Do Dropouts Benefit from Training Programs? Korean Evidence Employing Methods for Continuous Treatments

Chung Choe
Alfonso Flores-Lagunes
Sang-Jun Lee

DISCUSSION PAPER SERIES

IZA

Do Dropouts Benefit from Training Programs? Korean Evidence Employing Methods for Continuous Treatments

Chung Choe
CEPS/INSTEAD and K.U.Leuven

Alfonso Flores-Lagunes
University of Florida and IZA

Sang-Jun Lee
Korea Research Institute for Vocational Education and Training

ABSTRACT

Do Dropouts Benefit from Training Programs?
Korean Evidence Employing Methods for Continuous Treatments[*]

Failure of participants to complete training programs is pervasive in existing active labor market programs both in developed and developing countries. The proportion of dropouts in prototypical programs ranges from 10 to 50 percent of all participants. From a policy perspective, it is of interest to know if dropouts benefit from the time they spend in training since these programs require considerable resources. We shed light on this issue by estimating the average employment effects of different lengths of exposure to a program by dropouts in a Korean job training program. To do this, we employ parametric and semiparametric methods to estimate effects from continuous treatments using the generalized propensity score, under the assumption that selection into different lengths of exposure is based on a rich set of observed covariates. We find that participants who drop out later – thereby having longer exposures – exhibit higher employment probabilities one year after receiving training, and that marginal effects of additional exposure to training are initially fairly small, but increase sharply past a certain threshold of exposure. One implication of these results is that this and similar programs could benefit from providing incentives for participants to stay longer in the program.

Corresponding author:

Alfonso Flores-Lagunes
Department of Economics
State University of New York at Binghamton
P.O. Box 6000
Binghamton, NY 13902-6000
USA
E-mail: aflores@binghamton.edu

1. Introduction

Several countries throughout the world have instituted active labor market programs, such as job training programs, to help unemployed workers succeed in the labor market. For the most part, developed countries introduced this type of programs first (for references on the evaluation of these programs see, among others, Heckman, LaLonde, and Smith, 1999; Card, Kluve, and Weber, 2010; and references therein). More recently, developing countries have introduced similar programs, sometimes with the aid from multilateral organization such as the World Bank or the Inter-American Development Bank. Some of the work evaluating programs in developing countries include Attanasio, Kugler, and Meguir (2009); Betcherman, Olivas, and Dar (2004); Calderón-Madrid (2006); Card et al. (2007); Ibarraran and Rosas Shady (2009); Kluve, Lehmann, and Schmidt (1999); Lee and Lee (2005); Luboya and van Ours (1999); and Rodriguez-Planas and Jacob (2009), among others. In general, no consensus exists about the success of job training programs in both developed and developing countries. Some studies suggest that job training programs for the unemployed provide benefits to their participants in developed countries (e.g., Card, Kluve, and Weber, 2010; Betcherman, Olivas, and Dar, 2004), with the evidence generally less clear for developing countries (Betcherman, Olivas, and Dar, 2004). Recently, however, some evaluations of training programs in Latin America have offered more positive prospects of their effectiveness (e.g., Ibarraran and Rosas Shady, 2009; Attanasio, Kugler, and Meguir, 2009).

A pervasive feature of job training programs is that participants fail to complete the program. For instance, Heckman et al. (2000) report that the dropout rate in five major evaluations of U.S. job training programs range from 5 percent to 79 percent.[2] In the context of developing countries, De Crombrugghe, Espinoza, and Heijke (2009) report that the dropout rate in a selection of countries ranges from 10 to 50 percent, with an average dropout rate of 30 percent of the individuals enrolling in the job training program.[3] In contrast, the typical evaluation of job training programs either excludes dropouts from the analysis (e.g., Aedo and Nunez, 2004; Lee and Lee, 2005; Card et al., 2007) or estimates the effect of the program on participants regardless of whether individuals complete the program (e.g., Ibarraran and Rosas Shady (2009); Kluve, Lehmann, and Schmidt (1999); Luboya and van Ours (1999); Rodriguez-

---

[2] We note, however, that their Table I reports the fraction of individuals assigned to the treatment group that received the intended services. By this definition, individuals that never enroll are counted as dropouts. Our definition will only consider individuals that do enroll but do not complete the program.

[3] The programs they consider include six programs in Latin American countries, one in Germany, and one in the United States.

Planas and Jacob (2009), among others).[4] Just a limited number of papers give explicit consideration to dropouts, such as Heckman, Smith, and Taber (1998); Heckman et al. (2000); and Waller (2008) in developed countries, and Lee and Lee (2003, 2009) in a developing country.[5] Given the considerable amount of financial resources devoted to these programs—both from the own country and often from multilateral organizations in the case of developing countries—it is of interest from a policy perspective to analyze if dropouts show any benefit from the time they spend in job training programs.

In principle, if participants in a training program accumulate human capital during the time they are exposed to the program, it stands to reason that longer exposures will lead to improved labor market outcomes. If this applies to both completers of the program and dropouts, then the expectation is that dropouts receiving different dosages (lengths) of training will differ in their labor market outcomes too.[6] This argument provides a rationale for the expectation that dropouts benefit from participation in a training program despite not completing it, based on the idea that the longer the length of enrollment, the better the expected labor market outcomes. Therefore, our aim is to provide evidence on this prospect by estimating causal effects to receiving different dosages of training among dropouts from a job training program in a developing country.

We shed light on this issue by estimating the causal effects of a prototypical job training program in Korea on the future employment of participants who have dropped out at different times, thereby having different lengths of exposure.[7] Our estimation employs recent parametric and semiparametric methods to estimate the causal effects from continuous treatments. These methods allow us to estimate average casual effects of an additional week of exposure to the program before dropping out, under the assumption that selection into different lengths of training is based on a rich set of observed covariates (i.e., unconfoundedness).

---

[4] Some of these studies explicitly state that they have included dropouts in their analysis, while others simply do not mention whether dropouts are included in the analysis or not.

[5] De Crombrugghe, Espinoza, and Heijke (2009) also focus on dropouts, but they analyze dropout behavior (e.g., who drops out) and do not estimate the effects of the program.

[6] As a simple illustration in the context of the Korean program analyzed here, the objective of the Webmaster (training) Program is to train participants as web developers. In this case, participants with longer training spells would obtain more knowledge and skills regarding web development. Relative to individuals with shorter training spells, those with longer spells accumulate more skills that should lead in principle to a higher employment probability (our outcome variable).

[7] An alternative outcome that is typically used to evaluate training effectiveness is a measure of earnings, which is not available in our data. Kluve et al. (2007) also use employment probability as an outcome in evaluating the effects of length of enrollment in a German training program.

Recent work on the estimation of treatment effects under the assumption of unconfoundedness has focused on multivalued and continuous treatments. In particular, work by Imbens (2000) and Hirano and Imbens (2004) have extended the concept of the propensity score for binary treatments (e.g., Rosenbaum and Rubin, 1983) to the case of multivalued and continuous treatments, respectively. This extension allows employing the similar concept of "generalized propensity score" (GPS) to control for selection bias and estimate a "dose-response function" (DRF) of time spent in training on the outcome of interest. The key identifying assumption is that, conditional on observed variables, there are no unobserved factors that simultaneously determine both the training duration and the outcome of interest. To increase the confidence in this identifying assumption, one can carry out indirect tests in the spirit of Heckman and Hotz (1989), which we do. The estimators we employ control for the GPS in a nonparametric way, as in Flores et al. (forthcoming), in contrast to other existing parametric approaches (e.g., Hirano and Imbens, 2004; Kluve et al., 2007; Mitnik, 2008; and Bia and Mattei, 2008). In this regard, we document that controlling for the GPS in a nonparametric way is helpful in identifying instances where the parametric method may not be appropriate.

Methods for continuous treatments have been increasingly applied to the analysis of active labor market programs (Kluve et al., 2007; Mitnik, 2008; Flores et al., forthcoming), including conditional cash transfer programs (Ibarraran and Villa, 2010; Campos-Vazquez and Chiapa, 2011). In our case, applying these methods to analyze the effects of a program on dropouts is natural given that participants in job training programs who drop out at different times receive varying dosages of the training. This is the case in the Korean job training program that we analyze here, in which once a participant chooses a specific training trade, the length of enrollment to complete the program is fixed,[8] while those participants who do not complete the program drop out at different times, thereby generating a rich amount of variability in the lengths of enrollment. In our sample of trainees that drop out (which represent 31.6% of all job training participants), the average participation time in the program is 12 weeks, with a standard deviation of 6.7 weeks and inter-decile range of 19.7 weeks, yielding a nearly continuous measure of

---

[8] The complete duration of the training in the different trades available within our training program ranges from 3 to 48 weeks.

training spells (see Figure 1).[9] Therefore, our analysis exploits the rich variability in lengths of training to analyze the effect of the program on dropouts.

Our results indicate that the DRF of training duration on the employment probability for dropouts is small and essentially flat up to about 12 to 15 weeks of training, after which it becomes an increasing function with a steep slope. In addition, estimates of the derivative of the DRF—which can be interpreted as the marginal effect of an additional week of training (before dropping out) on the employment probability—indicate that the marginal effect bounces between positive and negative up until about the 12th week of training, after which it is positive but small until about the 17th week and then it becomes large, positive, and statistically significant. When we analyze sub-samples by gender, we find that the effectiveness of job trainings is larger among men than among women, although the smaller sample sizes for these groups result in imprecise inference. The remaining of the paper is organized as follows. Section 2 provides a description of the job training program in Korea that we analyze, along with the administrative data available to us and their summary statistics. Section 3 describes the econometric methods we employ to estimate the DRF under an unconfoundedness assumption. Section 4 describes the estimation results and indirectly assesses the validity of our identifying assumption. The final section provides concluding remarks.

2. The Korean Job Training Program and the Data Employed

A. The Job Training Program in Korea

Job training programs have a long history in many developed countries such as the United States and Europe. However, they have been less prevalent in developing countries. In Korea, job training programs were almost non-existent before the financial crisis of 1997−1998 that increased unemployment rates from 2.6 percent in 1997 to 7.0 percent in 1998. This led to the first ever large-scale job training program in 1997, and since then job training has been the mainstay in the Korean government's strategy to cope with unemployment.

All job training programs are publicly financed by the Korean Ministry of Labor. In each geographical region there are local government offices (Employment Information Service) that

---

[9] Our data provide the exact calendar time of the start and termination (completion or dropout) of the job training program. Thus, we use as our treatment variable a measure of participants' enrollment in the program in days—which we rescale to weeks by dividing the total number of days spent in the program by five.

plan the program details, taking into account the labor market issues in their areas, while the budget for the program is allocated by the federal Ministry of Labor to the local offices. About 130 of the local Employment Information Service offices are spread out across the nation. These offices oversee both the local training centers—mostly private agencies—and the types of courses provided by them. The final decision on whether a particular training center is eligible to become publicly funded is made by the Ministry of Labor.

