The Effects of Training Incidence and Duration on Labor Market Transitions*

Bernd Fitzenberger,[♮] Aderonke Osikominu,[♯] and Marie Paul[§]

[♮]University of Freiburg, ZEW, IZA, ROA, IFS, bernd.fitzenberger@vwl.uni-freiburg.de
[♯]University of Hohenheim, CESifo, IZA, a.osikominu@uni-hohenheim.de
[§]née Waller, University of Duisburg-Essen, RGS Econ, marie.paul@uni-due.de

This version: August 2015

Abstract: This paper develops a dynamic evaluation approach in discrete time to estimate the impact of training programs for the unemployed on employment transitions. Our framework accounts for the endogeneity of program incidence and duration and considers confounding caused by lagged outcomes and treatments, time-constant unobservables, as well as time-varying observed covariates. We specify a flexible bivariate random effects probit model for employment and training status that we estimate with Bayesian Markov Chain Monte Carlo (MCMC) techniques. Based on our estimates, we simulate different treatment effects of interest. Our estimation results imply positive effects of training on the employment probability of the treated, lying between 6 and 14 percentage points 2.5 years after program start. The effects are higher for women than for men and initially negative effects persist shorter in West Germany than in East Germany. Further, our results show that a longer planned enrollment length of 12 months as opposed to just 6 months leads to an increase in employment rates by 5 to 8 percentage points in the medium run.

Keywords: dynamic treatment effects, dynamic non-linear panel data models, MCMC, training, active labor market programs
JEL: J68, C33, C11

---

* We thank Jaap Abbring, John Ham, Per Johansson, Tiemen Woutersen as well as numerous seminar and conference participants for helpful comments and suggestions. This study is part of the project "Employment effects of further training programs 2000-2002 – An evaluation based on register data provided by the Institute of Employment Research, IAB ('Die Beschäftigungswirkungen der FbW-Maßnahmen 2000-2002 auf individueller Ebene – Eine Evaluation auf Basis der prozessproduzierten Daten des IAB'; IAB project number 6-531.1A). We gratefully acknowledge financial and material support by the IAB. Furthermore, Bernd Fitzenberger gratefully acknowledges financial support from the German Research Foundation through the project 'Statistical Modelling of Errors in Administrative Labour Market Data' and Aderonke Osikominu gratefully acknowledges financial support from the German Excellence Initiative as well as the Austrian Science Fund (FWF grant no. S 10304-G16). The usual caveat applies.

Contents

1 Introduction

Training programs are an important instrument in the tool box of active labor market policy. Yet, they are commonly not perceived as particularly effective.[1] One reason might be that labor market policy pursues heterogeneous, partly conflicting goals. Substantive skill development requires longer-term programs that may initially prolong unemployment. Thus, quick reintegration does not seem to be a viable goal of training, such programs rather aim at integration into high quality jobs.

A second important issue that complicates the evaluation of training programs comes from the methodological side. Standard statistical models applied in most of the literature on estimating the effects of training programs are static (Card et al. 2010). However, program start and continuation are the outcome of dynamic processes. Job-seekers who fail to find a job are eventually assigned to active labor market programs. Similarly, the realized training duration may depend on the success of job search during training.[2] For instance, lucky participants who receive a suitable job offer during training may drop out early, while the unlucky ones continue until the scheduled program end or even prolong participation for lack of job opportunities. Program start and continuation related to the development of employment outcomes after the beginning of unemployment raise endogeneity issues that are difficult to incorporate in static evaluation approaches that are commonly used in the literature.[3]

In this paper we devise an evaluation framework in discrete time that takes the dynamics of program start and continuation into account. Building on the seminal work by Robins (1997) our dynamic framework exploits no-anticipation and conditional randomization conditions for identification.[4] In particular, our no-anticipation condition states that potential employment outcomes associated with training sequences that coincide up to the current period are the same in the current period. Our conditional randomization condition states that conditional on time-constant unobservables, the history of potentially time-varying observed covariates, as well as

[1]See e.g. the surveys of Card et al. (2010), Heckman et al. (1999), and Martin and Grubb (2001).

[2]See Paul (2014) for an empirical analysis of endogenous training dropouts in Germany.

[3]Fredriksson and Johansson (2008) present a formal analysis of the bias that results when applying a static evaluation approach in the case of a dynamic assignment regime as described above.

[4]Robins (1997) refers to the no-anticipation condition as the consistency condition.

the training and employment history up to the current period assignment of training in the current period is as good as random, i.e. independent of future potential employment outcomes. Based on the no-anticipation and conditional randomization conditions we establish identification of dynamic causal effects assuming common support and knowledge of the distribution of the time-constant unobservables. In a next step, we outline how the distribution of the unobservables can be identified given data on employment and training sequences as well as observed covariates. Here we base our argument on Heckman and Navarro (2007) who study nonparametric identification of the distribution of unobservables in dynamic panel data models.

The key methodological innovation of our paper is that we clarify how the no-anticipation and conditional randomization conditions can be implemented when allowing for time-varying observed covariates and time-constant unobservables. These extensions are important from a substantive point of view. In our framework, employment outcomes and program participation may both respond to events that occur during the period of study, such as changes in local labor market conditions. Allowing for unobserved heterogeneity seems important because training programs tend to be assigned to people who experience a series of particularly negative labor market outcomes, which may be unlikely given their observed characteristics but could be related to below average unobserved characteristics.

We then apply our framework to evaluate the effects of training participation on the probability to be employed. The first treatment parameter we study is a classical treatment effect on the treated that evaluates the effect of participation against nonparticipation for those who participate at some point in time during unemployment. Our second treatment parameter of interest is the effect of assigning different planned enrollment lengths. A comparative assessment of different enrollment lengths for the same type of training is important for policymakers interested in an effective use of active labor market programs. Comprehensive training schemes typically range among the most expensive active labor market programs. During training, participants generally search less intensively for a new job (lock-in effect). Therefore, employment effects of training are typically negative in the short run, while positive effects may eventually unfold after completion of the program. Our analysis provides evidence as to whether longer participation associated with negative short-run effects results in sufficiently positive employment effects in the long run, or whether comparable long-run effects can be obtained with shorter programs at lower costs.

We focus on a large scale training program in Germany with a median duration of six months and for which enrollment lengths vary between a couple of weeks and more than one year. We specify a joint model for the transition rates into and out of employment and training using a very flexible bivariate random effects probit model. We account for the full employment and training history of each individual from the start of the first inflow into unemployment onwards. In addition, we account for heterogeneity in terms of time-varying observed covariates such as demographic characteristics and local labor market conditions. Our specification allows in a flexible way for state dependence and duration dependence in the transition rates as well as in the treatment effects, including numerous interactions with observed covariates. The rich administrative data allow us to integrate such flexibility into the model while performing separate estimations by gender and region (East and West Germany). In this respect, our approach is similar to matching analyses that rely on a rich and flexible specification of the observed heterogeneity across individuals. In contrast to matching methods, we account for time-invariant individual specific effects in the treatment equation and the outcome equation. Furthermore, we take account of both the potential dynamic selection given an observed employment and treatment sequence and the potential selection into and out of treatment with respect to the unobservable individual specific effects.

We use Bayesian Markov Chain Monte Carlo (MCMC) techniques that allow a numerically robust estimation of our flexible model specification.[5] We interpret the resulting estimates in a classical perspective focussing on posterior means and standard deviations. A major advantage of the MCMC technique is that it provides the posterior distribution of the individual specific effects. This allows us to account explicitly for the selection on unobservables based on their posterior distribution among the treated when calculating the posterior distribution of various treatment effects of interest, such as the average effect of treatment on the treated or the average effect of varying the planned program duration.

