were three pathways (advanced manufacturing, health care, and information technology) and six areas. In the main specification, we run a regression that pools all pathways/areas and estimates a single average treatment effect parameter, along with a single peer effect parameter and amount of individual treatment effect heterogeneity. Data constraints mean that we cannot precisely estimate the peer effect separately for every pathway/area - or equivalently we cannot fully interact all the parameters with pathway/area indicators - but we can allow for the average effect of the program and the individual treatment effect heterogeneity to vary across pathways and/or areas.[26] As shown in Table A6, doing so does not meaningfully change the peer effect estimates.

D Simulations

D.1 Testing for Exogeneity

At first blush, testing whether individuals are randomly assigned seems straightforward: simply test whether ones' peers' average characteristics are correlated with their own. Doing so via a simple regression is complicated by the fact that the peers' average characteristic is generally calculated using a leave-one-out approach, leading to negative correlation between individuals' own characteristic and their peers' average characteristics even when peers are randomly assigned.[27] In response to these issues, there is a small literature that discusses

[26] As a further check, we also ran the analysis separately for each pathway and, although noisily estimated, the point estimates on the peer effect parameter is positive in all three pathways.

[27] Allowing for an individual to be their own peer does not solve the problem, as it causes a positive correlation between the two measures even under random assignment.

approaches to correct this bias and proposes alternative tests that are asymptotically valid (Sacerdote, 2001; Guryan et al., 2009; Jochmans, 2021; Caeyers and Fafchamps, 2020).

While the test developed in Jochmans (2021) suggests that we cannot reject the assumption that individuals are randomly assigned to groups (p = 0.317), rather than appeal to asymptotics our preferred approach is test for the exogeneity of group assignment via simulation.[28] In particular, we note that if individuals endogenously sort into cohorts, we would expect the across-cohort variance in average labor market history to be significantly larger than if individuals were randomly assigned to treatment.[29]

We therefore test the null hypothesis that individuals are randomly sorted into cohorts by randomly re-assigning individuals to cohorts and then calculating the across-cohort variation in average labor market history based on these alternative assignment. We repeat this process 1,000 times to determine the distribution of the test statistic under the null hypothesis.

The result is shown in Figure A1, where the histogram shows the simulated distribution under the null and the red vertical line shows the realized outcome. As can be seen, the realized measure is consistent with random assignment of individuals to peers. Specifically, we find that 23.6% of the simulated across-cohort variance are larger than the realized outcome, which corresponds to a p-value of 0.476 in a two-sided hypothesis test.

Finally, we note that the proposed test has enough statistical power to reject the null in cases where there is not random assignment. Stated differently, we did not fail to reject the null only because we used a weak hypothesis test. For example, we use the same approach to test whether students are randomly assigned to middle schools, using their elementary school tests as the relevant peer characteristic. Even when restricting our sample to the same size as in this context, we reject the null at every level of statistical significance. In

[28] Particularly problematic is the fact that the asymptotic result in Jochmans (2021) involves letting the number of "urns," i.e., the groups within which you are testing for random assignment, grow. We are testing the null hypothesis that there is random assignment to cohorts, which implies that there is a single urn in our test. It is therefore a different asymptotic argument than needed for our standard errors to be valid, for example.

[29] This is similar in spirit, although different in practice, as the test in Lavy and Schlosser (2011).

fact, in none of the 1,000 simulations do we find the across-school variance in previous test scores to be larger than the observed across-school variance.

In addition to the simulation, which focuses on how individuals' sort based on their labor market history, we can also test whether there are any apparent peer effects when the outcome is replaced with a pre-determined characteristic. To do so, we run our main empirical specification as outlined in Section IV. However, instead of using individuals' actual labor market outcomes, we use their expected labor market outcomes as a function of their pre-randomization outcomes. As shown in Table A7, although we find peer effects when using individuals' actual outcomes we do not find peer effects when using their predicted outcomes. In addition, the fact that we find no peer effects in the first quarter (see Figure 3) also provides a placebo test of sorts, as it is likely too soon for the effects to appear in the first quarter. These results gives further credence to our empirical strategy.