Not all of the unemployed are eligible for the job training program. Instead, only those unemployed involuntarily whose previous employers paid for the unemployment insurance premium can choose to receive the unemployment insurance benefit (if had at least six months of premiums paid) or participate in the job training program (if had at least two months of premiums paid).[10] Those who are interested in job training should apply for a training program of her/his own interest through an Employment Information Service office. During training, participants receive a monthly compensation based on the type of training enrolled in, which is typically a fraction (about one-half) of the unemployment insurance benefit and roughly 50 percent of the last workplace monthly wage (Lee and Lee, 2003, 2005). Since each individual trainee chooses both the training center and training type (thereby implicitly choosing the time it takes to complete the training), the local offices do not play a significant role in deciding the individual's training type or her length of duration of the training.[11] The Korea Research Institute for Vocational Education and Training is in charge of evaluating the effectiveness of the training program, both at the local and national levels on a regular basis. Additional information about the job training program in Korea can be found in Na et al. (2007).

Previous studies have evaluated the Korean job training program, reporting mixed evidence. For instance, Lee and Lee (2005) analyze females by employing administrative data similar to our data (detailed below) but for a smaller time period (January 1999 to March 2000). They apply several duration techniques and find that the job training program increases the unemployment duration of female participants. Lee and Lee (2009) using the same sample above, along with sensitivity analyses for matching estimators, find similar negative results for the employment probability of individuals at the end of March 2000. Conversely, Yoo and Kang

---

[10] Overall, about 90% of the unemployed in South Korea are eligible for unemployment benefits or training.
[11] Nevertheless, a local case worker typically provides information and discusses the options of types of training with the unemployed individual (Lee and Lee, 2009). Exogenous variation in this factor likely contributes to the identification of the DRF, as explained below.

(2010) employ data from Korea's Economically Active Population Survey along with fixed effects and propensity score matching techniques. Their findings show that the job training program increases participants' earnings by an amount between 2.6 to 9.8 percent (depending on the method employed).

B. The Administrative Data of Job Training in Korea

Our study uses matched administrative data from two sources provided by the Center for Employment Information in the Ministry of Labor in South Korea: the job-training file and the employment insurance file. Both files include only individuals with documented unemployment insurance premium payments. Thus, workers at firms/factories that never pay for unemployment insurance premiums are not covered.

Similar data are employed in Lee and Lee (2003, 2005, 2009). However, compared to the data used by them, our data has two main advantages. First, ours include both male and female trainees (theirs only include females) and a longer time period, from January 2003 to May 2007 (their sample covers those who became unemployed in 1999 and completed job training by the end of 1999). This more comprehensive data allow us to implement an analysis by gender and thus examine if the estimated effects are heterogeneous in this dimension. Second, unlike the data analyzed by Lee and Lee (2003, 2005), ours report the accurate dates when the participants dropped out in case they did not finish the training program. With this information, we are able to construct the precise length of enrollment in job-training for each individual, including the dropouts, which were excluded from the analysis in Lee and Lee (2005).

C. Summary Statistics of the Data Employed

Our sample includes all individuals who enrolled in job training from 2003 to 2007 and dropped out of it by May 2007. We include individuals who were born between 1960 and 1989, and have information on the outcomes and covariates of interest. We measure the duration of training employing items in the administrative data that provide the exact starting and dropout dates of the training, rescaling them into weeks by assuming a 5-day workweek. We discard observations with training duration less than 2 weeks (which corresponds to the 1st percentile of

durations) and also those with training durations above the 99th percentile as atypical.[12] After

dropping these observations, our sample consists of 5,803 individuals who dropped out of

training. We analyze the causal effect of this duration on the employment probability one year

after exit from the program. The information on employment status is included in the

unemployment insurance file which, in our data, follows each individual until May 31, 2008. In

order to have at least one year between training termination and when the outcome is observed,

those participants who drop out from training after May 31, 2007 are not considered in our

sample.

The list of covariates available to us for use in the estimation of the generalized

propensity score (GPS) is listed in Table 1, with the exception of 80 indicator variables for the

local training office attended. We broadly classify them into demographic, education,

characteristics of their last job (the size in persons of the last workplace, whether the workplace

was a big conglomerate or not, reasons for quitting the previous job, and industry-type of the

previous job), duration of their unemployment spell before enrolling in the job training program,

characteristics of the training assigned (training types, training level, unit cost of training taken,

planned duration to complete the program, and match between training and previous occupation),

indicators for local government offices, and indicators for the year in which the training was

taken. Lastly, information on several types of unemployment rates at the time of training

termination and for the two months before the dropout date are also employed in our analysis.

Table 1 presents summary statistics of the outcome variable and observed covariates for

our sample of dropouts and the subsamples of female and male dropouts. Unconditional mean

statistics of the employment probability reveal that on average 78 percent of all dropouts are

employed one year after dropping out of job training, and similarly for the female and male

subsamples. Job training in Korea is generally a full-time program in which there is variability in

the time it takes to complete different types of training, which also varies by local office. The

duration of training for those who complete their corresponding program, in weeks, ranges from

3 to 48 weeks. The variable "planned weeks of training", which represents the time needed to

complete the actual training program the individual enrolled in, indicates that it is necessary for

the average dropout in our sample to enroll for 22.3 weeks to complete the (average) program.

---

[12] For instance, while the maximum duration in the sample is 44.7 weeks, after the removal of individuals with these
durations the maximum is only 26 weeks.

However, dropouts in the sample enroll only for an average of 12 weeks, with males enrolling slightly longer (12.1 weeks) than females (11.9 weeks). Thus, irrespective of gender, dropouts enroll in training, on average, for about only half of the period that is necessary to complete the program successfully. In our sample, 45 percent of the dropouts are male, and about 50 percent hold at least a college degree.

We employ several different types of unemployment rates to account for the labor market conditions that may influence a participant's decision to drop out of the job training program. Indeed, finding a new job seems to be an important reason why trainees drop out of training, as reported by Lee and Lee (2009, p. 91) based on conversations with case workers. Our unemployment variables were constructed by the Korean Statistical Information Service from the Economically Active Population Survey.[13] For our study, we use three monthly measures: 1) unemployment rates by gender and age group; 2) unemployment rates by gender and education; and 3) unemployment rates by gender and city where training was received (local unemployment rate). In particular, the local unemployment rate variable was constructed by matching city-level unemployment rates for the specific location of the training centers attended by each individual in our sample. Additionally, we match the timing of these unemployment rates to the individual's month of exit (dropout) from job training, as well as the two months previous to exiting the program. The inclusion of this rich and detailed set of unemployment variables allows us to control for the labor market conditions faced by the individual dropping out of training. Thus, in this respect, our identifying assumption implies that the arrival of a job offer will not be related simultaneously to training length and employment probability once we condition on the various unemployment rate measures and all other covariates. The summary statistics of the set of unemployment rates we employ are presented in Table 1.

Other variables that serve as important controls for confounding factors in the GPS specification are the reported individual reasons for becoming unemployed (self-employed, marriage, had child, injured, layoff, personal, workplace closed, work contract expiration, etc.). They help control for potential individual unobserved characteristics that could be related to both the future employment probability and the length of enrollment in job training. Similarly, the characteristics of the firm in which the previous job was held are important controls, such as the

---

[13] Economically Active Population Survey is a typical monthly-basis survey similar to the Current Population Survey in the United States.

size. In this respect, the average firm size is 90 employees (Ex-firm Size) and about 11 percent of the dropouts worked for a big conglomerate (Ex-firm Type: big firm) before they became unemployed and enrolled in training. Another characteristic of the job held previous to training enrollment is the industry-type of employment. From Table 1, it appears that there is a gender difference in the distribution of industry category of the previous job held by the dropouts: while 45 percent of males worked in manufacturing sector, only 28 percent of females engaged in that industry. Instead, the relative ratio of women in education or the public health sector is much higher than that of men. There is evidence that controlling for variables related to the previous job held by individuals entering training is important to remove selection bias under the unconfoundedness assumption in the context of binary treatment effect estimators (Lechner and Wunsch, 2011).

We also control for several characteristics of the type of training in which the individual enrolled. These variables help control for the selection of different lengths of enrollment before dropping out since they measure aspects of the quality of the match between the unobserved individual preference for a given training type and the actual training undertaken. A difference between the individual preference and the actual training undertaken could arise due to limited availability of the preferred training or by recommendation of a training official. The first set of these variables measure the match between the individual's background and the training type (same, similar or different). Not surprisingly, most individuals (67%) receive training that is different than their current background. A second aspect is the training level (basic, intermediate, or advanced), in which about half of the training participants enroll in the intermediate level. Finally, trainings are classified into fifteen different types, with the majority of participants enrolling in a few of them, such as machine/equipment, information/communication, service, and clerical. It can be seen from Table 1 that male trainees are more likely to participate in machine/equipment training, whereas women trainees show high participation rates in clerical and service.

Lastly, we enrich the specification of the GPS with two sets of fixed effects that account for unobserved confounding factors at different levels. The first set pertains to indicators for the local government offices (Korea Employment Information Service; not shown in Table 1) in which the individual undertakes the training. They are employed to control for time-invariant unobserved factors that are specific to the training office and/or additional local labor market

11

unobserved factors that may simultaneously influence the individual's length of enrollment and the employment probability. The second set is year fixed effects that are included to account for year-specific unobserved confounding factors that may simultaneously influence length of enrollment and employment probability but are common across all local government offices.

D. Discussion of the Plausibility of the Unconfoundedness Assumption

As mentioned before, our identification strategy employs the unconfoundedness or selection-on-observables assumption. This assumption states that, conditional on the rich set of observable covariates described before, there are no unobserved factors (known as confounders) that simultaneously influence both the training length before dropping out and the employment probability one year after dropping out. This identification assumption is inherently untestable, although we offer indirect evidence of its validity using "placebo tests" in section 4.C below. In this subsection we discuss some factors that we argue increase the plausibility of the unconfoundedness assumption in the context of our application, as well as the potential sources of exogenous variability that we exploit in identifying the DRF and its derivative.

The traditional source of bias in studies employing the unconfoundedness assumption relates to potential confounders that are static (e.g., demographic characteristics, unobserved ability, and self-discipline, among others). We believe that the rich set of control variables available to us adequately control for this source of bias. A second important source of bias when employing methods for continuous treatments to analyze the causal effect of a duration variable (e.g., training length) is the presence of potential "dynamic confounders" (Heckman and Navarro, 2007; Flores et al., forthcoming). These potential confounders can arise due to factors whose realization occur after the start of training and are unobserved to the analyst. A relevant example of these factors are individual performance measures such as tests within the training program that can discourage the individual from continuing training and also affect her future employment probability. In this regard, the institutional characteristics of the training program in Korea help us reduce the importance of a number of these potential dynamic confounders. Specifically, there are no formal tests within the program that determine continuation, likely removing a significant source of dynamic confounders. Thus, in sum, a combination of the richness of our data and the institutional details of the training program increases our confidence that the unconfoundedness assumption is satisfied in our application, which will be indirectly assessed in section 4.C.