Our main findings are as follows. We estimate positive effects of training on the

[5] See Chib (2001) for a survey on MCMC methods and for recent applications in labor economics see Buchinsky et al. (2010), Horny et al. (2012), and Troske and Voicu (2010).

employment probability unfolding three to four quarters after program start in all four subsamples considered. Our results suggest that participating in training improves the employment probability of the participants by 6 to 14 percentage points 2.5 years after program start. Further, participants benefit from being assigned to programs with a longer planned enrollment length. The employment rate associated with assignment to a twelve-month program as opposed to a six-month program is 5 to 8 percentage points higher. Thus, longer training programs show higher long-run employment gains, which may justify the higher costs involved.

The remainder of this paper is structured as follows. The next Section discusses related dynamic evaluation frameworks in the literature. Section 3 presents our evaluation framework. Section 4 describes the institutional setup, the data and the empirical implementation of our model. Section 5 discusses the main results and further sensitivity analyses. Section 6 concludes. The appendix provides further details on the data, the implementation of the estimation approach, and detailed estimation results.

2 Related Literature on Dynamic Treatment Effects

Estimating the effect of dynamic treatment start and continuation on the evolution of the outcome of interest has intrigued researchers for quite some time (see also Abbring and Heckman, 2007, for an overview). Robins (1986) has pioneered the conceptual foundations. Robins (1997, henceforth RO) provides an overview for general treatment sequences in discrete time. Identification of dynamic causal effects relies on (i) a no-anticipation condition (denoted consistency condition by RO), i.e. that for two treatment sequences in discrete time which are the same up to some period, potential outcomes are also the same up to this period; (ii) a sequential randomization condition regarding treatment assignment in the current period conditional on the history of covariates and outcomes so far; and (iii) a support condition, i.e. that, in any period of interest, there exist both treated and nontreated units that have experienced the same treatment and outcome history so far. Under these assumptions, Robins (1997) shows that identification of the distribution of an outcome trajectory under a counterfactual treatment sequence is possible based on period-to-period conditional probabilities for the outcome in the next period conditional on the history of outcomes and treatments so far (his so-called "g-computation formula"). The approach of Robins (1997) is sufficiently general to handle both the case of program start and continuation under selection on observables.

In the economics literature, research in this area has expanded especially since the early 2000s. Early contributions include Ham and LaLonde (1996, henceforth HL) and Eberwein, Ham, and LaLonde (1997, henceforth EHL). They jointly model discrete time hazard rates between the labor market states unemployment, training, and employment. Their empirical analysis relies on data in which program assignment is randomized at some baseline point. HL and EHL emphasize however the importance of selectivity in program take up and drop out as well as take up of employment after initial randomization. To address the selection on unobservables they allow the hazard rates to depend on a common unobserved heterogeneity term. They further propose to use randomization at baseline as well as time-varying covariates (e.g. local unemployment rates) as source of exogenous variation supporting identification. They do not provide an explicit further account of the conditions that allow them to identify causal effects.

A key aspect of dynamic treatment assignment highlighted in the more recent literature evaluating active labor market programs (e.g. Abbring and van den Berg, 2003; Fredriksson and Johansson, 2008; Sianesi, 2004) is that eligibility for program participation is tied to the labor market status. In many countries, job-seekers are eligible for program participation as long as they remain unemployed. Thus, program assignment and exit from unemployment are two competing risks. If exit from unemployment occurs first, the waiting time until treatment is censored. Conversely, if assignment to treatment occurs first, the exit rate from unemployment may change as a consequence of the treatment. Because of this feature of program assignment, the no-anticipation condition plays a crucial role for identification of causal effects. Researchers have proposed different strategies ranging from matching methods to duration models in discrete or continuous time to estimate causal effects in this setup.

Sequential matching techniques mimic a sequential randomization into treatment of individuals who have been unemployed for the same elapsed duration. Analogously to RO's sequential randomization condition, current treatments and potential outcomes associated with current and future treatments are assumed to be independent conditional on observed covariates and elapsed unemployment duration. While applied studies typically do not provide an explicit discussion, the nonanticipation condition is typically implied by the presumed unconfoundedness assumption required for identification (see discussion in Abbring and Heckmann, 2007). The sequential matching approach allows one to estimate the average effect of receiving treatment at a given elapsed unemployment duration versus not now, implying the possibility of receiving treatment at a later time, i.e. the effect of treatment now versus waiting (Sianesi, 2004). However, as this approach focuses on identification at given elapsed durations and does not specify a model for the selection over time, it is not possible to estimate a causal effect averaged across starting elapsed unemployment durations or a causal effect of treatment against the alternative of no treatment at all.[6]

Fredriksson and Johansson (2008, henceforth FJ) suggest a hybrid approach that combines sequential matching with the estimation of a discrete-time hazard rate model. FJ focus on the effect of starting treatment at some time during an unemployment spell on the exit from unemployment.[7] FJ model the hazard rate from open unemployment to treatment and the hazard rate from unemployment to employment. FJ assume unconfoundedness of the two hazards conditional on the elapsed duration in unemployment and time-constant observed covariates. Their unconfoundedness assumption is a special case of the no-anticipation and sequential randomization conditions of RO. FJ then show how to recover the nontreatment hazards for the treated based on the observed exit rate of the not yet treated conditional on elapsed unemployment duration and observed covariates. The product of the counterfactual period-to-period survival rates in unemployment provides the expected remaining duration of unemployment without treatment. In fact, this Kaplan-Meier type estimator is a version of RO's g-computation formula.

Another related strand of the literature uses continuous-time duration models to estimate dynamic treatment effects. In particular, Abbring and van den Berg (2003) develop an evaluation framework in continuous-time that relies on analogous conditional no-anticipation and randomization conditions as in the discrete time setting. Osikominu (2013) provides a recent extension and application of this approach al-

[6]Recent applications of this approach to the evaluation of training programs include Biewen et al. (2014), Dyke et al. (2006), and Hotz et al. (2006).

[7]FJ do not explicitly relate their approach to RO because they argue that dynamic treatment assignment, i.e. treatment may only be observed when individuals are still unemployed, is not covered by the early work of Robins (1986).

lowing e.g. for time-varying observed covariates and heterogeneous treatment effects. The continuous-time approach allows for selection on time-constant unobservables. The conditions for identification of the joint distribution of the unobservables differ however from the analysis in discrete-time. Assuming separability of the different components that influence the hazard rates, i.e. occurrence dependence, duration dependence, covariates as well as unobserved heterogeneity (i.e. mixed proportional hazard rates), the distribution of the unobserved heterogeneity terms can be identified nonparametrically with minimal time-constant variation of observed covariates. With richer data that include time-varying covariates or repeated observations on each cross-sectional unit, only separability in the influence of unobserved heterogeneity is required.

Heckman and Navarro (2007, henceforth HN) discuss how a dynamic treatment that corresponds to a stopping time can be modeled in a structural dynamic discrete choice and outcome model. Their conceptual setup differs from the literature studying dynamic evaluation of active labor market programs (e.g. Abbring and van den Berg, 2003, FJ, and Sianesi, 2004). In the latter literature, the no-anticipation assumption plays a crucial role because future treatments are not observed after exit from the baseline outcome state (unemployment). Thus, the waiting times until treatment start and until exit from the baseline state are two competing risks. The competing risks aspect is absent in HN. HN allow for anticipation effects of future treatment on the current potential outcome. Further, they provide a detailed analysis of the conditions that allow one to identify the joint distribution of potentially time-varying unobservables influencing choices and outcomes in a semiparametric or nonparametric way.