D.2 Estimated Peer Effects Under Alternative Assignments

To confirm that our results and hypothesis tests are valid, we conduct a simulation to determine how often we find peer effects under alternative assignments. In this simulation, we randomly assigned individuals to cohort/treatment groups, use those simulated groups to compute the individuals' peer quality, and then re-estimate the main specification. Note that in this simulation, we hold fixed an individuals' *own* cohort/treatment assignment. We do so because we want to test the null hypothesis that there are no peer effects in the program, not the sharp null hypothesis that treatment status (and therefore groupings) do not matter for outcomes.

The results of this simulation are shown in Figure A2. Like in Figure A1, the histograms show the simulated distribution under the null and the red vertical lines show the realized outcome. Unlike in Figure A1, which tested the null hypothesis that individuals were randomly sorted into groups, here we find that the estimated effect is unlikely to be explained by chance. We find that only a handful of potential peer groups suggest the presence of peer effects larger than realized peer group. Specifically, we find that only 1.7, 10.4, and

2.1 percent show larger peer effects, depending on whether the outcome is employment probability, log wage, or total wage, respectively. Furthermore, in all three cases the mean estimate of peer effects under these hypothetical groupings is approximately zero, which suggests that the results presented above are not biased.

D.3 Simulation with Added Noise to Variable

As discussed in Feld and Zölitz (2017) and Angrist (2014), measurement error in the covariates can increase the peer effect estimates when the peer groups are not formed under random assignment. Of course, in our context the peer groups are in large part formed by the random assignment of individuals to treatment, at least within cohort. Since there is non-random assignment of individuals to cohort, however, we follow Feld and Zölitz (2017)'s suggestion and conduct a Monte Carlo simulation to ensure that measurement error does not bias upward the coefficient estimates.

We start the simulation by adding an extra error term to each individual's labor market history, with the error term drawn from a normal distribution with mean zero and standard deviation that varies from zero to one. We use this noisy measure of each individual's labor market history as the new measure of θ_i. Given this, we then estimate the regression specified in Equation (13) to generate new parameter estimates.

We repeat this procedure 250 times for each standard deviations ranging from 0.25 to 1. Figure A3 shows the average peer effect estimate of these simulations for each standard deviation. As the standard deviation of the error term increases, the size of the estimated peer effect decreases.

E Appendix Tables and Figures

E.1 Appendix Tables

Table A1: IV Regressions

	(1) Log Wage	(2) Total Wage	(3) Employed	(4) Total Wage Conditional on Employment
Impact of Peers' Ave. Labor Market History on	2.275***	1561.9*	0.246***	1254.6
Treatment Effect	(0.867)	(882.9)	(0.0929)	(925.0)
Impact of Own Labor Market History on	-0.299	-452.0**	-0.0287	-782.5***
Treatment Effect	(0.221)	(209.8)	(0.0230)	(216.8)
Average Treatment Effect	0.859***	895.7***	0.0721**	1272.6***
	(0.248)	(268.2)	(0.0323)	(302.8)
Cohort Fixed Effects	X	X	X	X
Time Period Fixed Effects	X	X	X	X
Strata Fixed Effects	X	X	X	X
Estimator	IV Regression	IV Regression	IV Regression	IV Regression
Sample	Post-Training	Post-Training	Post-Training	Post-Training
Number of Clusters	174	174	174	139
Observations	1632	1632	1632	1095

*** p<1%, ** p<5%, * p<10%. An observation is a individual-time period. The standard errors are clustered at the individual-level for the individuals not assigned to the treatment and the training cohort for the individuals assigned to the treatment. While labeled "log wage," we use an inverse hyperbolic sine function to allow for zero value in the outcome. In this regression, we use the assigned treatment assignment as an IV for the realized treatment assignment.