Finally, for our identification and estimation strategy to work well, we need to have sources of exogenous variation, that is, factors that are related to the variation in training duration before dropping out but not to future employment variability, conditional on the rich set of control variables. These factors can be thought of as "unobserved instruments" (Busso, DiNardo, and McCrary, 2008). Based on the institutional arrangement of the training program we analyze and the details of the data, there are a number of factors that can produce this exogenous variability. One of them is the availability of particular types of training at the specific time the individual enters the program, which could cause a mismatch between an individual's preferences and the type of training actually undertaken, leading to variability in training duration. Another is the potential influence that case workers can exert in the training choice (and thus duration) through the information and conversations they have with individuals. Similarly, personality mismatches between participants and instructors in the program (conditional on observed covariates) may also lead to exogenous variation in training duration before dropping out. As a final example, the random (conditional on our unemployment rate measures and other covariates) arrival of job offers is another potential factor that could be related to training duration but not to the employment probability one year after training termination.

3. Econometric Methods

We index the individuals in our sample by $i = 1, ..., N$ and denote by $Y_i(t)$ the potential outcome of individual i under treatment level $t \in \Im$, where $\Im$ is an interval and t denotes the duration of training before dropping out in our application. We aim to estimate the average dose-response function (DRF) denoted by $\mu(t) = E[Y_i(t) \mid \Psi]$ where Ψ denotes the set of individuals that enroll in training and drop out before completing it. For notational simplicity, we leave implicit the conditioning on Ψ in what follows. The observed variables for each individual i are a vector of covariates X_i (those described in the previous section), the level of the treatment received T_i, and the observed outcome for the level of the treatment actually received $Y_i = Y_i(T_i)$.

Our key identifying assumption to estimate the DRF is that selection into different treatment levels is *weakly* unconfounded given the covariates, which was introduced in Hirano

and Imbens (2004) for the case of a continuous treatment:[14]

$$Y_i(t) \perp T_i \mid X_i \text{ for all } t \in \mathfrak{I}. \qquad (1)$$

This assumption states that the level of the treatment received (T_i) is independent of the potential

outcome $Y_i(t)$ conditional on observed covariates, and it is a natural extension of the commonly

used unconfoundedness assumption in the binary-treatment literature (e.g., Heckman, LaLonde

and Smith, 1999; Imbens, 2004). Importantly, this assumption rules out any systematic

"selection" into levels of the treatment based on unobservable characteristics (the confounders

described in the previous section) not captured by observable ones. Under this assumption, the

average DRF can be derived by estimating average outcomes in subpopulations defined by the

covariates and different levels of the treatment. However, as the number of covariates increases,

it becomes difficult to simultaneously adjust for all covariates contained in X.

In the case of a binary treatment, the propensity score—defined as the conditional

probability of receiving treatment given the covariates—is commonly used to estimate average

treatment effects under unconfoundedness. Rosenbaum and Rubin (1983) show that adjusting the

propensity score eliminates selection bias between treated and untreated individuals if selection

into treatment is based on observable factors. This result implies that it is enough to control a

scalar variable as opposed to adjusting for all covariates, leading to more flexible ways to

estimate treatment effects. Another advantage of propensity score methods is that they allow

detecting and deleting observations in the treatment and control groups for which it is not

possible to find comparable units in the opposite group, thus avoiding a source of bias (e.g.,

Imbens, 2004).

Hirano and Imbens (2004) extend the propensity score methodology to continuous

treatments employing the GPS to reduce the conditioning set to one, just as in the binary case.

The GPS is the conditional density of the treatment given the covariates:

$$r(t,x) = f_{T|X}(t \mid X = x). \qquad (2)$$

Let $R_i = r(T_i, X_i)$ denote the conditional density at the treatment actually received, and let

$R_i^t = r(t, X_i)$ denote the family of random variables indexed by t. Clearly, for those units with

$T_i = t$ we have $R_i = R_i^t$. Hirano and Imbens (2004) show that, as in the binary case, the GPS has

---

[14] They refer to this assumption as weak unconfoundedness since it does not require joint independence of all
potential outcomes but instead requires conditional independence to hold for each value of the treatment.

a "balancing property" in that, loosely speaking, $X \perp 1\{T = t\} \mid r(t, X)$, and that weak

unconfoundedness given the covariates implies weak unconfoundedness given the GPS, i.e.,

$f_T(t \mid R_i^t, Y_i(t)) = f_T(t \mid R_i^t)$. This last result allows the estimation of the average DRF by using the

GPS to remove selection bias. In particular, they show that under the assumption in (1) we can

identify the average DRF as

$$(i) \quad \beta(t, r) = \mathrm{E}[Y_i(t) \mid R_i^t = r] = \mathrm{E}[Y_i \mid T_i = t, R_i = r]$$

$$(ii) \quad \mu(t) = \mathrm{E}[\beta(t, R_i^t)]. \tag{3}$$

This result suggests estimating the DRF at t using a partial mean, which is an average of a

regression function over some of its regressors while holding others fixed (Newey, 1994). In this

case the regressor that is fixed in the second averaging is the treatment level t. Hence, the DRF

function can be estimated using the GPS by following two steps. The first involves estimation of

the conditional expectation of Y on T and R, $E[Y_i \mid T_i = t, R_i = r]$, while the second is to average

this conditional expectation over R_i^t to get the value of the DRF at t.

This estimation procedure can be undertaken in a number of different ways. For instance,

Hirano and Imbens (2004), Kluve et al. (2007), and Bia and Mattei (2008), among others,

implement this procedure by assuming a flexible parametric form for the regression function of Y

on T and R by estimating the regression:

$$E[Y_i \mid T_i, \hat{R}_i] = \alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 T_i^3 + \alpha_4 \hat{R}_i + \alpha_5 \hat{R}_i^2 + \alpha_6 \hat{R}_i^3 + \alpha_7 T_i \cdot \hat{R}_i \tag{4}$$

where $\hat{R}_i$ is an estimator of R_i (to be discussed in the following section). Then, the DRF at t is

estimated by averaging (4) over the distribution of $\hat{R}_i^t$, which is an estimator of R_i^t:

$$\hat{\mu}(t)_{PPM} = \frac{1}{N} \sum_{i=1}^{N} [\hat{\alpha}_0 + \hat{\alpha}_1 t + \hat{\alpha}_2 t^2 + \hat{\alpha}_3 t^3 + \hat{\alpha}_4 \hat{R}_i^t + \hat{\alpha}_5 \hat{R}_i^{t2} + \hat{\alpha}_6 \hat{R}_i^{t3} + \hat{\alpha}_7 t \cdot \hat{R}_i^t].$$

We will obtain results employing this estimator, which we refer to as the parametric partial mean

(PPM) estimator.

A more flexible way to implement the methodology is by modeling the conditional

expectation $\beta(t, r)$ nonparametrically, since there is no reason to commit ex-ante to any

particular functional form and a misspecification could result in inappropriate bias removal. Here,

we specifically employ Flores et al.'s (forthcoming) semiparametric estimator based on inverse

weighting by the GPS. This estimator is derived from a result in Flores (2005) showing that,

under the assumption in (1), the DRF at t can be identified as:

$$\mu(t) = E\left[\frac{\omega(T_i, X_i; t)Y_i}{E[\omega(T_i, X_i; t) \mid X_i]}\right] \qquad (5)$$

where $\omega(T, X; t)$ is a function of the treatment and the covariates such that $E[\omega(T, X; t) \mid X]$ exists and is different from zero. The DRF can be estimated by setting $\omega(T_i, X_i; t) = 1(T_i = t)$ and using nonparametric methods to deal with the continuous nature of the treatment. Additionally, the weights for the estimator are normalized to add up to one, as it is commonly done in the binary case (Imbens, 2004). The general form of the semiparametric inverse weighting (IW) estimator of the DRF at t is:

$$\hat{\mu}(t) = \frac{\sum_{i=1}^{N} \tilde{K}_{h,X}(T_i - t) \cdot Y_i}{\sum_{i=1}^{N} \tilde{K}_{h,X}(T_i - t)}$$

where $\tilde{K}_{h,X}(T_i - t) = K_h(T_i - t) / \hat{R}_i^t$ and K_h is a kernel function with bandwidth h. This IW estimator is simply the usual local constant regression (or Nadaraya-Watson) estimator where each individual's kernel weight is divided by her GPS at t.

Our specific implementation of the IW estimator uses a local linear regression of Y on T and the weighted kernel function $\tilde{K}_{h,X}(T_i - t)$ in order to avoid boundary bias and facilitate the computation of estimated derivatives. Let $S_j(t) = \sum_{i=1}^{N} \tilde{K}_{h,X}(T_i - t)(T_i - t)^j$ and

$D_j(t) = \sum_{i=1}^{N} \tilde{K}_{h,X}(T_i - t)(T_i - t)^j Y_i$. Then, the IW estimator employed can be written as:

$$\hat{\mu}(t)_{IW} = \frac{D_0(t)S_2(t) - D_1(t)S_1(t)}{S_0(t)S_2(t) - S_1^2(t)}. \qquad (6)$$

We employ a normal kernel and choose a global bandwidth based on the procedure proposed by Fan and Gijbels (1996), which is based on estimating the unknown terms appearing in the optimal global bandwidth by employing a global polynomial of order p plus 3, with p being the order of the local polynomial fitted.[15]

---

[15] This bandwidth selector has been previously used in economics (e.g., Ichimura and Todd, 2007), especially an adaptation of it to the regression discontinuity context (e.g., Lee and Lemieux, 2010).

4. Estimation Results

A. Estimation of the GPS

The first step to implement the estimators from the previous section consists of modeling the conditional distribution of the training duration before dropping out (T_i) given the covariates, that is, the GPS. Instead of committing ex-ante to any one specification for the GPS, we follow Flores et al. (forthcoming) and estimate a number of flexible generalized linear models (McCullagh and Nelder, 1989) and choose the model that best fits our data. Let $g\{E(T)\} = X\gamma$, with T the continuous treatment variable, X the covariates, γ a vector of coefficients, and g a "link function". Various specifications can be obtained by choosing a distribution F for T and a functional form for g. For example, a log-normal specification is obtained as a special case with F as normal and g as the log function. We estimate several plausible specifications by maximum likelihood and choose a model that fits the data best according to the deviance measure of McCullagh and Nelder (1989), the Akaike Information Criteria (AIC) and the value of the log-likelihood function (the models estimated, along with these goodness of fit measures, are presented in the Internet Appendix).[16]

The variables included in the generalized linear models are all those listed in Table 1 (including the indicators for local training office attended), along with a full set of interactions of each variable with indicator variables for gender, interactions of education categories with the local unemployment rate variable, and higher order polynomials (up to cube) of several continuous variables (age, planned training spell, and unit cost of training). For all three samples, we employ a gamma model with a log link to model the GPS.[17] All estimated coefficients of the GPS model for each sample, along with their respective robust standard errors, are contained in the Internet Appendix. We briefly mention here some of the estimated coefficients of the GPS

---

[16] The distributions considered were the log-normal, inverse Gaussian, and gamma distributions. Within the inverse Gaussian and gamma distributions, we employed link functions corresponding to the identity; inverse powers 1, 1.5, 2; and a log link. We employ AIC to choose across the different distributions and the other two measures to choose among link functions (e.g., Hardin and Hilbe, 2007).