Our analysis contributes to this literature in several ways. First, we provide a general formal treatment of dynamic treatment evaluation that highlights the links between different assumptions and specifications that coexist in a seemingly unconnected way in the previous literature. Second, we consider both confounding through time-constant unobservables and time-varying observed covariates, which is important from a substantive point of view. We clarify the different assumptions on the evolution of outcomes, treatments, observed covariates, and unobservables over time that are necessary to identify causal effects in a dynamic setup. Moreover, we present novel evidence on how the impact of a given type of training program varies with the length of enrollment. Such evidence is important for policymakers interested in balancing short-run costs and long-run gains of program participation.

3 Evaluation Framework

3.1 Identification

This section develops a dynamic causal framework to evaluate the effect of training incidence and duration on the probability to be employed in a given month. We consider a setup like this: People can either be unemployed or employed. Unemployed may participate in training programs offered by the local employment agency. In particular, the unemployed and the caseworker at the local employment agency meet repeatedly during the unemployment spell. At any such occasion, the caseworker decides whether to assign the unemployed to a program or to postpone participation to the future, waiting further how job search evolves. Program participation may start at random points in time during unemployment. Once an unemployed takes up a new job he/she is not eligible for training anymore. Thus, an unemployed in open, i.e. untreated, unemployment is exposed to two risks that compete to end open unemployment: start of a training program and start of a new job. Training programs are characterized by different planned durations that are determined at the moment of program start. The realized training duration may differ from the planned duration because participants may drop out early or prolong training participation. Drop out and continuation are potentially related to the success of job search.

We adopt a dynamic version of the potential outcome approach to causal inference (Neyman, 1923; Roy, 1951; Rubin, 1974; Robins, 1986). Let Q_t denote a binary random variable for training status in period t, $t = 1, \ldots, T$, with $Q_t = 1$ if somebody participates in training and $Q_t = 0$ otherwise. Note that we will use an inflow sample into unemployment such that t denotes the time since the start of the first unemployment spell. In our dynamic setup, a treatment of interest corresponds to a sequence of training states. For instance, a training participation that starts in period $t = k$, $2 < k < T - 5$, and lasts until period $k + 3$ is characterized by the sequence $Q_1 = 0, \ldots, Q_{k-1} = 0, Q_k = 1, Q_{k+1} = 1, Q_{k+2} = 1, Q_{k+3} = 1, Q_{k+4} = 0, \ldots, Q_T = 0$. We denote such sequences by q, i.e. q is a $T \times 1$ vector of zeroes and ones. At the same time, when participation in a training program starts, the caseworker assigns a planned program duration P which can differ from the actual length of program participation.

Next, we define the potential outcomes. Let $E_t(q)$ be a binary random variable corresponding to the potential employment status in period t associated with training sequence q. $E_t(q) = 1$ if somebody is employed and $E_t(q) = 0$ otherwise. We use $E_t(0)$ to denote the potential employment status under no treatment, i.e. $q = (0, \dots, 0) = \mathbf{0}$. Further, we adopt the convention that $E^t(q)$ denotes the sequence of potential employment states from period 1 through t and similarly q^t denotes the treatment sequence until period t. A variable without a time index, e.g. $E(q)$ or q, corresponds to its sequence from period 1 through T. Let E_t be the actual employment status. For somebody receiving treatment q, we have that $E \equiv E(q)$ while the other potential outcomes are counterfactual.

Now, we characterize the probability to be employed and to participate in training in period t conditional on the information available up to that time. Let $\boldsymbol{X}_t$ denote a vector of possibly time-varying random variables and $\boldsymbol{\alpha} = (\alpha_E, \alpha_Q)$ a vector of time-constant random variables. While $\boldsymbol{X}_t$ represents covariates that can be observed by both the job-seeker/caseworker and the researcher, $\boldsymbol{\alpha}$ is not observable for the researcher.[8] The timing of events within each period t is that $\boldsymbol{X}_t$ is determined first, then E_t, and finally Q_t. The unobserved individual specific effects $\boldsymbol{\alpha}$ are determined before the first period, thus characterizing the individual history before the inflow into unemployment defining our sample (see section 4.2). We take account of both the potential dynamic selection given an observed employment and treatment sequence and the potential selection into and out of treatment with respect to $\boldsymbol{\alpha}$.

Let $\mathcal{I}_t \equiv \{E^{t-1}, Q^{t-1}, \boldsymbol{X}^t, \boldsymbol{\alpha}\}$ denote the information available when the employment decision is made in period t. In the first period, we set $\mathcal{I}_1$ equal to $\{\boldsymbol{X}_1, \boldsymbol{\alpha}\}$.[9] Formally, $\mathcal{I}_t$ is the σ-algebra generated by the random variables $\{E_{r-1}, Q_{r-1}, \boldsymbol{X}_r, \boldsymbol{\alpha}, 0 < r \leq t\}$. Similarly, denote by $\mathcal{J}_t \equiv \{E^t, Q^{t-1}, \boldsymbol{X}^t, \boldsymbol{\alpha}\}$ the information available when the training decision is made in period t. We set $\mathcal{J}_1$ equal to $\{E_1, \boldsymbol{X}_1, \boldsymbol{\alpha}\}$. Now, we

[8] Below we will further distinguish between observed covariates that affect the employment status, $\boldsymbol{X}_{E,t} \subseteq \boldsymbol{X}_t$, and covariates that affect training participation, $\boldsymbol{X}_{Q,t} \subseteq \boldsymbol{X}_t$. This means that potentially only a subset of the available covariate information is relevant for employment decisions and similarly also for training.

[9] Since we consider a population of individuals who start an unemployment episode, the initial states are identical for everybody, i.e. $E_{i0} = 1$ and $Q_{i0} = 0$ for all individuals i.

can model the conditional employment and training probabilities as follows:

$$\begin{aligned} \Pr(E_t(q) = 1 \,|\, \mathcal{I}_t) &= \Pr\left[\psi_E(E^{t-1}, Q^{t-1}, \boldsymbol{X}_t) + \alpha_E + \varepsilon_{E,t} > 0\right] \\ \Pr(Q_t = 1 \,|\, \mathcal{J}_t) &= \Pr\left[\psi_Q(E^{t}, Q^{t-1}, \boldsymbol{X}_t) + \alpha_Q + \varepsilon_{Q,t} > 0\right] \end{aligned} \tag{1}$$

where $\psi_E()$ and $\psi_Q()$ are index functions that summarize the dependence on the history so far, i.e. on the observed variables in $\mathcal{I}_t$ or $\mathcal{J}_t$. The error terms $\varepsilon_{E,t}$ and $\varepsilon_{Q,t}$ are mutually independent martingale difference sequences with respect to $\mathcal{I}_t$ and $\mathcal{J}_t$, respectively. Thus, $\varepsilon_{E,t}$ and $\varepsilon_{Q,t}$ are purely transitory with no serial dependence. Since training participation is only possible during unemployment we set $\Pr(Q_t = 1 \,|\, \mathcal{J}_t) = 0$ if $E_t = 1$. Further, we set $\Pr(Q_t = 1 \,|\, \mathcal{J}_t) = 0$ if somebody has exited training in a previous period, i.e. $Q^{t-1} \neq \mathbf{0}$ and $Q_{t-1} = 0$, because the institutional regulations rule out repeated training participation within a period of three years.