Table A2: Residualization Alternatives

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	8)
	Log Wage	Total Wage	Employed	Total Wage Conditional on Employment	Log Wage	Total Wage	Employed	Total Wage Conditional on Employment
Impact of Peers' Ave. Labor Market History on Treatment Effect	1.940**	1293.1*	0.212**	911.7	1.568*	708.0	0.180*	-89.54
	(0.763)	(781.0)	(0.0818)	(852.4)	(0.884)	(948.0)	(0.0933)	(1014.5)
Impact of Own Labor Market History on Treatment Effect	-0.220	-419.7*	-0.0195	-771.4***	-0.0116	-352.8	0.00545	-764.6***
	(0.223)	(219.4)	(0.0227)	(220.2)	(0.288)	(263.7)	(0.0305)	(254.7)
Treatment Effect at Averages	0.760**	906.3***	0.0724**	1257.3***	0.802***	729.3**	0.0608*	942.0***
	(0.294)	(266.3)	(0.0326)	(305.2)	(0.234)	(326.3)	(0.0320)	(358.1)
Cohort Fixed Effects	X	X	X	X	X	X	X	X
Time Period Fixed Effects	X	X	X	X	X	X	X	X
Strata Fixed Effects	X	X	X	X	X	X	X	X
Other Controls	Predicted Outcome	Predicted Outcome	Predicted Outcome	Predicted Outcome	None	None	None	None
Sample	Post-Training	Post-Training	Post-Training	Post-Training	Post-Training	Post-Training	Post-Training	Post-Training
Number of Clusters	174	174	174	139	174	174	174	139
Observations	1632	1632	1632	1095	1632	1632	1632	1095

*** p<1%, ** p<5%, * p<10%. An observation is a individual-time period. The standard errors are clustered at the individual-level for the individuals not assigned to the treatment and the training cohort for the individuals assigned to the treatment. While labeled "log wage," we use an inverse hyperbolic sine function to allow for zero value in the outcome. Unlike the main specification, these outcomes are not resididualized prior to the regression.

Table A3: Interacting Fixed Effects

	(1) Log Wage	(2) Total Wage	(3) Employed	(4) Total Wage Conditional on Employment
Impact of Peers' Ave. Labor Market History on Treatment Effect	3.128***	2561.9***	0.324***	1938.6*
	(0.969)	(965.6)	(0.105)	(1139.3)
Impact of Own Labor Market History on Treatment Effect	-0.310	-409.0**	-0.0272	-685.6***
	(0.233)	(206.4)	(0.0240)	(207.3)
Average Treatment Effect	1.018***	1156.9***	0.0939***	1789.8***
	(0.294)	(279.2)	(0.0330)	(287.7)
Cohort-by-Strata Fixed Effects	X	X	X	X
Time Period Fixed Effects	X	X	X	X
Sample	Post-Training	Post-Training	Post-Training	Post-Training
Number of Clusters	164	164	164	129
Observations	1618	1618	1618	1076

*** p<1%, ** p<5%, * p<10%. An observation is a individual-time period. The standard errors are clustered at the individual-level for the individuals not assigned to the treatment and the training cohort for the individuals assigned to the treatment. While labeled "log wage," we use an inverse

Table A4: Alternative Specification

	(1) Log Wage	(2) Total Wage	(3) Employed	(4) Total Wage Conditional on Employment
Impact of Peers' Ave. Labor Market History on Treatment Effect	2.730***	1992.0**	0.299***	1161.7
	(0.926)	(947.8)	(0.0995)	(1088.8)
Impact of Own Labor Market History on Treatment Effect	-0.266	-429.5*	-0.0252	-772.8***
	(0.224)	(217.6)	(0.0230)	(224.2)
Average Treatment Effect	0.864***	978.7***	0.0841***	1331.1***
	(0.290)	(267.0)	(0.0320)	(294.2)
Cohort-by-Treatment Group Ave. Labor Market History	-1.053	-898.2	-0.120	-27.08
	(0.849)	(732.3)	(0.0948)	(824.8)
Cohort Fixed Effects	X	X	X	X
Time Period Fixed Effects	X	X	X	X
Strata Fixed Effects	X	X	X	X
Sample	Post-Training	Post-Training	Post-Training	Post-Training
Number of Clusters	174	174	174	139
Observations	1632	1632	1632	1095