[17] According to the AIC measure, the log-normal model appears to be best among those considered. However, the GPS estimated with this model has trouble satisfying the balancing property, and thus was discarded in favor of models with a Gamma distribution. Among the models with a Gamma distribution, those with a log link and an inverse power 1 link have a very similar deviance measure and value of the log likelihood function, and both satisfy the balancing property. Since all results employing these two links are almost identical we decided to report the results employing the log link. Note that a gamma model with log link and scale parameter equal to one is equivalent to an exponential regression model, commonly used in duration analysis. However, our GLM model does not restrict the scale parameter to one, and thus it is more general.

model for the entire dropout sample.

The planned duration of the particular type of training chosen shows a positive association with the actual length of training before dropping out, and has a strong predictive power. Other covariates such as training cost, length of unemployment spell, ex-firm size, and advanced training level are all positively associated with training duration before dropping out, although some of these factors are statistically insignificant. The indicators for the reasons to quit the previous job, training type, and training centers play a significant role in explaining the enrollment length before dropping out. In general, significant differences are found in the coefficients across genders, especially in those for some education categories, reasons to quit the previous job, training type, and training centers. Lastly, most of the estimated coefficients of the unemployment variables have negative signs, implying that participants tend to shorten the length of training before dropping out when unemployment rates are relatively high. This sign is consistent with the notion that the value of finding a job during a high unemployment rate period is high relative to a period with low unemployment rates, thus providing trainees with an additional incentive to drop out of training to take that job.

Given that the GPS is employed to make comparisons of individuals with different values of T but the same values of the GPS, it is important to verify that no values of the GPS are so extreme that individuals with comparable values of the GPS are impossible to find. For these extreme values, inference using the GPS will be poor, resulting in a source of bias. Therefore, we follow common practice (e.g., Dehejia and Wahba, 2002; Imbens, 2004; Gerfin and Lechner, 2002; Lechner, 2002; Flores and Mitnik, 2008) and restrict the analysis to those individuals for which reliable inference can be obtained by concentrating the analysis on the subset of individuals for which the common overlap support condition is satisfied.

To measure the extent of overlap in the support of different levels of the treatment, we utilize a straightforward extension of a method employed in the binary-treatment case (e.g., Dehejia and Wahba, 2002) which was suggested in Flores et al. (forthcoming). Let the quintile each individual belongs to be denoted by $Q_i = \{1, 2, 3, 4, 5\}$. For each quintile q, we compute the value of the GPS at the median level of the treatment in that quintile for all individuals, call it $\hat{R}_i^q$. The common-support region with respect to quintile q is obtained by comparing the support of the distribution of $\hat{R}_i^q$ for those individuals with $Q_i = q$ to that of individuals with $Q_i \neq q$. Let

18

CS_q denote the common-support subsample with respect to quintile q. Then, we define CS_q as:

$$CS_q = \left\{ i : \hat{R}_i^q \in [\max\{\min_{\{j:Q_j=q\}} \hat{R}_j^q, \min_{\{j:Q_j \neq q\}} \hat{R}_j^q\}, \min\{\max_{\{j:Q_j=q\}} \hat{R}_j^q, \max_{\{j:Q_j \neq q\}} \hat{R}_j^q\}] \right\}.$$

We restrict our sample (for each group analyzed) by keeping only those individuals that are comparable across all five quintiles simultaneously. Hence, our common-support subsample is given by: $CS = \bigcap_{q=1}^{5} CS_q$.

The resulting common-support restricted samples show that a non-trivial number of observations are dropped from the original samples in Table 1, which reflects the difficulty of finding comparable individuals in terms of the GPS within each of the samples. The observations dropped due to common support restrictions are (percentage of observations dropped in parentheses) 1,283 observations (22.1%) for the entire dropout sample, 854 observations (26.8%) for female dropouts, and 505 observations (19.3%) for male dropouts. Thus, not surprisingly, there appears to be more heterogeneity in terms of the estimated GPS for females than for males. Given that using observations outside the common support of the GPS can result in misleading inference, we concentrate on the samples that exclude these observations for the rest of the analysis.

An important characteristic of the estimated GPS that needs to be verified is its balancing property: the GPS "balances" the covariates within strata defined by the values of the GPS, such that, within strata, the "probability" that $t = T$ does not depend on the value of X. This balancing property can be employed to empirically assess the adequacy of the estimated GPS in a similar way in which it is done in the binary treatment case with the propensity score (e.g., Dehejia and Wahba, 2002; Smith and Todd, 2005). The approach we follow here for a continuous treatment consists of estimating the same GLM model (gamma with a log link) for T as in the GPS that includes all covariates plus the GPS up to a cubic term (the unrestricted model).[18] Then, a likelihood ratio (LR) test is employed to compare that model with a restricted one that sets the coefficients of all covariates to zero. If the GPS sufficiently balances the covariates employed, then they should have little explanatory power conditional on the GPS. We find this to be the case in Table 2 as the LR tests strongly indicate that the restricted model is overwhelmingly preferred in all samples (top panel). Conversely, using the same unrestricted model, we test whether the GPS coefficients are all simultaneously equal to zero. This restriction is strongly

---

[18] We use a cubic specification of the GPS to make it consistent with the specification of the PPM estimator in (4).

rejected in the bottom panel of Table 2, speaking to the importance of the role played by the GPS.[19] We regard this as evidence that the balancing property of our estimated GPS is satisfied.

B. Estimation of the Dose-Response Function and its Derivative

We compute results employing the PPM and IW estimators, as well as the ordinary least squares (OLS) estimator for comparison. The DRF estimates are obtained at 99 evenly-divided values of the length of training for each of the samples analyzed. We generate a series of figures that plot the DRF of length of enrollment in job training (in weeks) on the employment probability one year after dropping out of the program. This outcome measure fixes at one year the amount of time that elapses between the end of the training spell and when the employment probability is measured. We also generate plots of the derivative of the DRF, which represents the "marginal" return to additional time spent in training.[20] Our results are accompanied by 95% (point-wise, percentile-based) confidence bands obtained with 1,000 bootstrap replications that account for all estimation steps, including the estimation of the GPS and the imposition of the common support condition.

To conserve space, we present here a selection of plots that provide the main insights of our analysis, while the full set of plots can be found in the Internet Appendix to the paper. We start with plots of the DRF and estimated derivatives of the DRF using the OLS, PPM, and IW estimators considered for the entire sample of dropouts (Figure 2). We do not include confidence bands in this figure for readability. Figure 2(a) gives a general idea of the shape of the DRF using each of the estimators and allows us to point out that there are differences among the alternative estimators for the DRF in this application, which likely arise due to the differences in flexibility of each. The main differences occur for low values of the length of training (weeks 2 to 7) and more markedly after around the 14th week of training. Looking at the IW estimator (the most flexible one), the DRF is fairly flat until around 14 weeks in training, after which it shows an upward trend. As for the estimated derivatives of the DRF in Figure 2(b), there are also evident differences among the three estimators (again, likely arising due to their relative flexibility). In

---

[19] All these models are estimated with the common-support restricted sample.

[20] For the PPM estimator the derivative at t is obtained as the "forward" change of one additional week of training: $\hat{\mu}(t+1) - \hat{\mu}(t)$. This is the usual approach when using the PPM estimator (e.g., Bia and Mattei, 2008). For the IW estimator, the derivative estimate at t equals the slope coefficient of the linear term from a local quadratic regression of Y on T using the re-weighted kernel defined in section 3, $\tilde{K}_h(T_i, X_i; t)$. We choose the appropriate bandwidth by using the procedure described in Fan and Gijbels (1996).

general, the figure shows that the marginal returns to additional training before dropping out initially decline with the length of training before dropping out, then they hover around zero until after about the 17th week of training when they increase sharply (except OLS). For the subsequent analysis, we concentrate on the more flexible PPM and IW estimators.

Figure 3 presents plots of the estimated DRFs employing both the PPM (top panel) and IW (bottom panel) estimator for the three different samples analyzed, along with 95 percent confidence bands. In general, for all three samples, the employment probability one year after dropping out from the program is relatively constant up to a certain threshold, becoming then a sharp increasing function of the length of time spent in training. While we do not observe sharp differences between the parametric (PPM) and semiparametric (IW) estimates, overall, the DRFs using the IW estimator are a bit wigglier, especially for females. This could be the result of the likely greater heterogeneity within females with respect to the observed covariates, which is better accounted for by the semiparametric estimator. Consistent with Figure 2, the entire dropout sample shows initially a rather flat trend, but the employment probability after 1 year post-training sharply increases after around 17 weeks of training. A very similar pattern holds for both the female and male dropout sample, with the exception that for males, the upswing starts earlier (around the 14th week). One implication of these DRFs is that dropouts obtain considerably positive returns only if they remain in the program for a certain amount of time (over 14−17 weeks).

Figure 4 presents, for each of the samples, plots of the estimated derivative of the DRF calculated using the PPM (top panel) and the IW (bottom panel) estimators, along with 95% confidence bands. The interpretation of these derivatives corresponds to the marginal effect of an additional week of training on the employment probability one year after dropping out of the program. Several features in the figures are worth pointing out. First, note that the two estimators are fairly similar in the entire dropout sample, while they show more differences in the female and male subsamples (as was the case with the estimated DRF in Figure 3). This is likely due to a combination of the smaller number of observations in the two subsamples and the larger heterogeneity in observed covariates within the female subsample, as mentioned before. We continue to concentrate our discussion on the more flexible IW estimator in what follows. Second, there are substantial ranges in all of the samples where the estimated derivatives are positive. Consistent with the estimated DRFs, these estimated positive derivatives occur more

21

consistently after a certain threshold of weeks of training. The range of weeks of training where positive marginal effects are estimated are 2−8 and 12−25 for the entire dropout sample, 6.5−7.5, 9−13 and 17.5−24.5 for females, and 2−5.5 and 13.5−26 for males. Third, despite the evidence that most estimated derivatives are positive, especially for the entire sample and for males, most of these point estimates are not statistically different from zero. The weeks where positive and statistically significant estimated derivatives occur are 20−24 for the entire sample, 19−22 for females, and 22−22.5 for males. Importantly, however, in none of the samples are there estimated marginal effects that are negative and statistically significant (except for a half-week range at 10.5−11 for males). The general lack of precision of these estimated marginal effects is likely due to the fine grid of treatment levels we employed in their estimation (99 values of the range of training durations).