3.1.1 Contrasting Alternative Treatments

We first compare the probability to be employed under treatment q, $\Pr[E_t(q) = 1 \,|\, \mathcal{I}_t]$, with the probability to be employed under the alternative treatment q', $\Pr[E_t(q') = 1 \,|\, \mathcal{I}_t]$. We now discuss the conditions under which such a contrast identifies a causal treatment effect. The first condition is the no-anticipation condition or, in the terminology of RO, the consistency condition:

No-Anticipation Condition (NAC):

$$E_t(q) = E_t(q') \quad \text{if} \quad q^{t-1} = (q')^{t-1} \tag{2}$$

For two treatment sequences q and q', the expression $q^{t-1} = (q')^{t-1}$ means that the treatment sequences coincide up to period $t-1$, i.e. $q_r = q'_r$ for all $r < t$. Under no-anticipation, the mapping $E_t(q) \equiv y_t(E^{t-1}, q, \boldsymbol{X}_t, \alpha_E, \varepsilon_{E,t})$ reduces to $y_t(E^{t-1}, q^{t-1}, \boldsymbol{X}_t, \alpha_E, \varepsilon_{E,t})$ – with slight abuse of notation – and equals $y_t(E^{t-1}, (q')^{t-1}, \boldsymbol{X}_t, \alpha_E, \varepsilon_{E,t}) \equiv E_t(q')$ if $q^{t-1} = (q')^{t-1}$. Thus, the no-anticipation condition ensures that, conditional on the treatment history observed so far, there is a unique correspondence between current potential employment status and current observed employment status.[10] The no-anticipation condition further implies that current potential outcomes are independent of future treatments conditional on

[10]Therefore, conditioning on $(E^{t-1}, Q^{t-1} = q^{t-1})$ is equivalent to conditioning on $(E^{t-1}(q), Q^{t-1} = q^{t-1})$.

the treatment history observed so far. The no-anticipation condition implies the following weak version:[11]

$$\Pr[E_t(q) = 1 \,|\, \mathcal{I}_t] = \Pr[E_t(q') = 1 \,|\, \mathcal{I}_t] \quad \text{if} \quad Q^{t-1} = q^{t-1} = (q')^{t-1} \tag{3}$$

Thus, future values of time-varying covariates must evolve independently of current values of potential outcomes, because otherwise independence of future treatments and current potential outcomes would not hold given the current information set only ($\mathcal{I}_t$), see also equation (5) below. Weak no-anticipation implies that $\Pr[E_t(q) = 1 \,|\, E^{t-1}, Q = q, \boldsymbol{X}, \boldsymbol{\alpha}] = \Pr[E_t(q) = 1 \,|\, E^{t-1}, Q^{t-1} = q^{t-1}, \boldsymbol{X}^t, \boldsymbol{\alpha}]$.

The second condition is the following sequential randomization condition, which is analogous to RO equation (3.2).

Sequential Randomization Condition 1 (SRC.1):

$$Q_t \perp\!\!\!\perp [E_{t+1}(q), \ldots, E_T(q)] \,|\, \mathcal{J}_t\,, \tag{4}$$

and

$$Q_t \perp\!\!\!\perp (\boldsymbol{X}_{t+1}, \ldots, \boldsymbol{X}_T) \,|\, \mathcal{J}_t\,. \tag{5}$$

where $\perp\!\!\!\perp$ represents statistical independence. This strong sequential randomization condition assumes that current treatment status does not anticipate the potential future employment states and the future covariates conditional on the history so far and conditional on unobserved heterogeneity.[12]

The strong sequential randomization condition gives rise to the following weak version:[13]

$$\Pr[Q_t = 1 \,|\, E(q), Q^{t-1} = q^{t-1}, \boldsymbol{X}^t, \boldsymbol{\alpha}] = \Pr[Q_t = 1 \,|\, E^t, Q^{t-1} = q^{t-1}, \boldsymbol{X}^t, \boldsymbol{\alpha}] \tag{6}$$

[11] The first line of the unconfoundedness assumption in FJ translates into our notation as follows: $\Pr[E_t(q) = 1 \,|\, E^{t-1}(q) = \mathbf{0}, Q^{t-1} = q^{t-1}, \boldsymbol{X}_1] = \Pr[E_t(q') \,|\, E^{t-1}(q') = \mathbf{0}, Q^{t-1} = (q')^{t-1}, \boldsymbol{X}_1]$ if $q^{t-1} = (q')^{t-1} = \mathbf{0}$. Analogous to our weak NAC, FJ's condition ensures that the period-by-period hazard rates for the nontreated consistently estimate the corresponding counterfactual hazard rates for the treated.

[12] Regarding future covariates, one could alternatively allow $\boldsymbol{X}_t$ to depend on intermediate outcomes as e.g. Lechner and Miquel (2010) do. This would imply, however, that one could not estimate the effect of choosing arbitrarily different treatment sequences for those who are in one sequence.

[13] Our weak SRC.1 in eq. (6) parallels the second line of the unconfoundedness assumption in FJ that reads, translated into our notation: $\Pr[Q_t = 1 \,|\, E(q), E^t = \mathbf{0}, Q^{t-1} = \mathbf{0}, \boldsymbol{X}_1] = \Pr[Q_t = 1 \,|\, E^t = \mathbf{0}, Q^{t-1} = \mathbf{0}, \boldsymbol{X}_1]$.

This way of stating the sequential randomization condition makes clear that the sequential randomization condition precludes that conditional on $(E^t, Q^{t-1} = q^{t-1}, \boldsymbol{X}^t, \boldsymbol{\alpha})$ future values of the time-varying covariates $\boldsymbol{X}_r$, $r > t$, influence Q_t because future potential outcomes may depend upon $\boldsymbol{X}_r$.

The third condition is a support condition guaranteeing that, whatever the history of observed and unobserved covariates and the treatment sequence so far, the probability of starting or continuing training lies strictly between zero and one (conditional upon being nonemployed and that a training program has not yet been completed). Formally, this means:

Support Condition 1 (SC.1):

(7) $$0 < \Pr(Q_t = 1 \,|\, \mathcal{J}_t) < 1\,,$$

if $E_t = 0$ and $Q^{t-1} = \mathbf{0}$ (treatment has not started yet) or $Q^{t-1} = (0, \ldots, 0, 1, \ldots, 1)$ (treatment has started but not ended).

Since $E_t = 0$ is a necessary condition for $Q_t = 1$, the support condition implies that the probability to be nonemployed, $\Pr(E_t = 0 \,|\, \mathcal{I}_t)$, is larger than zero. Together these two conditions represent a special case of condition (3.3) in RO that he calls the identifiable treatment sequences. The support condition further implies that, if two treatment sequences q and q' coincide up to period $t-1$, i.e. $q^{t-1} = (q')^{t-1}$, and $q_t = 1$ and $q'_t = 0$, then there is a positive probability to start the treatment in t or to continue the treatment in t conditional on all other variables.[14]

Under the three conditions NAC, SRC.1, and SC.1, we can identify the one-period-ahead causal effect, $\Pr[E_{t+1}(q) = 1 \,|\, \mathcal{I}_{t+1}] - \Pr[E_{t+1}(q') = 1 \,|\, \mathcal{I}_{t+1}]$, for a given employment history $E_t = e_t$ and two treatment sequences q and q' with $q^{t-1} = (q')^{t-1}$ and $Q_t = 1$ while $Q'_t = 0$. Specifically, under the NAC and the SRC.1 we have that:

$$\Pr[E_{t+1}(q) = 1 \,|\, \mathcal{I}_{t+1}] = \Pr(E_{t+1} = 1 \,|\, E^{t-1} = e^{t-1}, E_t = 0, Q^{t-1} = q^{t-1}, Q_t = 1, \boldsymbol{X}^{t+1}, \boldsymbol{\alpha})$$

and

$$\Pr[E_{t+1}(q') = 1 \,|\, \mathcal{I}'_{t+1}] = \Pr(E_{t+1} = 1 \,|\, E^{t-1} = e^{t-1}, E_t = 0, Q^{t-1} = q^{t-1}, Q_t = 0, \boldsymbol{X}^{t+1}, \boldsymbol{\alpha})\,.$$

To show that the potential employment probabilities on the left hand side equal the actual employment probabilities on the right hand side, note that the NAC implies