*** p<1%, ** p<5%, * p<10%. An observation is a individual-time period. The standard errors are clustered at the individual-level for the individuals not assigned to the treatment and the training cohort for the individuals assigned to the treatment. Cohort-by-Treatment Group Ave. Labor Market History is the average labor market history of an individuals' peer group, defined as their cohort-by-treatment group. The "Impact of Peers' Ave. Labor Market History on Treatment Effect" is the coefficient of this measure interacted with a treatment dummy. While labeled "log wage," we use an inverse hyperbolic sine function to allow for zero value in the outcome.

Table A5: Labor Market History Components

(a) Total Wages

	(1) Total Wage	(2) Total Wage	(3) Total Wage	(4) Total Wage	(5) Total Wage	(6) Total Wage
Impact of Peers' Ave. Labor Market History on Treatment Effect	1921.1**	804.2	1300.3	921.2	1082.5	1390.8*
	(880.2)	(1021.3)	(815.7)	(910.3)	(902.3)	(777.3)
Impact of Own Labor Market History on Treatment Effect	-287.9	-431.8	-164.6	-439.3	-323.6	-430.9**
	(305.7)	(454.0)	(101.4)	(310.0)	(327.4)	(217.7)
Average Treatment Effect	908.6***	2043.1***	899.9***	911.7***	935.9***	0.0838**
	(275.3)	(729.7)	(276.5)	(274.9)	(265.3)	(0.0324)
Cohort Fixed Effects	X	X	X	X	X	X
Time Period Fixed Effects	X	X	X	X	X	X
Strata Fixed Effects	X	X	X	X	X	X
Sample	Post-Training	Post-Training	Post-Training	Post-Training	Post-Training	Post-Training
Labor Market History Measure	Employed At Screening	Total Wage at Screening	Log Wage at Screening	Wage in Two Years Prior to Screening	Fraction of Quarters Employed in Two Years Prior to Screening	Average Labor Market History
Number of Clusters	174	174	174	174	174	174
Observations	1632	1632	1632	1632	1632	1632

*** p<1%, ** p<5%, * p<10%. An observation is a individual-time period. The standard errors are clustered at the individual-level for the individuals not assigned to the treatment and the training cohort for the individuals assigned to the treatment.

(b) Log Wages

	(1) Log Wage	(2) Log Wage	(3) Log Wage	(4) Log Wage	(5) Log Wage	(6) Log Wage
Impact of Peers' Ave. Labor Market History on Treatment Effect	2.811***	1.303	2.030**	1.417	1.438	2.026***
	(0.860)	(1.047)	(0.823)	(0.955)	(0.947)	(0.756)
Impact of Own Labor Market History on Treatment Effect	-0.120	-0.210	-0.0990	-0.301	-0.167	-0.268
	(0.357)	(0.510)	(0.103)	(0.301)	(0.316)	(0.225)
Average Treatment Effect	0.866***	0.719**	1.449*	0.752**	0.755**	0.814***
	(0.293)	(0.304)	(0.802)	(0.302)	(0.301)	(0.293)
Cohort Fixed Effects	X	X	X	X	X	X
Time Period Fixed Effects	X	X	X	X	X	X
Strata Fixed Effects	X	X	X	X	X	X
Sample	Post-Training	Post-Training	Post-Training	Post-Training	Post-Training	Post-Training
Labor Market History Measure	Employed At Screening	Total Wage at Screening	Log Wage at Screening	Wage in Two Years Prior to Screening	Fraction of Quarters Employed in Two Years Prior to Screening	Average Labor Market History
Number of Clusters	174	174	174	174	174	174
Observations	1632	1632	1632	1632	1632	1632

*** p<1%, ** p<5%, * p<10%. An observation is a individual-time period. The standard errors are clustered at the individual-level for the individuals not assigned to the treatment and the training cohort for the individuals assigned to the treatment. While labeled "log wage," we use an inverse hyperbolic sine function to allow for zero value in the outcome.