An alternative way to summarize the derivative estimates of the DRF in Figure 4 is to consider the average derivative over different ranges of treatment levels. We present some of these average derivatives (along with indicators of their statistical significance) for selected ranges of the treatment in Table 3. The ranges, for each sample, are the number of weeks in training that corresponds to the intervals 1−99, 1−50, 25−75, and 50−99 of the evenly-spaced levels of training in which the original range of lengths of training was divided to compute the DRF. The corresponding actual weeks of training are indicated in the table for each of the samples. Hence, for instance, the average derivative for the entire dropout sample between the 50th and 99th levels of training—that correspond to 13.71 to 25.31 weeks of training—is a highly statistically significant 0.0126. This estimate implies that, within this range of training lengths, an additional week of training for dropouts is causally related to an increase of 1.26 percentage points in the employment probability one year after dropping out of the program. Other statistically significant (at the 90 percent or better) average derivatives are the full range of training lengths (1−99) for the entire sample of dropouts with a point estimate of 0.008, the 1−99 range for the male sample with a point estimate of 0.0084, and the 50−99 range for the male sample with a point estimate of 0.0152. None of the estimated average derivatives is statistically significant for females.

It is also worth noting some additional features of the estimated average derivatives reported in Table 3. First, the differences between the estimated marginal returns to training for males and females are considerable. For instance, the estimated average derivative for the entire

range of training durations (1−99) for males is three times higher than that for females, while the average derivative for the range 50−99 for males is about fifty percent higher than that for females. Second, although we find two instances of a negative estimated average derivative (for females in the 1−50 range and males in the 25−75 range), these are very small and statistically insignificant. Thus, the evidence provided by the average derivatives is firmly towards positive marginal effects from longer training durations before dropping out. Third, the size of the estimated average marginal returns from additional weeks in training is much larger in the second half of the training period (after week 13, or the 50−99 range), which is consistent with the previous results presented in Figures 2−4.

C. Placebo Tests Results

In order to gauge the adequacy of the identifying assumption underlying our methodology, we perform placebo tests in the spirit of Heckman and Hotz (1989) considering outcomes that cannot possibly be affected by the treatment (training duration before dropping out) unless selection bias is inherent in our methodology. We consider four variables measured before the start of training for this exercise. The first two are individual employment status prior to randomization: (a) employment one year prior to training and (b) employment six months prior to training. Additionally, we consider (c) the type of previous employer—whether the previous employer is a big firm. Finally, we consider (d) a variable that measures the duration of the last unemployment spell prior to enrolling in training.

While placebo tests are commonly employed in studies estimating the effect of a binary treatment, their interpretation has to be clarified in our framework of a continuous treatment represented by training duration. In their traditional application, these tests are informative about unobserved confounders that are related to both the outcome of interest and the placebo outcome. As discussed in section 2.D, our identifying assumption can also be violated due to the presence of dynamic confounders. Since these confounders are revealed only after the start of treatment and our placebo outcomes are measured before treatment, the type of placebo tests we employ can only provide indirect evidence on the presence of dynamic confounders to the extent that they are correlated to both pre-treatment uncontrolled factors and the placebo outcomes. As a result, these tests are less informative in our setting than in the traditional one. Nevertheless, in both

23

settings, the presence of statistically significant estimated effects on the placebo outcome represent strong evidence against the validity of the identifying assumption.

Figure 5 plots the estimated derivative of the DRF for the different samples (similar to those in Figure 4) employing the first placebo outcome in employment one year prior to enrolling in training.[21] Similar figures for the other three placebo outcomes are provided in the Internet Appendix. Compared to Figure 4, the effects are of a smaller magnitude and they are generally not statistically significant except in a small range of treatment values in the sample of females (9−9.7 and 12.3−13.8 weeks using the IW estimator). Using any of the other placebo outcomes as well as employing the PPM estimator yields essentially the same conclusions: only very few estimated derivatives are statistically significant (positive or negative). As before, looking at estimated average derivatives nicely summarizes these results. These are shown in Table 4 for all placebo outcomes and samples, where none of the average derivatives is statistically different from zero at conventional significance levels for any of the ranges of the treatment considered. Additionally, the effects on the two employment probabilities before the treatment are very small relative to those on the actual outcome. We conclude that the placebo tests do not provide evidence against our identifying assumption.

D. Implications for Policy

The results presented in this section have implications for policy. They show that dropouts from this Korean job training program do benefit from longer training spells in terms of their employment probability one year after dropping out, but only after a certain training duration threshold. Spending an additional week of training does not lead to higher employment probability for short training durations. However, after spending about 12−15 weeks in the program, the marginal effect of an additional week has a positive effect on the employment probability, becoming statistically significant after about the 18th−22th week (depending on the sample). In other words, re-employment benefits of the program for participants who drop out require a minimum amount of time spent in training. Thus, policy makers should encourage trainees to stay in the program for as long as they can—ideally completing the program—but definitely more than 12−15 weeks, which corresponds to about half of the average duration

---

[21] To obtain these estimates the GPS is recomputed excluding the placebo outcome being analyzed from its specification to avoid controlling directly for the outcome variable in the estimated GPS.

needed to complete the (average) training program. This could be accomplished, for example, by providing incentives for enrollees to commit to the program for at least that amount of time.

5. Conclusion

We provide estimates of the causal effects of the length of time spent in training on the probability of employment one year after training termination for a sample of dropouts who enrolled in a Korean job training program. Estimates of the effects of job training programs on participants who drop out short of completion are important given the documented pervasiveness of high dropout rates in prototypical training programs throughout the world and the fact that these programs receive a considerable amount of resources from individual countries and multinational organizations. Our main findings suggest that dropouts do benefit from longer training spells compared to shorter ones, but only after a certain duration threshold. This information is important for policymakers and stakeholders of job training programs as they assess the effectiveness of such programs and design new ones.

To obtain our estimates of the dose-response function (DRF) of length of training on the employment probability, we employ recently developed methods for the estimation of causal effects from continuous treatments. The key identifying assumption is that, conditional on a rich set of observed covariates, there are no unobservable factors that are related to both the length of training chosen by the individual and her outcome of interest (future employment probability). We estimate the DRF and its derivative employing both a parametric and a semiparametric method, and provide indirect evidence of the validity of their underlying identifying assumption in the application's context.

Our estimates for the entire sample of dropouts indicate that the DRF of training duration on the employment probability is relatively flat, turning into an increasing function only after a certain threshold of the training duration (about 14 weeks). In addition, we estimated the derivative of the DRF, which can be interpreted as the marginal effect of an additional week of training on the employment probability. These estimates reveal that the marginal effects hover around zero for short durations of training up until about the 14th week, after which they become positive and eventually statistically significant. When summarizing the results in the form of average derivatives for different ranges of training durations, we find that there are positive and significant effects for the entire sample in the second half of the training durations range (13 to 26

25

weeks), reinforcing the notion that there are positive and significant marginal effects after a certain threshold of duration is met. Finally, we separately analyze subsamples corresponding to male and female dropouts. The main differences between the two is that males show higher positive estimated marginal effects than do females and that the estimated average derivatives for males are larger (sometimes considerably) and statistically significant.

Overall, this study provides evidence on the value of job training programs for participants that do not complete them. Our results have implications for policy in that, given that positive marginal effects appear only after a minimum length of enrollment is attained, policy makers should encourage trainees to stay in this program for as long as they can and definitely more than 12−15 weeks. Finally, while the results only apply directly to the Korean training program we analyzed, similar results could be found in comparably designed training programs in other countries.

6. References

Aedo, C. and Nuñez, S. (2004), "The Impact of Training Policies in Latin America and the Caribbean: The Case of Programa Joven," Research Network Working Paper R-483, Inter-American Development Bank.

Attanasio, O., Kugler, A., and Meghir, C. (2009), "Subsidizing Vocational Training for Disadvantaged Youth in Developing Countries: Evidence from a Randomized Trial," IZA Discussion Paper No. 4251.

Betcherman, G., Olivas, K., and Dar, A. (2004), "Impacts of Active Labor Market Programs: New Evidence from Evaluations with Particular Attention to Developing and Transition Countries," Social Protection Discussion Paper Series, The World Bank.

Bia, M. and Mattei, A. (2008), "A STATA Package for the Estimation of the Dose-Response Function through Adjustment for the Generalized Propensity Score," *The Stata Journal*, 8, 354-373.

Busso, M., DiNardo, J., and McCrary, J. (2008), "Finite Sample Properties of Semiparametric Estimators of Average Treatment Effects," Mimeo, University of Michigan.

Campos-Vazquez, R. and Chiapa, C. (2011), "Estudio Sobre la Temporalidad en el Programa [Oportunidades] y Nuevos Esquemas de Apoyo y Corresponsabilidad para las Familias," Mimeo, El Colegio de Mexico.

Calderón-Madrid, A. (2006), "Revisiting the Employability Effects of Training Programs for the Unemployed in Developing Countries," RES Working Paper R-522, Inter-American Development Bank.

Card, D., Ibarraran, P., Regalia, F., Rosas D., and Soares, Y. (2007), "The Labor Market Impacts of Youth Training in the Dominican Republic: Evidence from a Randomized Evaluation," NBER Working Paper No. 12883.

Card, D., Kluve, J., and Weber, A. (2010), "Active Labor Market Policy Evaluations: A Meta-Analysis," *The Economic Journal*, 120, F452-F477.

De Crombrugghe, D., Espinoza, H., and Heijke, H. (2009), "Dropout Behavior in Youth Job Training Programmes: The Case of Projoven-Peru," Mimeo, Maastricht University.

Dehejia, Rajeev H. and Wahba, Sadek (2002), "Propensity Score-Matching Methods for Nonexperimental Causal Studies," *Review of Economics and Statistics*, 84(1), 151-161.

Fan. J. and Gibels, I. (1996), *Local Polynomial Modeling and Its Applications*, Chapman and Hall: London.

Flores, C. (2005), "Estimation of Dose-Response Functions and Optimal Doses with a Continuous Treatment," Doctoral Dissertation, University of California at Berkeley.

Flores, C. and Mitnik, O. (2009), "Evaluating Nonexperimental Estimators for Multiple Treatments: Evidence from Experimental Data," Mimeo, Department of Economics, University of Miami.

Flores, C., Flores-Lagunes, A., Gonzalez, A., and Neumann, T. (forthcoming), "Estimating the Effects of Length of Exposure to Instruction in a Training Program: The Case of Job Corps," *The Review of Economics and Statistics*.

Gerfin, M. and Lechner, M. (2002), "A Microeconometric Evaluation of the Active Labour Market Policy in Switzerland," *Economic Journal*, 112(482), 854-893.

Hardin, J, and Hilbe, J. (2007), *Generalized Linear Models and Extensions*, 2nd Ed. College Station: Stata Press.