[14] Furthermore, here and in the following we assume full overlap in the support of the time varying covariates $\boldsymbol{X}^t$ between individuals with different treatment sequences (see HN for similar considerations). Similarly, FJ require full overlap in the support of the time constant covariates (see also footnote 11).

that $E^{t-1}(q) = E^{t-1}(q') = E^{t-1}$ for fixed Q^{t-1} with $Q^{t-1} = q^{t-1} = (q')^{t-1}$. Then, SRC.1 allows one to exchange the conditioning on $Q_t = 1$ with $Q_t = 0$. Thus, the conditional contrast

$$\Pr(E_{t+1} = 1 \,|\, E^{t-1} = e^{t-1}, E_t = 0, Q^{t-1} = q^{t-1}, Q_t = 1, \boldsymbol{X}^{t+1}, \boldsymbol{\alpha}) - \\ \Pr(E_{t+1} = 1 \,|\, E^{t-1} = e^{t-1}, E_t = 0, Q^{t-1} = q^{t-1}, Q_t = 0, \boldsymbol{X}^{t+1}, \boldsymbol{\alpha})$$

identifies a conditional causal effect. This effect could be estimated from a sample, if SC.1 holds true and the distribution of $\boldsymbol{\alpha}$ is known.

Next, we can use the result for the one-period-ahead counterfactual employment probability to identify the joint probability of a counterfactual employment trajectory extending multiple periods ahead. The g-computation formula in RO (RO Theorem 3.1, AH equation (3.2)) shows how to reconstruct recursively the counterfactual probabilities for all future periods. Specifically, for $(q')^{t-1} = q^{t-1}$, we can recover the probabilities $\Pr(E_{t+k}(q') = 1 \,|\, E^{t+k-1} = e^{t+k-1}, Q = q, \boldsymbol{X}^{t+k}, \boldsymbol{\alpha})$, for $k = 2, 3, \ldots$, based on $\Pr(E_{t+k} = 1 \,|\, E^{t+k-1} = e^{t+k-1}, Q^{t+k-1} = (q')^{t+k-1}, \boldsymbol{X}^{t+k}, \boldsymbol{\alpha})$, which are identified based on individuals with observed treatment sequence $(q')^{t+k-1}$ and employment sequence e^{t+k-1} up to period $t+k-1$. Then, the joint probability of a sequence of potential employment states under the counterfactual treatment q', $E_{t+1}(q'), \ldots, E_{t+k}(q')$, can be reconstructed as follows:

$$\Pr(E_{t+1}(q') = e_{t+1}, \ldots, E_{t+k}(q') = e_{t+k} \,|\, E^t = e^t, Q = q, \boldsymbol{X}^{t+k}, \boldsymbol{\alpha}) \; = \\ \prod_{l=1}^{k} \Pr(E_{t+l} = e_{t+l} \,|\, E^{t+l-1} = e^{t+l-1}, Q^{t+l-1} = (q')^{t+l-1}, \boldsymbol{X}^{t+l}, \boldsymbol{\alpha}) \; .$$

Recall that the two treatment sequences q and q' have started to deviate in period t. Among those with $E^t = e^t$ and receiving treatment sequence q', we can recursively condition (starting in period $t+1$) on the employment path of interest up to the previous period and then identify the counterfactual transition probability of interest based on the implied subgroup. The product of the conditional one-period-ahead employment probabilities equals the joint probability of the employment trajectory from period $t+1$ to $t+k$.

Further, we can construct the unconditional (of E^{t+k-1}) counterfactual probability to be employed in period $t+k$, $K \geq 1$, by integrating over the set of possible employment histories, $\boldsymbol{\mathcal{E}}_{t+k-1}$. Formally, we can write this as:

$$\begin{aligned}
&\Pr[E_{t+k}(q') = 1 \,|\, Q = q, \boldsymbol{X}^{t+k}, \boldsymbol{\alpha}] \;= \\
&\sum_{\{j \,:\, e_j^{t+k-1} \in \boldsymbol{\mathcal{E}}_{t+k-1}\}} \Big\{ \Pr[E_{t+k}(q') = 1 \,|\, E^{t+k-1} = e_j^{t+k-1}, Q = q, \boldsymbol{X}^{t+k}, \boldsymbol{\alpha}] \\
&\qquad \cdot \Pr[E^{t+k-1} = e_j^{t+k-1} \,|\, Q = q, \boldsymbol{X}^{t+k}, \boldsymbol{\alpha}] \Big\} \;.
\end{aligned} \tag{8}$$

Again, equation (8) is a special case of equation (3.8) in Theorem 3.1 of RO. The support condition SC.1 guarantees for treatment sequence q that in large samples there will always be comparison individuals with the treatment sequence q'.

3.1.2 Contrasting Alternative Planned Durations

In order to model the causal effect of planned training duration on actual training status, we rely on the framework in the previous subsection which provides causal contrasts between different alternative treatment sequences. Define the start of treatment as $s \equiv \{t : Q^{t-1} = \mathbf{0}, Q_t = 1\}$ and $s \equiv \infty$ for untreated individuals. Conditional on the history so far and conditional on unobserved hetorogeneity, we assume that the planned duration P_s for program start in period $t = s$ is determined randomly in that month, i.e. $P_s = p$ is observed in s. Put differently, the planned duration may depend upon the time-varying covariates and the employment history up to that point of time, as well as upon the unobserved heterogeneity. We model the impact of planned duration on actual treatment sequences. Thus, we identify the causal effect of planned duration on employment outcomes through the impact of planned duration on actual treatment and the estimates of the causal impact of actual treatment sequences on employment derived in the previous subsection.

Our setup allows for an arbitrary dependence between P_s and the transitory error term $\varepsilon_{Q,s}$ in the period when the treatment sequence starts. After program start, P_s augments the history of covariates in subsequent time periods $t = s + \theta$ $(\theta \geq 1)$. Its impact on the actual treatment sequence is accounted for in $\psi_Q(E^t, Q^{t-1}, \boldsymbol{X}_t, P_s)$ for $t > s$, where – with slight abuse of notation – the set of arguments of $\psi_Q(.)$ in equation (1) and both information sets $\mathcal{I}_t, \mathcal{J}_t$ are augmented by P_s. Correspondingly, P_s is independent of future transitory error terms $\varepsilon_{Q,t}$.

Analogous to the previous subsection, we assume a sequential randomization condition and a support condition for the planned duration.

Sequential Randomization Condition 2 (SRC.2):

(9) $$P_s \perp\!\!\!\perp [E_{s+1}(q), ..., E_T(q)] \mid \mathcal{J}_s \text{ if } Q^{s-1} = \mathbf{0} \text{ and } Q_s = 1\,,$$

and

(10) $$P_s \perp\!\!\!\perp (\boldsymbol{X}_{s+1}, \ldots, \boldsymbol{X}_T) \mid \mathcal{J}_s \text{ if } Q^{s-1} = \mathbf{0} \text{ and } Q_s = 1\,.$$

The first part of this condition holds in particular for the actual treatment sequence $Q = q$ observed with $q^{s-1} = \mathbf{0}$ and $q_s = 1$. This implies that planned duration does not affect employment beyond its impact on the actual treatment sequence.[15]

For a causal contrast between two different values the random variable P_s may take, say p and p', we assume the following support condition.

Support Condition 2 (SC.2):

(11) $$0 < \Pr(P_s = \tilde{p} \mid \mathcal{J}_s) < 1\,, \tilde{p} = p, p'\,, \text{ if } Q^{s-1} = \mathbf{0}, Q_s = 1\,.$$

The sequential randomization condition SRC.2 ensures that we identify the causal effect of different planned durations on future employment by contrasting individuals with the same history, the same observed covariates, and the same individual specific effects. The support condition SC.2 ensures that we can observe similar individuals with different planned durations. The identifying conditions in this subsection assume that it suffices to condition on the individual specific effects in the employment equation and the treatment equation to account for selection on time-invariant unobservables.