(c) Employed

	(1) Employed	(2) Employed	(3) Employed	(4) Employed	(5) Employed	(6) Employed
Impact of Peers' Ave. Labor Market History on Treatment Effect	0.301***	0.150	0.218**	0.163	0.160	0.219***
	(0.0905)	(0.110)	(0.0886)	(0.104)	(0.104)	(0.0811)
Impact of Own Labor Market History on Treatment Effect	-0.0101	-0.0196	-0.00927	-0.0305	-0.0143	-0.0254
	(0.0379)	(0.0550)	(0.0105)	(0.0319)	(0.0335)	(0.0231)
Average Treatment Effect	0.0838**	0.0677**	0.137	0.0720**	0.0721**	0.0784**
	(0.0324)	(0.0335)	(0.0833)	(0.0331)	(0.0331)	(0.0323)
Cohort Fixed Effects	X	X	X	X	X	X
Time Period Fixed Effects	X	X	X	X	X	X
Strata Fixed Effects	X	X	X	X	X	X
Sample	Post-Training	Post-Training	Post-Training	Post-Training	Post-Training	Post-Training
Labor Market History Measure	Employed At Screening	Total Wage at Screening	Log Wage at Screening	Wage in Two Years Prior to Screening	Fraction of Quarters Employed in Two Years Prior to Screening	Average Labor Market History
Number of Clusters	174	174	174	174	174	174
Observations	1632	1632	1632	1632	1632	1632

*** p<1%, ** p<5%, * p<10%. An observation is a individual-time period. The standard errors are clustered at the individual-level for the individuals not assigned to the treatment and the training cohort for the individuals assigned to the treatment.

(d) Conditional Wages

	(1) Total Wage Conditional on Employment	(2) Total Wage Conditional on Employment	(3) Total Wage Conditional on Employment	(4) Total Wage Conditional on Employment	(5) Total Wage Conditional on Employment	(6) Total Wage Conditional on Employment
Impact of Peers' Ave. Labor Market History on Treatment Effect	1110.5	602.2	1073.0	633.0	883.5	1142.9
	(1080.3)	(959.1)	(879.3)	(929.8)	(948.7)	(846.4)
Impact of Own Labor Market History on Treatment Effect	-824.2***	-727.1*	-299.8***	-665.1*	-626.6	-773.0***
	(302.2)	(369.8)	(103.5)	(377.5)	(411.8)	(223.9)
Average Treatment Effect	1292.3***	1274.2***	3348.6***	1241.9***	1235.7***	1330.1***
	(299.3)	(296.9)	(810.3)	(319.0)	(317.4)	(297.6)
Cohort Fixed Effects	X	X	X	X	X	X
Time Period Fixed Effects	X	X	X	X	X	X
Strata Fixed Effects	X	X	X	X	X	X
Sample	Post-Training	Post-Training	Post-Training	Post-Training	Post-Training	Post-Training
Labor Market History Measure	Employed At Screening	Total Wage at Screening	Log Wage at Screening	Wage in Two Years Prior to Screening	Fraction of Quarters Employed in Two Years Prior to Screening	Average Labor Market History
Number of Clusters	139	139	139	139	139	139
Observations	1095	1095	1095	1095	1095	1095

*** p<1%, ** p<5%, * p<10%. An observation is a individual-time period. The standard errors are clustered at the individual-level for the individuals not assigned to the treatment and the training cohort for the individuals assigned to the treatment.

Table A6: Pathway and Area-Specific Treatment Effects

(a) Pathway Specific Treatment Effects

	(1) Log Wage	(2) Total Wage	(3) Employed	(4) Total Wage Conditional on Employment
Impact of Peers' Ave. Labor Market History on Treatment Effect	2.122***	1391.3**	0.237***	1119.1
	(0.665)	(660.3)	(0.0721)	(688.3)
Cohort Fixed Effects	X	X	X	X
Time Period Fixed Effects	X	X	X	X
Strata Fixed Effects	X	X	X	X
Pathway-Specific Individual Treatment Effects	X	X	X	X
Sample	Post-Training	Post-Training	Post-Training	Post-Training
Number of Clusters	174	174	174	139
Observations	1632	1632	1632	1095

*** p<1%, ** p<5%, * p<10%. An observation is a individual-time period. The standard errors are clustered at the individual-level for the individuals not assigned to the treatment and the training cohort for the individuals assigned to the treatment. While labeled "log wage," we use an inverse hyperbolic sine function to allow for zero value in the outcome.