Heckman, J., Hohmann, N., Smith, J., and Khoo, M. (2000), "Substitution and Dropout Bias in Social Experiments: A Study of an Influential Social Experiment," *Quarterly Journal of Economics*, 115(2), 651-694.

Heckman, J., and Hotz, J. (1989), "Choosing Among Alternative Nonexperimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training," *Journal of the American Statistical Association*, 84:408, 862-874.

Heckman, J., LaLonde, R., and Smith, J. (1999), "The Economics and Econometrics of Active Labor Market Programs." In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, 3A. Amsterdam, New York and Oxford: Elsevier Science North-Holland, 1865-2097.

Heckman, J., and Navarro, S. (2007), "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*, 136:2, 341-396.

Heckman, J., Smith, J., and Taber, C. (1998), "Accounting for Dropouts in Evaluations of Social Programs," *Review of Economics and Statistics*, LXXX, 1-14.

Hirano, K. and Imbens, G. (2004), "The Propensity Score with Continuous Treatments." In Andrew Gelman and Xiao-Li Meng (eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. West Sussex: John Wiley and Sons, 73-84.

Ibarrarán, P. and Rosas Shady, D. (2009), "Evaluating the Impact of Job Training Programs in Latin America: Evidence from IDB Funded Operations," Mimeo, Inter-American Development Bank.

Ibarrarán, P. and Villa, J. (2010), "Labor Insertion Assessment of Conditional Cash Transfer Programs: A Dose-Response Estimate for Mexico's Oportunidades," Mimeo, Inter-American Development Bank.

Ichimura, H. and Todd, P. (2007), "Implementing Nonparametric and Semiparametric Estimators." In J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, 6B. Amsterdam, New York and Oxford: Elsevier Science North-Holland, 5369-5468.

Imbens, G. (2000), "The Role of the Propensity Score in Estimating Dose-Response Functions," *Biometrika*, 87, 706-710.

Imbens, G. (2004), "Nonparametric Estimation of Average Treatment Effects under Exogeneity: A Review," *Review of Economics and Statistics*, 86(1), 4-29.

Kluve, J., Lehmann, H., and Schmidt, C. (1999) "Active Labor Market Policies in Poland: Human Capital Enhancement, Stigmatization, or Benefit Churning?" *Journal of Comparative Economics* 27 (1): 61-89.

Kluve, J., Schneider, H., Uhlendorff, A. and Zhao, Z. (2007), "Evaluating Continuous Training Programs Using the Generalized Propensity Score," IZA Discussion Paper No. 3255.

Lechner, M. (2002), "Program Heterogeneity and Propensity Score Matching: An Application to the Evaluation of Active Labor Market Policies." *Review of Economics and Statistics*, 84(2), 205-220.

Lechner, M. and Wunsch, C. (2011), "Sensitivity of Matching-Based Program Evaluations to the Availability of Control Variables," IZA Discussion Paper No. 5553.

Lee, M. and Lee, S. (2003), "Analyzing Effects of Job-Trainings Suffering Dropouts with an Optimal Multiple Matching," Mimeo, Office of Research, Singapore Management University.

Lee, M. and Lee, S. (2005), "Analysis of Job-Training Effects on Korean Women," *Journal of Applied Econometrics*, 20, 549–562.

Lee, M. and Lee, S. (2009), "Sensitivity Analysis of Job-Training Effects on Reemployment for Korean Women," *Empirical Economics*, 36(1), 81-107.

Lee, D. and Lemieux, T. (2010), "Regression Discontinuity Designs in Economics,"*Journal of Economic Literature*, 48:2, 281-355.

Lubyova, M. and van Ours, J. (1999), "Effects of Active Labor Market Programs on the Transition Rate from Unemployment into Regular Jobs in the Slovak Republic," *Journal of Comparative Economics*, 27, 90-112.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. 2nd Ed. Chapman and Hall/CRC.

Mitnik, O. (2008), "Intergenerational Transmission of Welfare Dependency: The Effects of Length of Exposure," Working Paper, University of Miami.

Na, Y., Jeong, W., Lee, S. and Lee, J. (2007), "The Study to Improve the Vocational Training System in South Korea." Korea Research Institute for Vocational Education and Training.

Newey, W. (1994), "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233-253.

Rodriguez-Planas, N. and Jacob, B. (2010), "Evaluating Active Labor Market Programs in Romania," *Empirical Economics*, 38(1), 65-84.

Rosenbaum, P. and Rubin, D. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70(1), 41-55.

Smith, J. and Todd, P. (2005), "Does Matching Overcome Lalonde's Critique of Nonexperimental Estimators?" *Journal of Econometrics*, 125(1-2), 305-353.

Waller, M. (2008), "Further Training for the Unemployed: What Can We Learn About Dropouts from Administrative Data?" FDZ Methoden Report No. 4/2008, Institute for Employment Research, Nuremberg, Germany.

Yoo, G. and Kang, C. (2010), "The Impacts of Vocational Training on Earnings in Korea: Evidence from the Economically Active Population Survey," *KDI Journal of Economic Policy*, 32(2), 29-53.

Figure 1: Histograms of the length of enrollment in job-training for each group

Figure 2: Dose response function (DRF) and its estimated derivative on employment probability one year post job-training using alternative estimators
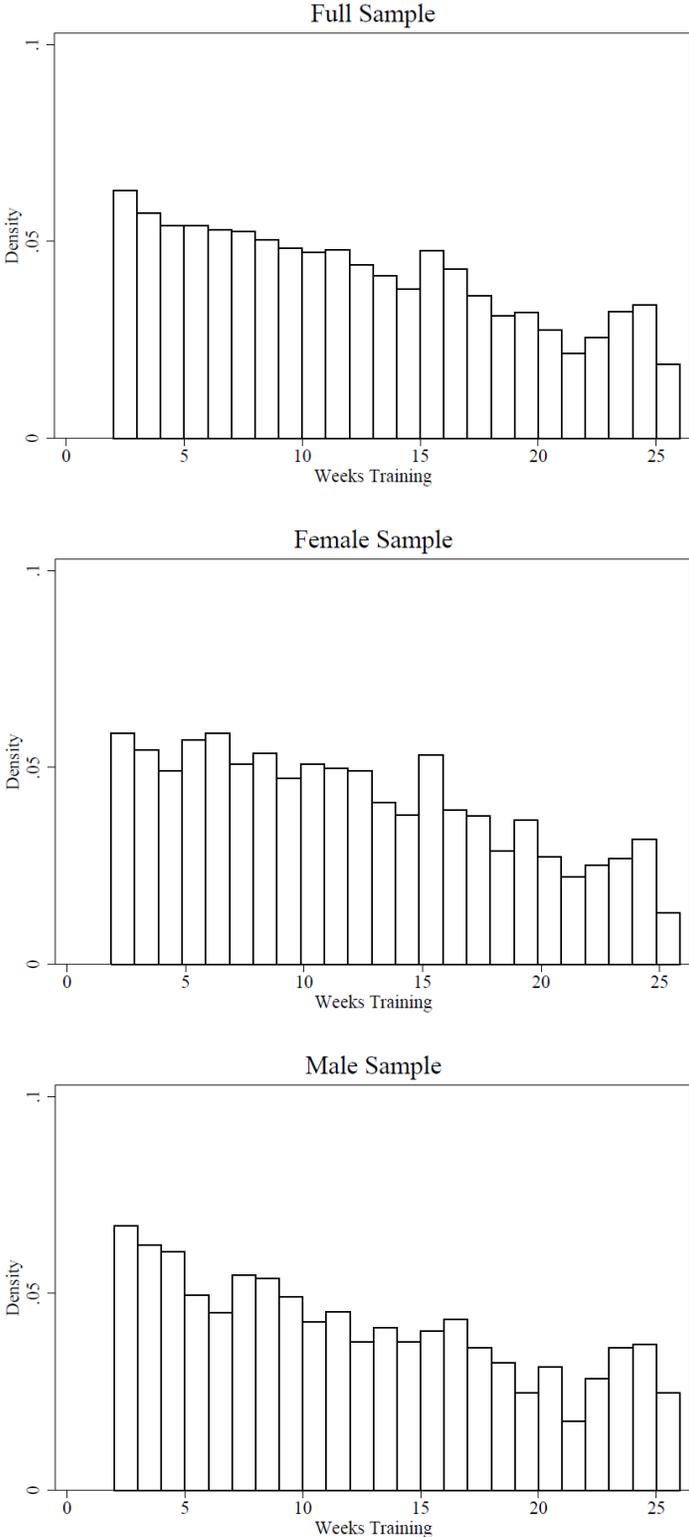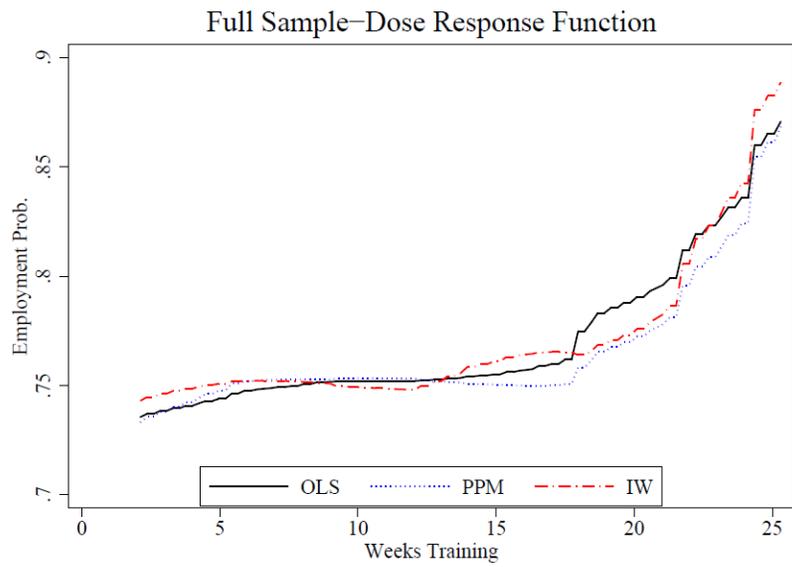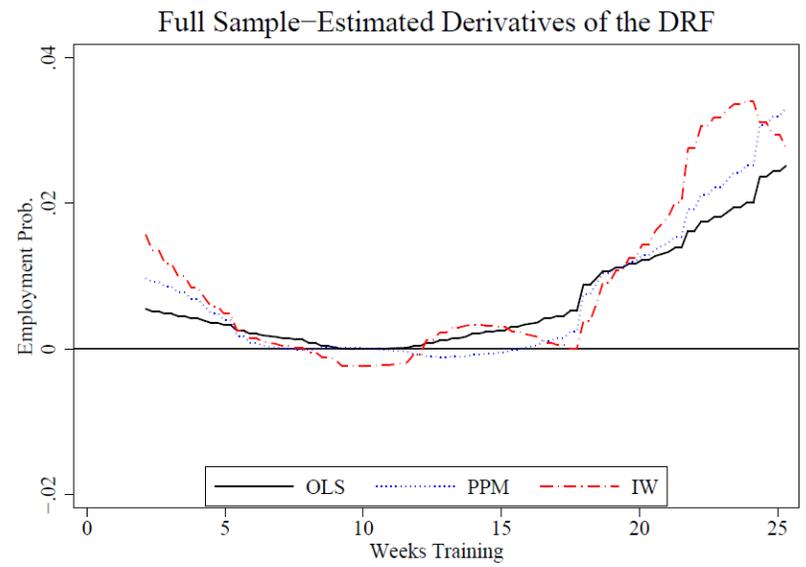
Figure 3: Dose response function (DRF) on employment probability one year post job-training using alternative estimators
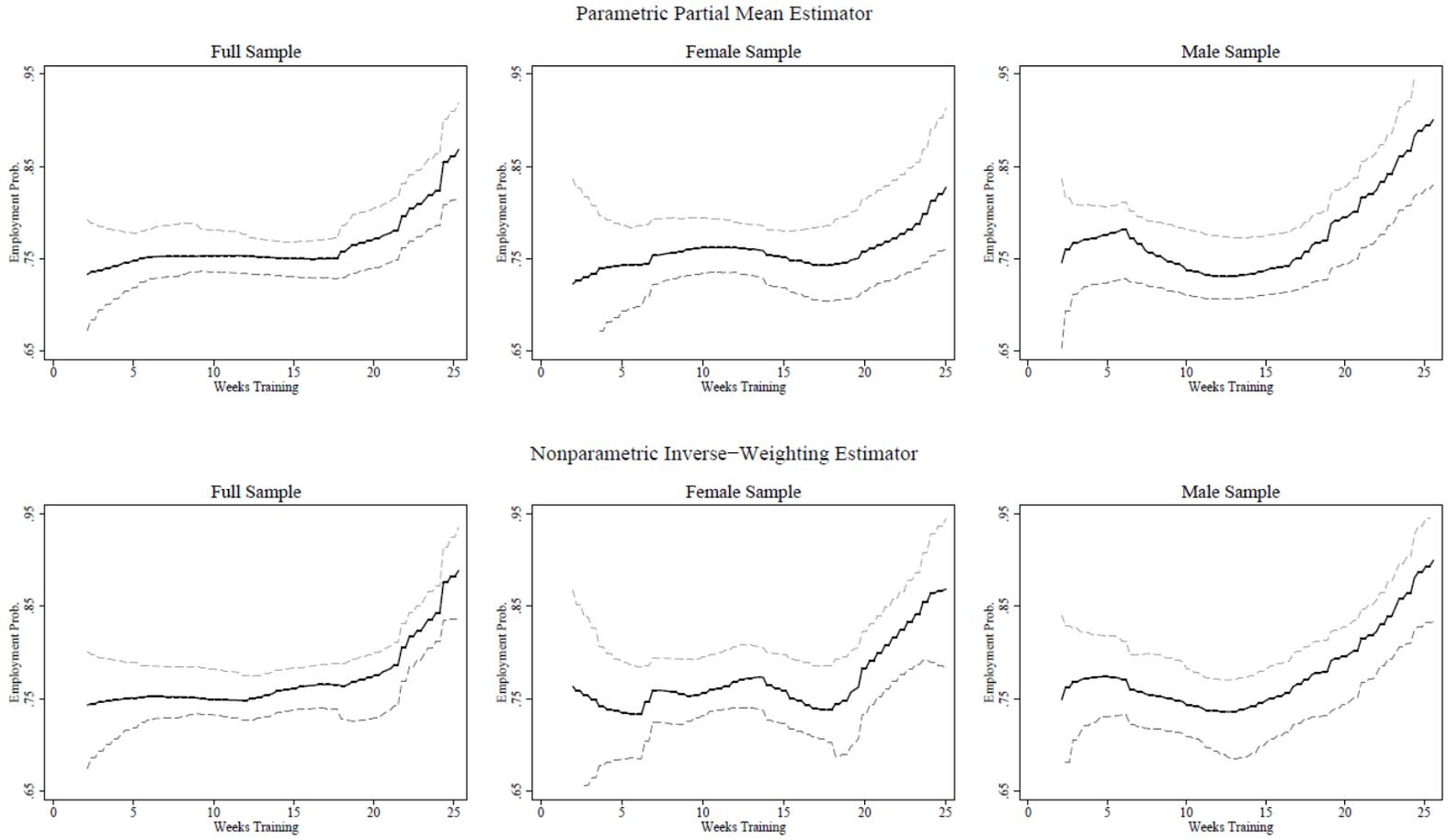