3.1.3 Identification of Unobservables

So far we have assumed knowledge of the joint distribution of $(\alpha_E, \alpha_Q) = \boldsymbol{\alpha}$ to establish identification of the causal effects for the treated both for a certain training sequence against no training at all and for a change in planned duration of training given training has started. To complete the dynamic evaluation framework, we now discuss how the joint distribution of the unobservables can be identified nonparametrically given data on $\{E_t, Q_t, \boldsymbol{X}_t\}$. The discussion in HN concerning

[15]Note that we cannot use planned duration as an instrument to control for the endogeneity of the actual treatment sequence. This is because planned duration is observed only when a treatment sequence has started and remains constant afterwards.

Theorem 3 provides a precise account for a bivariate dynamic discrete choice model like ours. The argument goes as follows. Denote by $\nu_{j,t} \equiv \alpha_j + \varepsilon_{j,t}$, $j = E, Q$, the composite error term. Starting with the first period, $t = 1$, vary the employment index, $\psi_E(\boldsymbol{X}_1)$, over its full support and trace out the marginal distribution of the employment unobservable, $\nu_{E,1}$.[16] Next, fix the employment index such that the probability to be nonemployed in period 1 approaches one. In this limit set, one can identify the marginal distribution of the unobservable in the training equation, $\nu_{Q,1}$, by varying the training index, $\psi_Q(E_1, \boldsymbol{X}_1)$, over its full support. Jointly varying the employment and training indices of the first period allows one to trace out the joint distribution of the first period unobservables. Then, move on to the next period and proceed in the same sequential way. Specifically, to trace out the marginal distribution of the unobservable in the second period employment equation, $\nu_{E,2}$, fix the first period training and employment indices to obtain an appropriate limit set and then vary the second period employment index, $\psi_E(E_1, Q_1, \boldsymbol{X}_2)$, over its full support. Continue the sequential procedure until the last period.

This argument makes clear that nonparametric identification relies on the ability to vary the index functions of the observed covariates and lagged employment and training states independently across equations and time periods over their full support. Thus, a fully nonparametric analysis requires at least two regressors, that vary sufficiently strongly across time, where the current value of one of them is excluded from one of the two index functions. Put formally, $\text{supp}(\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T) = \text{supp}(\boldsymbol{X}_1) \times \ldots \times \text{supp}(\boldsymbol{X}_T)$, where supp() denotes the support, and there exist subvectors $\boldsymbol{X}_{E,t} \subset \boldsymbol{X}_t$ and $\boldsymbol{X}_{Q,t} \subseteq \boldsymbol{X}_t$ such that $\boldsymbol{X}_{E,t} \neq \boldsymbol{X}_{Q,t}$ and $\text{supp}(\psi_E(E^{t-1}, Q^{t-1}, \boldsymbol{X}_{E,t}), \psi_Q(E^t, Q^{t-1}, \boldsymbol{X}_{Q,t})) = \text{supp}(\psi_E(E^{t-1}, Q^{t-1}, \boldsymbol{X}_{E,t})) \times \text{supp}(\psi_Q(E^t, Q^{t-1}, \boldsymbol{X}_{Q,t}))$ for all t.[17] While we have some equation-specific time-varying variables (especially information on local labor market conditions and local supply of training programs), the requirements for nonparametric identification seem too strong in our application that involves 50 time periods.[18] Therefore, we impose the permanent-transitory structure on $\nu_{E,t}$

[16] It is assumed here that the support of the index, $\text{supp}(\psi_j(E^{t-1}, Q^{t-1}, \boldsymbol{X}_t))$, $j = E, Q$, contains the support of the respective composite error, i.e. $\text{supp}(\nu_{j,t}) \subseteq \text{supp}(\psi_j(E^{t-1}, Q^{t-1}, \boldsymbol{X}_t))$.

[17] The requirements regarding the variation across equations and across time can be relaxed, if one is willing to impose further structure on how the index functions depend on observed variables, see HN for details.

[18] Further time-varying regressors that we include are the remaining entitlement to unemployment benefits, remaining planned program duration, season of the year and number of months since the

and $\nu_{Q,t}$, where the dependence across time and across equations is generated by the time-constant α_E, α_Q.[19] With this simplification, we could identify the joint distribution of the unobservables semiparametrically even with just two periods of data.

3.2 Treatment Effects of Interest

In our empirical analysis, we estimate and report the following two treatment parameters, which are identified based on the discussion above. Our first estimand of interest is the average effect of treatment on the treated (ATT), where the treatment is participating in a training program, i.e. $Q = (Q_1, \ldots, Q_T) \neq \mathbf{0}$, characterized by a planned duration $P_s = p$ as observed (recall that s defines the start of treatment). The alternative is not participating in a training program during the observation period, i.e. $Q = \mathbf{0}$ and P_s is not observed ($s \equiv \infty$). This corresponds to the effect of training incidence integrated over the observed distribution of planned durations. We report treatment effects aligned by the time since program start, which we denote by $\theta \equiv \{t - s : t \geq s\}$.

Denote by $\Delta_\theta^C(q)$ the average effect of training against no training for those who undergo training sequence q in period $\theta = 1, 2, \ldots$ since program start:[20]

$$\Delta_\theta^C(q) \equiv \mathbb{E}[E_{s+\theta}(q) - E_{s+\theta}(0) \,|\, Q = q] \,.$$

If we take the expectation of $\Delta_\theta^C(q)$ with respect to the observed distribution of a random training sequence $Q \neq \mathbf{0}$ (this defines the treated within our observation window), we obtain the first treatment effect of interest as ($\mathbb{E}_Q$ defines the expectation among the treated)

$$\Delta_\theta^C = \mathbb{E}_Q[\Delta_\theta^C(Q)] \,. \tag{12}$$

As a second treatment parameter we study the causal effect of assigning different

inflow.

[19]Dynamic evaluation approaches in continuous-time, e.g. Abbring and van den Berg (2003) and Osikominu (2013), rely on a similar assumption.

[20]Note that $\mathbb{E}[E_t(q) - E_t(0) \,|\, Q = q]$ is equal to zero for $t < s$ because of the no-anticipation assumption (NAC) and $\mathbb{E}[E_s(q) - E_s(0) \,|\, Q = q] = 0$ by the assumption regarding the timing of events within a period that employment is determined first and then if the person remains unemployed then it is decided whether treatment starts.

planned program durations to those who receive treatment.[21] At the start of training, participants are assigned a planned program duration, P_s, that remains constant thereafter. It may differ from the realized training duration because participants may drop out or prolong participation.

We can construct the counterfactual probability to continue with training when a different planned training duration is assigned, provided that there is overlap in the support of p and p' as described above. In the period in which training starts, the observed realized training status and the counterfactual realized training status are both equal to one. From the period following the program start onwards, the counterfactual probability to continue with training is obtained recursively, conditioning upon continuation of training until the previous period. Denote by $\Delta_\theta^P(p, p', s)$ the average effect of assignment to a program with planned duration p against assignment to a program with planned duration p' for those with program start in period s (assuming $q^{s-1}(p) = (q')^{s-1}(p') = 0$ and $q_s(p) = (q')_s(p') = 1$), which is given by

$$\Delta_\theta^P(p, p', s) \equiv \mathbb{E}[E_{s+\theta}(q(p)) - E_{s+\theta}(q'(p')) \,|\, Q^{s-1} = \mathbf{0}, Q_s = 1] \ .$$

Thus, our second parameter of interest for the effect of assigning planned duration p versus p' is:

$$\Delta_\theta^P(p, p') = \mathbb{E}_S[\Delta_\theta^P(p, p', S)] \tag{13}$$

where the expectation is with respect to the observed distribution of random training starts S among the treated $Q \neq \mathbf{0}$.