(b) Area Specific Treatment Effects

	(1) Log Wage	(2) Total Wage	(3) Employed	(4) Total Wage Conditional on Employment
Impact of Peers' Ave. Labor Market History on Treatment Effect	2.280***	1716.1**	0.250***	1372.1*
	(0.690)	(684.7)	(0.0744)	(715.1)
Cohort Fixed Effects	X	X	X	X
Time Period Fixed Effects	X	X	X	X
Strata Fixed Effects	X	X	X	X
Area-Specific Individual Treatment Effects	X	X	X	X
Sample	Post-Training	Post-Training	Post-Training	Post-Training
Number of Clusters	174	174	174	139
Observations	1632	1632	1632	1095

*** p<1%, ** p<5%, * p<10%. An observation is a individual-time period. The standard errors are clustered at the individual-level for the individuals not assigned to the treatment and the training cohort for the individuals assigned to the treatment. While labeled "log wage," we use an inverse hyperbolic sine function to allow for zero value in the outcome.

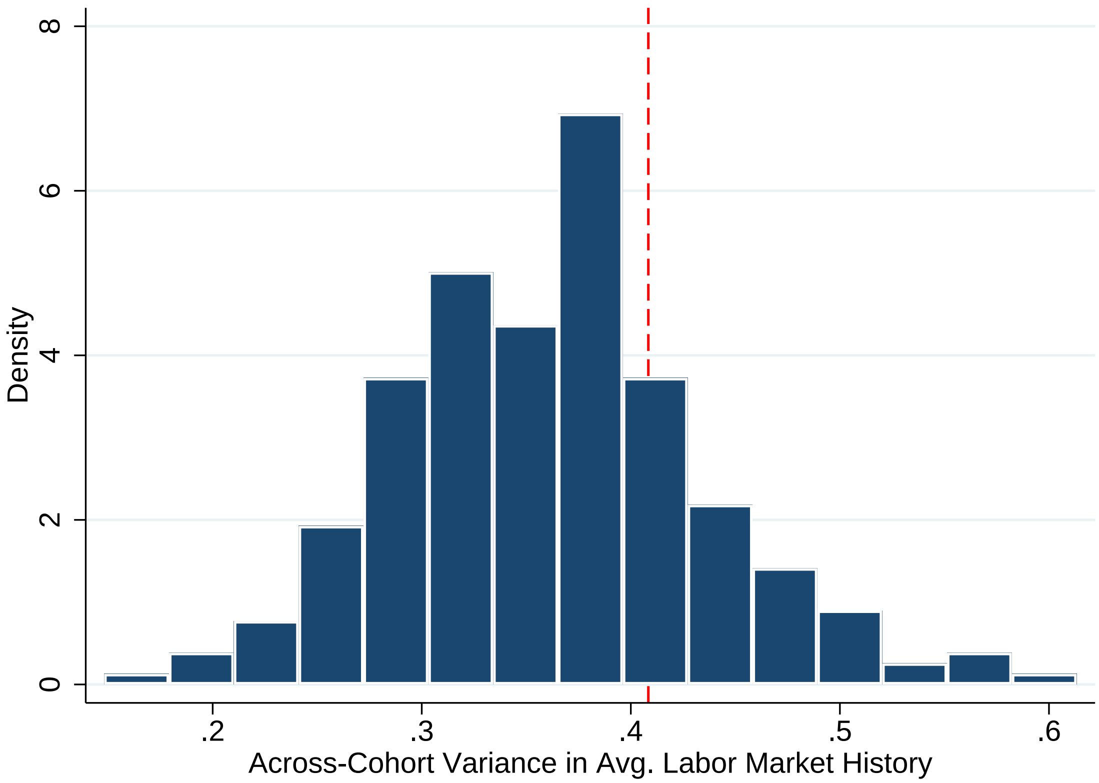
Table A7: Expected Outcomes Based on Pre-Randomization Measures