Parametric Partial Mean Estimator

Figure 4: Estimated derivative of the DRF on employment probability one year post job-training using alternative estimators

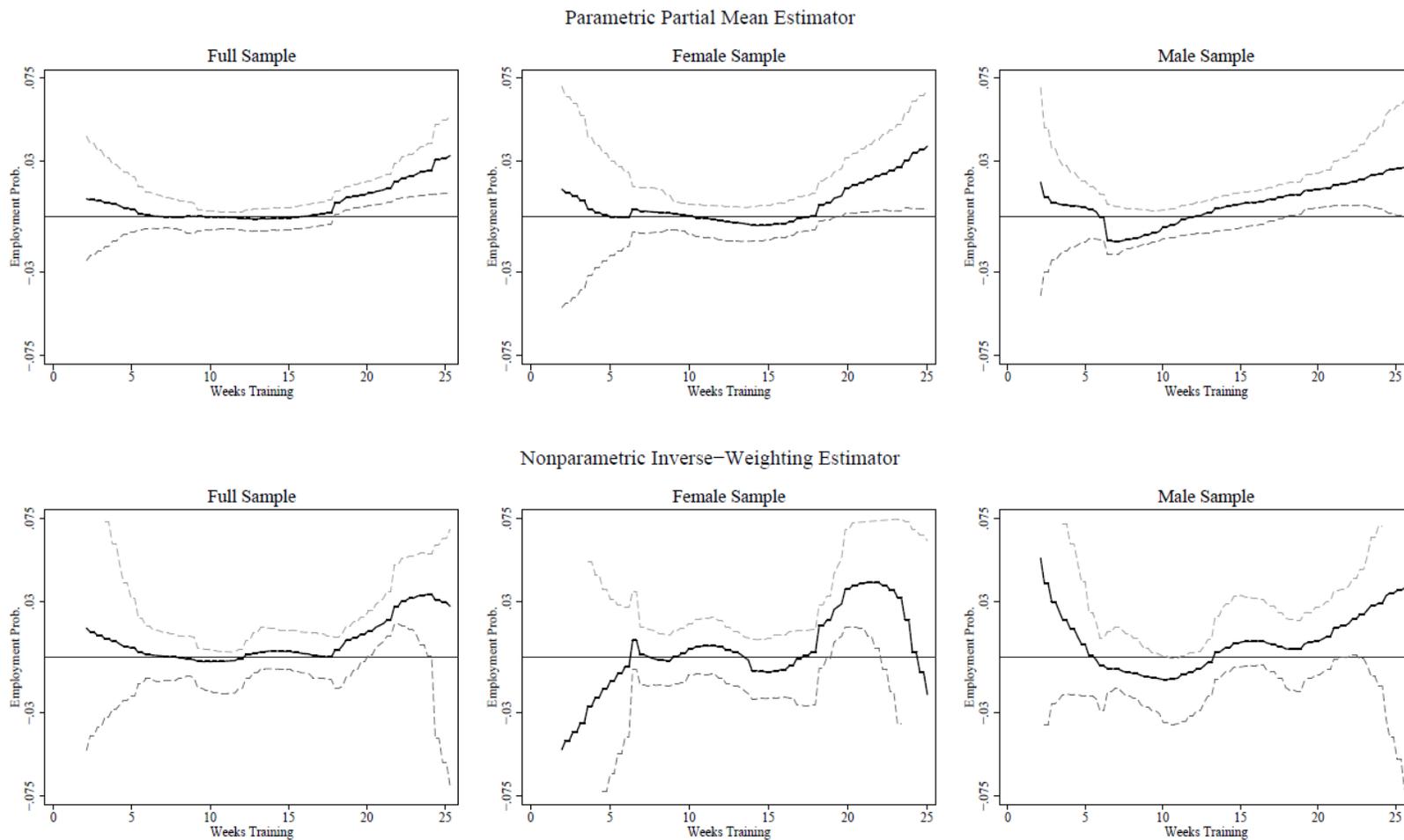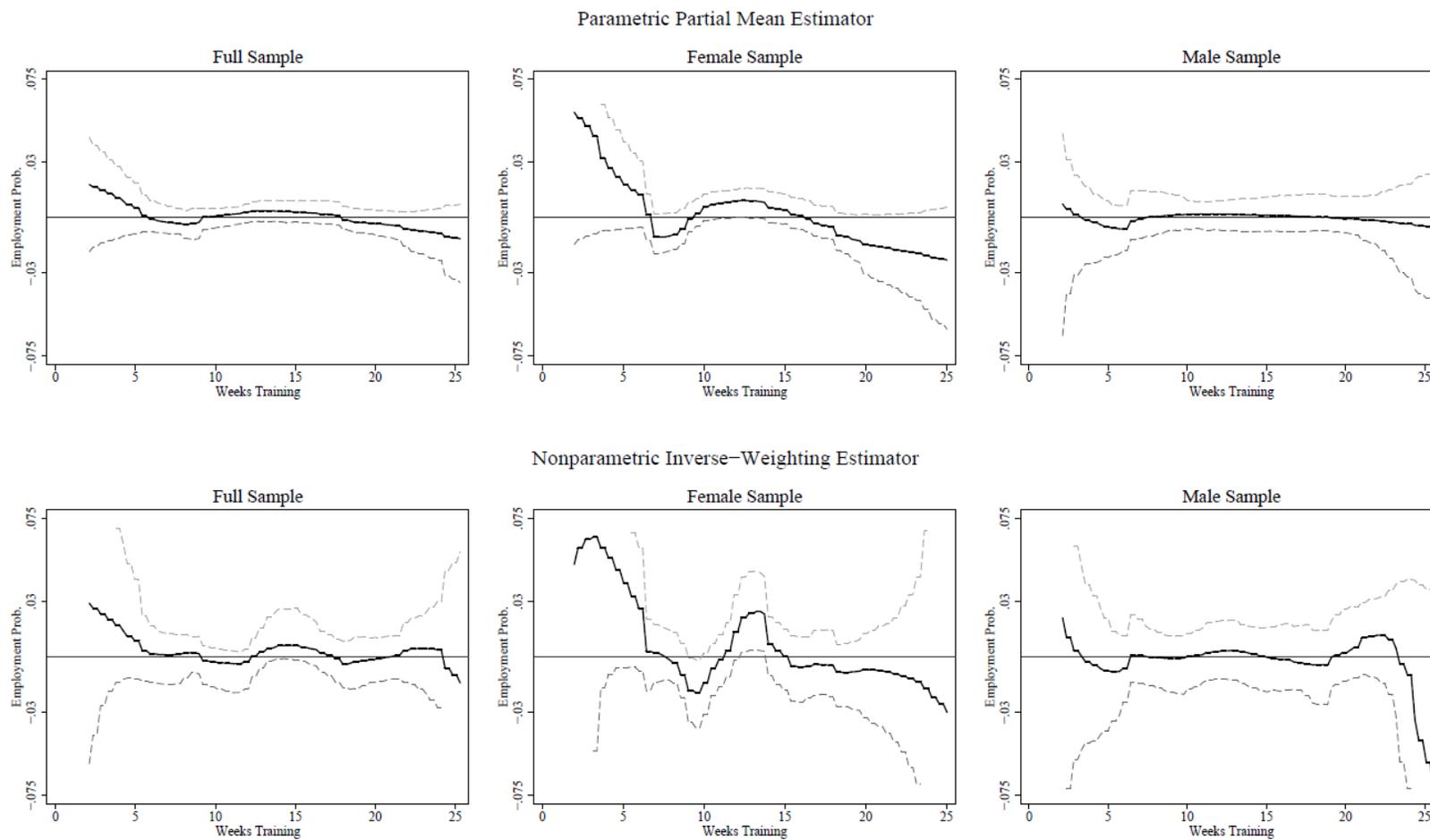Figure 5. Estimated derivative of the DRF on a placebo outcome (employment one year prior to training)