4 Institutional Background and Empirical Implementation

4.1 Training in Germany

Training schemes have traditionally dominated active labor market policy in Germany. Legislation distinguishes three main types of training, further training (*Berufliche Weiterbildung*), retraining (*Berufliche Weiterbildung mit Abschluss in einem*

[21] While our evaluation framework would also allow us to estimate the causal effect of varying the realized duration of training, we focus on the planned duration because the latter is controlled by caseworkers and policymakers.

anerkannten Ausbildungsberuf), and short-term training (*Trainingsmaßnahmen und Maßnahmen der Eignungsfeststellung*). Figure 1 shows the evolution of entries into the three different training programs in West and East Germany during the period 1999 to 2007. Until 2000, enrollment into further training (henceforth also referred to as long-term training) was around 260,000 in West Germany and 170,000 in East Germany. A policy reorientation favoring programs supposed to activate the unemployed in the short run led to a decline in further training and retraining and a sharp increase in short-term training. In 2004, participation in further training was about 100,000 in West Germany and about 50,000 in East Germany. The corresponding figures for short-term training were 800,000 and 400,000, respectively, up from around 200,000 in 1999. After a low in 2005, participation recovered somewhat in 2006 and 2007.

— Insert figure 1 about here. —

The main goal of active labor market policy in Germany is to reintegrate unemployed individuals into employment. In this study we focus on further training programs. They are used to adjust the skills of the unemployed to changing requirements of the labor market and possibly to changed individual conditions of employability (due to health problems for example). Further training courses typically last several months to one year and are usually conducted as full-time programs. Teaching takes place in class rooms or on the job in training firms. The course curriculum may also include internships. Typical examples of further training schemes are courses on IT based accounting or on customer orientation and sales approach. Similar to the much longer retraining schemes, that lead to a complete new degree within the German apprenticeship system, further training programs aim at improving the human capital and productivity of the participant. Short-term training, in contrast, primarily aims at improving job search and lasts typically about four weeks.

In order to become eligible for training, job seekers have to register personally at the local employment agency. This involves a counseling interview with a caseworker. In principle, they have in addition to fulfill a minimum work requirement and be entitled to unemployment benefits. However, there are exceptions to this rule. The most important criterion is that the training scheme has to be considered necessary by the caseworker for the unemployed to find a new job. Participation in training can occur at any time during an unemployment spell.

Until 2003, training measures were assigned by the caseworker. This was often done in agreement with the job seeker, considering his or her willingness to receive training and to work in a specific field. The final decision was subject to the discretion of the caseworker. Assignment into programs was to a large extent driven by the supply of courses that were booked in advance for a year by the employment agencies from training providers. Assignments to training often occurred at very short notice in order to fill course capacities and to keep up job search incentives (Schneider et al., 2006).[22]

During training most participants receive a subsistence allowance of the same amount as the unemployment compensation they would receive otherwise. Participants not eligible for subsistence allowance may receive similar payments from the European Social Fund. In addition, travel and child-care costs may be covered by the employment agency.

Once a particular program has been assigned, participation is mandatory. Noncompliance is in general sanctioned with a temporary suspension of unemployment compensation. The planned duration of the further training programs considered in this paper is eight months on average. However, not all participants who start a program complete it. In fact, according to Paul (2014), one out of five participants who have started a program and attended it for at least one week drop out before having reached 80% of the planned duration. About half of the dropouts start employment soon after quitting a program. In many cases this behavior is encouraged by the employment agency because in general employment has priority over participation in active labor market programs. Exceptions from this rule are possible if completing the program is deemed necessary for a stable placement. Those dropping out for other reasons are often not sanctioned. As opposed to dropouts, it also happens in some cases that participation in training is prolonged. Due to dropout and possible prolongment of participation the actual duration of training is endogenously

[22] In 2003, the assignment procedure changed to a system in which the job seeker receives a voucher that specifies the length, content, and objective of the training program. The job seeker is then supposed to choose a suitable course from a pool of certified providers. The 2003 reform intended to improve the targeting of training programs. However, potential participants continued to be uncertain about the actual start of a course because it turned out that training providers tended to collect vouchers until a critical number of participants was reached or they shortly canceled scheduled courses if there were too few participants (Kühnlein and Klein, 2003, Schneider et al., 2006). Moreover, during the first quarter of 2003, the old and new assignment system coexisted. 93% of the programs in our analysis sample start before 2003. An additional 2% starts in the first quarter of 2003.

determined.

4.2 Constructing a Panel Data Set

For the empirical analysis, we construct a panel data set from a rich administrative database, the Integrated Employment Biographies Sample (IEBS). The IEBS is a 2.2% random sample from a merged data file containing individual data records collected in four different administrative processes: the IAB Employment History (*Beschäftigten-Historik*), the IAB Benefit Recipient History (*Leistungsempfänger-Historik*), the Data on Job Search Originating from the Applicants Pool Database (*Bewerberangebot*), and the Participants-in-Measures Data (*Maßnahme-Teilnehmer-Gesamtdatenbank*). The data contain detailed daily information on employment subject to social security contributions, receipt of transfer payments during unemployment, job search, and participation in different active labor market programs.[23]

We consider an inflow sample into unemployment consisting of individuals who became unemployed between the first of July 1999 and the end of December 2000, after having been continuously employed for at least 125 days. Entering unemployment is defined as the transition from non-subsidized employment to non-employment plus subsequently (not necessarily immediately) some contact with the employment agency, either through benefit receipt, program participation, or a job search spell. In order to exclude individuals eligible for specific labor market programs targeted to youths and individuals eligible for early retirement schemes, we only consider persons aged between 25 and 53 years at the beginning of their unemployment spell.

We aggregate the spell information in the original data into calendar months. We follow a person in the sample from the month of his or her first inflow into unemployment over the next 49 months or until the end of 2004, whichever occurs first. For 72% of the individuals in the sample we observe the full sequence of 50 months. The sequences of the remaining individuals are shorter either because we observe less than 50 months from their inflow until the end of 2004, or because we censor the time path of individuals when they enter a long-term active labor market program other than training. We ignore participation in short-term training and do not censor employment sequences in this case.

[23]For further information on the data see Appendix A.

We distinguish the two outcome states non-subsidized employment (henceforth denoted as employment) and non-employment as alternative states. We aggregate the employment information measured at a daily level into months as follows. First, for short gaps of a length up to 15 days between sequences of longer employment or non-employment spells we extend the longer spells through the gap. Second, we map the start of non-employment and employment spells to the monthly employment dummy in the following way. If a transition to non-employment occurs during a calendar month, the employment dummy is set to zero during this month. It continues to equal zero in the following month if the elapsed duration of non-employment at the end of the month exceeds 30 days. From the third month of non-employment onwards, the employment dummy is set to zero if the share of days in non-employment exceeds one half. Third, we adjust our procedure in order to take account of short employment spells that otherwise would be dropped.

Participation in training is coded as follows. We construct a dummy variable that equals one in the month in which the job seeker starts a training program and attends it for at least 27 days. In order to model the duration of the training program we apply the same rules as for the employment dummy above to the qualification dummy. Because not only the start of a program but also the program status in each following month is used for the estimation, it is important to use reliable information on the realized program duration. We correct the reported end dates of training programs using the correction procedures proposed in Waller (2008). Participation can already occur in the first month we observe for an individual.