	(1) Expected Log Wage	(2) Expected Total Wage	(3) Expected Employed	(4) Total Wage Conditional on Employment
Impact of Peers' Ave. Labor Market History on Treatment Effect	-0.508	-703.2	-0.0517	-1253.2*
	(0.595)	(633.6)	(0.0606)	(696.4)
Impact of Own Labor Market History on Treatment Effect	-0.267	-95.90	-0.0338	-213.1
	(0.249)	(255.2)	(0.0257)	(302.5)
Average Treatment Effect	0.367*	134.2	0.0441**	-80.32
	(0.197)	(176.1)	(0.0215)	(197.4)
Cohort Fixed Effects	X	X	X	X
Time Period Fixed Effects	X	X	X	X
Strata Fixed Effects	X	X	X	X
Sample	Post-Training	Post-Training	Post-Training	Post-Training
Number of Clusters	174	174	174	139
Observations	1632	1632	1632	1095

*** p<1%, ** p<5%, * p<10%. An observation is a individual-time period. The standard errors are clustered at the individual-level for the individuals not assigned to the treatment and the training cohort for the individuals assigned to the treatment. While labeled "log wage," we use an inverse hyperbolic sine function to allow for zero value in the outcome. All variables are measured as the "expected" outcome, with the expectations formed using a random forest regression with pre-randomization measures as covariates.

E.2 Appendix Figures

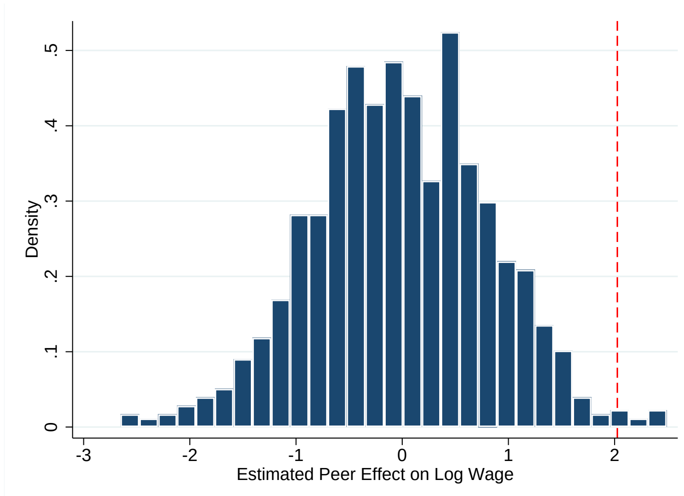
Figure A1: Exogeneity Test Results



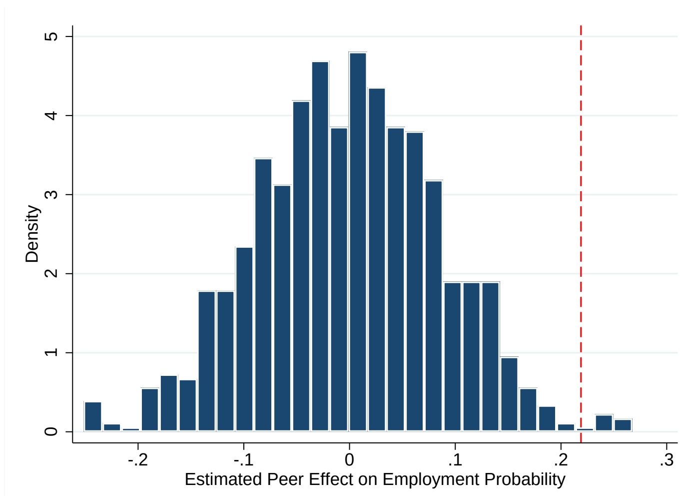
Note: The above figure shows the results of a simulation, in which we randomly assign individuals to cohorts and then compute the across-cohort variance in average labor market history. We repeat this simulation 1,000 times and plot a histogram of the results in the blue bars, with the red dotted line illustrating the across-cohort variance based on the actual cohorts that individuals are in. See Appendix D.1 for more details.