34

	Full Sample		Female Sample		Male Sample	
	Mean	Std Dev	Mean	Std Dev	Mean	Std Dev
Employment prob 1 year post-training	0.779	0.415	0.776	0.417	0.782	0.413
Weeks training	12.024	6.668	11.870	6.488	12.061	6.872
Planned weeks training	22.335	4.730	22.376	4.690	22.261	4.796
Male	0.449	0.497	0.000	0.000	1.000	0.000
Age	31.469	6.106	31.214	6.225	31.830	5.948
Middle school graduates	0.008	0.091	0.007	0.085	0.010	0.097
High school graduates	0.482	0.500	0.469	0.499	0.501	0.500
College graduates	0.259	0.438	0.275	0.447	0.239	0.426
University graduates & over	0.251	0.433	0.248	0.432	0.251	0.434
Unemployment spell before joining training (days)	227.821	286.003	219.812	280.372	238.971	293.891
Unemployment rates						
By gender/age (month of exit)	4.582	2.441	3.874	1.870	5.435	2.758
By gender/age (1 month before exit)	4.630	2.485	3.941	1.931	5.460	2.807
By gender/age (2 months before exit)	4.662	2.509	3.980	1.954	5.483	2.841
By gender/educ. (month of exit)	3.961	0.836	3.868	0.788	4.078	0.881
By gender/educ. (1 month before exit)	4.046	0.872	3.974	0.842	4.137	0.902
By gender/educ. (2 months before exit)	4.127	0.915	4.070	0.890	4.200	0.944
By gender/region (month of exit)	3.531	1.072	3.154	0.949	3.995	1.029
By gender/region (1 month before exit)	3.597	1.074	3.236	0.968	4.040	1.029
By gender/region (2 month before exit)	3.660	1.085	3.321	1.007	4.080	1.029
Reason: self-employed	0.119	0.324	0.080	0.271	0.168	0.374
Reason: marriage/baby	0.035	0.185	0.053	0.224	0.014	0.116
Reason: injured/aging	0.026	0.158	0.027	0.161	0.024	0.153
Reason: layoff	0.004	0.060	0.003	0.056	0.005	0.068
Reason: personal	0.382	0.486	0.346	0.476	0.427	0.495
Reason: workplace close	0.033	0.178	0.037	0.189	0.027	0.164
Reason: layoff	0.031	0.174	0.032	0.177	0.031	0.174
Reason: salary delay or changes in work condition	0.020	0.140	0.020	0.140	0.019	0.137
Reason: others related to company	0.272	0.445	0.301	0.459	0.236	0.425
Reason: contract expiration	0.077	0.267	0.101	0.302	0.048	0.214
Previous firm: big firm	0.114	0.318	0.154	0.361	0.064	0.244
Ex-firm size/1000	0.090	0.746	0.096	0.707	0.082	0.790
Industry: agriculture	0.002	0.045	0.001	0.035	0.003	0.055
Industry: manufacturing	0.357	0.479	0.278	0.448	0.455	0.498
Industry: electric, gas	0.002	0.047	0.002	0.047	0.002	0.048
Industry: construction	0.069	0.253	0.061	0.240	0.078	0.269
Industry: retail/whole	0.139	0.346	0.139	0.346	0.138	0.345
Industry: accommodation	0.023	0.149	0.027	0.162	0.018	0.134
Industry: trans./ware.	0.044	0.205	0.026	0.160	0.066	0.248
Industry: communications	0.006	0.077	0.006	0.075	0.006	0.078

Industry: finance/ins.	0.017	0.131	0.024	0.153	0.009	0.093
Industry: real est. rent	0.023	0.151	0.024	0.153	0.021	0.145
Industry: others	0.126	0.332	0.133	0.339	0.118	0.323
Industry: administ, n'tl. defense	0.014	0.117	0.020	0.141	0.005	0.073
Industry: education	0.062	0.242	0.092	0.290	0.024	0.154
Industry: public health	0.074	0.263	0.125	0.330	0.015	0.123
Industry: health-realted	0.011	0.103	0.012	0.107	0.008	0.089
Industry: public	0.030	0.172	0.029	0.169	0.032	0.176
Training match: same	0.126	0.332	0.128	0.334	0.124	0.330
Training match: similar	0.208	0.406	0.231	0.422	0.182	0.386
Training match: different	0.666	0.472	0.641	0.480	0.694	0.461
Training unit cost	26.960	3.406	26.788	2.984	27.191	3.897
Training level: basic	0.313	0.464	0.310	0.462	0.321	0.467
Training level: intermediate	0.474	0.499	0.506	0.500	0.435	0.496
Training level: advanced	0.213	0.410	0.185	0.388	0.243	0.429
Training type: textile	0.029	0.168	0.051	0.220	0.003	0.052
Training type: chemistry	0.007	0.082	0.001	0.031	0.014	0.118
Training type: machine/equp	0.172	0.378	0.043	0.202	0.331	0.471
Training type: construction	0.042	0.200	0.039	0.193	0.045	0.207
Training type: electricity	0.017	0.130	0.001	0.031	0.037	0.189
Training type: electronic	0.049	0.215	0.037	0.189	0.065	0.246
Training type: info./comm.	0.247	0.431	0.244	0.430	0.248	0.432
Training type: manufacturing of trans equip	0.002	0.039	.	.	0.003	0.059
Training type: industrial appl.	0.082	0.274	0.092	0.289	0.067	0.250
Training type: crafts	0.011	0.103	0.015	0.120	0.006	0.075
Training type: service	0.187	0.390	0.255	0.436	0.104	0.305
Training type: clerical	0.147	0.354	0.214	0.410	0.067	0.250
Training type: fanance/insurance	0.004	0.067	0.004	0.061	0.005	0.073
Training type: health	0.002	0.047	0.004	0.064	0.000	0.000
Training type: environment	0.003	0.051	0.001	0.025	0.005	0.070
Training year: 2003	0.194	0.396	0.186	0.389	0.204	0.403
Training year: 2004	0.226	0.418	0.235	0.424	0.214	0.410
Training year: 2005	0.167	0.373	0.170	0.376	0.167	0.373
Training year: 2006	0.322	0.467	0.312	0.464	0.331	0.470
Training year: 2007	0.091	0.287	0.097	0.296	0.085	0.279
Observations	5,803		3,192		2,620	

Table 2: Balancing Tests

	Total Sample	Female Sample	Male Sample
Unrestricted model: T on GPS, GPS^2, GPS^3, and X's			
Test restriction that X's can be excluded from the unrestricted model			
LL Restricted	-15,980.53	-8,308.63	-7,448.34
LL Unrestricted	-16,040.70	-8,338.78	-7,471.83
Test Statistic	120.34	60.30	46.98
p-value	1.00	1.00	1.00
Number of Restrictions	276	148	144
Test restriction that GPS coefs can be excluded from the unrestricted model			
LL Restricted	-15,980.53	-8,308.63	-7,448.34
LL Unrestricted	-16,028.45	-8,333.30	-7,477.68
Test Statistic	95.85	49.35	58.68
p-value	0.00	0.00	0.00
Number of Restrictions	3	3	3
N	4,520	2,338	2,115

Table 3: Average Derivatives for Selected Levels Employing the IW Estimator

	1-99	1-50	25-75	50-99
	2.11-25.31 weeks	2.11-13.71 weeks	7.79-19.63 weeks	13.71-25.31 weeks
Total Sample	0.0080*	0.0033	0.0001	0.0126***
	1.97-25.02 weeks	1.97-13.5 weeks	7.61-19.38 weeks	13.5-25.02 weeks
Female Sample	0.0028	-0.0049	0.0014	0.0104
	2.11-25.59 weeks	2.11-13.85 weeks	7.86-19.84 weeks	13.85-25.59 weeks
Male Sample	0.0084*	0.0015	-0.0010	0.0152**

Note: *, **, and *** indicate that the estimates are statistically significant at the 90, 95, and 99 percent level, respectively. These significance tests are computed using the bootstrap estimates employing the percentile method.

Table 4: Average Derivatives for Selected Levels Employing the IW Estimator - Placebo Outcomes

	Placebo outcome	1-99	1-50	25-75	50-99
		2.11-25.31 weeks	2.11-13.71 weeks	7.79-19.63 weeks	13.71-25.31 weeks
Total Sample	Employed 1 year before training	0.0030	0.0058	0.0008	0.0002
	Employed 6 mth before training	0.0035	0.0056	0.0009	0.0015
	The previous firm is a big firm	-0.0021	-0.0058	0.0009	0.0018
	Uneployment spell before training	-1.6925	-2.8322	-0.8357	-0.6113
		1.97-25.02 weeks	1.97-13.5 weeks	7.61-19.38 weeks	13.5-25.02 weeks
Female Sample	Employed 1 year before training	0.0035	0.0147	-0.0007	-0.0074
	Employed 6 mth before training	-0.0052	0.0111	-0.0011	-0.0212
	The previous firm is a big firm	-0.0039	-0.0089	0.0010	0.0013
	Uneployment spell before training	-2.9803	-8.0595	-0.0644	1.8895
		2.11-25.59 weeks	2.11-13.85 weeks	7.86-19.84 weeks	13.85-25.59 weeks
Male Sample	Employed 1 year before training	-0.0040	-0.0011	-0.0002	-0.0067
	Employed 6 mth before training	-0.0008	-0.0022	-0.0009	0.0006
	The previous firm is a big firm	-0.0010	-0.0067	-0.0017	0.0046
	Uneployment spell before training	2.1205	-1.0823	-0.9813	5.2597