Figure A2: Estimated "Peer Effects" Under Alternative Assignments

(a) Log-Wage

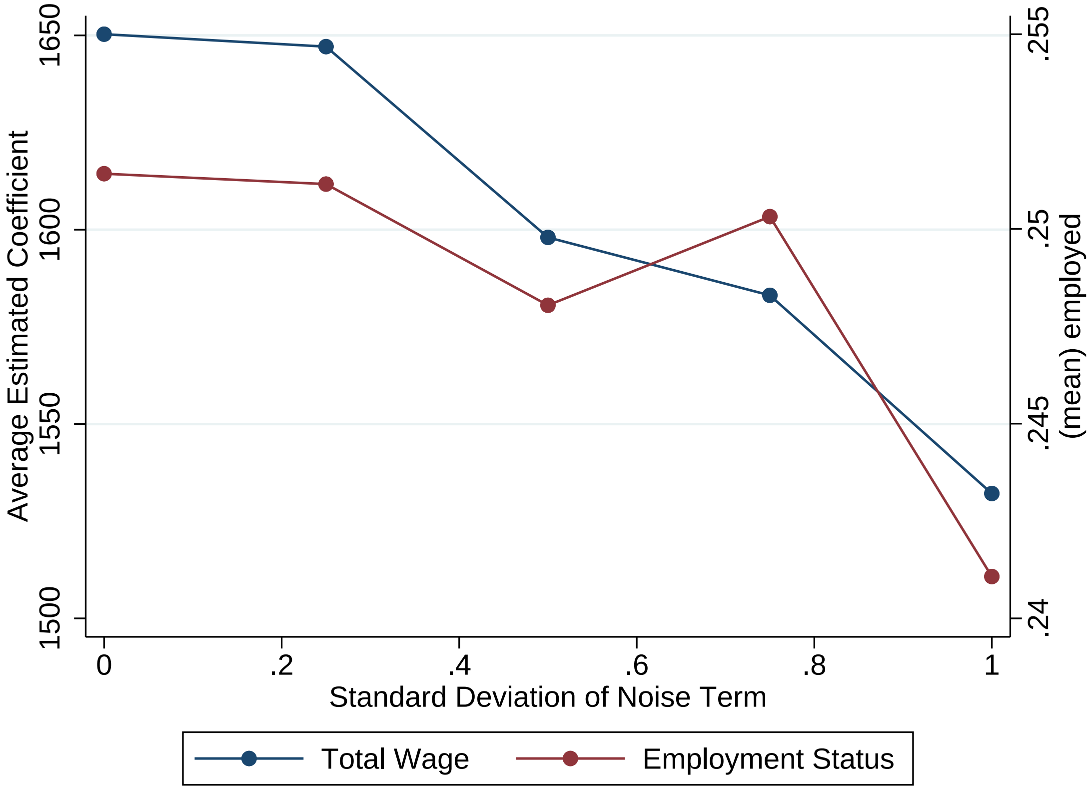


(b) Employment Probability



Note: These figures show the result of a simulation, in which we randomly assign individuals to cohorts/treatment, use those simulated groups to compute the individuals' peer quality, and then re-estimate the main specification using these as individuals' peer characteristics. We repeat this simulation 1,000 times and plot a histogram of the peer effect estimates using these fake peer assignments in the blue bars. The red dotted line illustrates the peer effect estimates using the true peer groups. See Apendix D.2 for more details.

Figure A3: Measurement Error Simulation



Note: The above figure shows the results of a simulation, where additional measurement error was added to each individuals' labor market history and then estimation was conducted in the same manner as is done in the main analysis. The x-axis shows the standard deviation of the added measurement error, while the y-axes show the estimated coefficients.