The definition of the monthly employment and training dummy variables mimics the timing of events. We assume that within a given month the employment status is determined before the training status. When a person is non-employed in a given month he/she may start a training program in the same month. When a program participant exits to employment in a given month, even though he has been in the program at the beginning of this month, the training dummy changes to zero in that month. Consequently, our empirical analysis imposes a lag in the causal effect of training, such that training in month t is only allowed to affect employment from month $t + 1$ onwards.

The panel data set is completed by adding personal, occupational and regional information. Time-varying information on regional labor market conditions is matched with a lag to the current calendar date. For instance, the local unemployment rate in calendar month $t-1$ is matched to calendar month t in our panel. Thus within a given month, the timing is such that covariate values are determined first, then the employment status and last the training status. The empirical analysis is carried out separately for males and females and West and East Germany.

4.3 Descriptive Analysis

Table 1 gives an overview of the four samples and their basic characteristics. On average we observe 39 to 45 months per person, with the number of non-employment months ranging from 22 to 26. This corresponds to 1.6 to 2.2 unemployment spells and 0.9 to 1.5 employment spells on average per person. One in ten to one in five persons participate in training throughout the observation period with participation rates being higher in East Germany and among females.

— Insert table 1 about here. —

Planned and realized enrollment lengths in training vary widely. Figure 2 provides histogram plots of planned and realized program durations in the four samples. The height of the bars records the fraction of cases with a program duration corresponding to the value given on the horizontal axis. Realized and in particular planned durations display spikes at certain round dates like six months or one year. The share of realized durations lying below half a year is higher than that of planned durations. This indicates that some trainees drop out before the scheduled program end.

— Insert figure 2 about here. —

Figure 3 illustrates the evolution of the employment and training rates from the month of inflow into unemployment onwards. In the calendar month of the inflow, all individuals are defined as non-employed. The employment rates subsequently recover, but those of females remain at a slightly lower level than those of males. While participation rates peak at five percent in West Germany, they peak at nine percent in East Germany.

— Insert figure 3 about here. —

Figure 4 gives a first impression of the likely order of magnitude of the treatment effects. It shows the actual employment rate and estimates of the counterfactual employment rate associated with starting a training program in a given month versus no training start until that month for the treated individuals, where treated and matched controls are only aligned in time. Treatment status is a time-varying variable. This means that training participants who enrol later are counted as controls for those who enrol in an earlier month. The matching is performed with respect to the calendar month of the first inflow and the elapsed unemployment duration in the current unemployment spell. No adjustments are made for other potential sources of selection bias. West German females show the largest employment differences 24 to 36 months after program start, which amount to more than 15 percentage points. The initial lock-in periods characterized by negative employment effects are substantially longer in East Germany than in West Germany.

— Insert figure 4 about here. —

4.4 Econometric Specification

We now describe the implementation and specification of our model (1). We specify a system of two probit equations with correlated random effects α_E, α_Q and period-specific error terms $\varepsilon_{E,t}, \varepsilon_{Q,t}$. According to our conditions for identification, $\varepsilon_{E,t}, \varepsilon_{Q,t}$ are independent over time and across equations. A further key assumption relates to the separability between the impact of observed variables (covariates as well as the lagged dependent variables) and the error terms. A nonzero correlation between the individual specific effects (α_E, α_Q) gives rise to a spurious dependence between training and employment status even if the treatment effect is zero. Our specification of the unobservables is similar to other dynamic treatment effect approaches, in particular those by Abbring and van den Berg (2003) and HN, that also assume separability between the effects of observed and unobserved model components and that a low dimensional set of time constant unobserved heterogeneity terms generates the dependence over time and across equations.[24] All error terms are assumed to be normally distributed.

[24]Unlike HN, we do not model α_E to be a function of treatment status.

Our specification of the impact of the history since the inflow into unemployment (E^{t-1}, Q^{t-1}) on (E_t, Q_t) accounts for heterogeneous effects across different employment and treatment sequences and over time, thus accounting for duration dependence and state dependance in a very flexible way. Furthermore, we have investigated a number of possible interaction effects between the covariates considered and the lagged dependent variables and the variables relating to time. In light of its very flexible structure, we view our model as being nonparametric in the influence of these variables. In this respect, our approach is similar to matching analyses that rely on a rich and flexible specification of the heterogeneity with respect to observables. In contrast to matching analyses, we allow for selection into and out of training based on unobservables.

Specifically, we would like to emphasize the following three aspects regarding the flexibility of our model. First, we stratify the data by gender and region (West and East Germany) and run separate estimations for each of the four strata. Second, many variables are indicators referring to different categories of finely coded discrete variables. As far as the data permit, we specify fully saturated models. Third, for the continuous regressors we use polynomials and various interactions with the discrete variables (in particular with E^{t-1}, Q^{t-1}) in our specification.

Next we provide further details on the specification of the employment and training equation.[25] Consider first the employment equation. In order to model the employment dynamics we introduce employment lags up to the order of 49 (i.e. $E_{t-1}, E_{t-2}, \ldots, E_{t-49}$) as explanatory variables for current employment status. A lagged variable only kicks in if the inflow into unemployment has not been too recent for the corresponding lag to be available, i.e. the jth lag kicks in if $t-j \geq 1$. We use separate lags for E_{t-1} to E_{t-6} and E_{t-12} but sum up the other lags into a linear spline with four segments. This way we account for the entire employment history since the inflow into unemployment, thus accounting for both state dependence and duration dependence in the most flexible way, based on our discrete time data. In addition, we control in a flexible way for the elapsed number of months an individual is in the panel, t, and the elapsed duration in the current employment or non-employment spell, denoted by $\tau_{E,t}$ separately by last months employment state. Furthermore, we include a vector of observed characteristics, $\boldsymbol{X}_{E,t}$, in the employment equation. In particular, we use information on education, age, previous occupation, part-time

[25] The details can be found in the table 2 results in Appendix C.

status and earnings in the previous jobs, number of months employed in the last three years before the inflow into unemployment, health, children, labor market characteristics of the residential municipality, season and year. We allow for a large number of interaction effects between lagged employment status, elapsed duration, and covariates. Finally, we add three dummy variables capturing the duration of the remaining claim on unemployment benefits, the residential county's unemployment rate in the previous year, and two variables accounting for the policy style of the labor office (the number of individuals in job creation schemes and the number of entries into training both in relation to the number of unemployed individuals in the previous year).

The employment equation further includes lagged training participation whose impact is modeled in a flexible way. The dummy variable Q_{t-1} indicates whether the individual attended a training program in the previous month. If this dummy equals one, lagged training is depicted by a dummy if participation so far has lasted two months $[Q_{t-1} = 1] \times [Q_{t-2} = 1] \times [Q_{t-3} = 0]$, three months $[Q_{t-1} = 1] \times [Q_{t-3} = 1] \times [Q_{t-4} = 0]$, four months $[Q_{t-1} = 1] \times [Q_{t-4} = 1] \times [Q_{t-5} = 0]$ and so forth. We distinguish explicitly the effects on the exit from nonemployment in the next month while being in training from the effect on future employment in subsequent months after training has ended. To do so, we add a dummy D_t indicating whether an individual has ever participated in training since the inflow month. For trainees who have already exited the program, i.e. individuals having received training before month $t-1$ for whom the dummy on participation in training in the last month Q_{t-1} is zero, D_t is equal to one. We account for the training history by interacting D_t with $\tau_{Q,t}$, which in this case indicates the completed duration, ${\tau_{Q,t}}^2$ as well as with variables capturing the number of months which have passed since the end of training. An interaction effect of D_t, $\tau_{Q,t}$, and the time since end is also added. Effect heterogeneity with regard to individual characteristics is accounted for by interacting D_t (and in addition $\tau_{Q,t}$) with some covariates. To distinguish between the effect of training in the past on finding a job or on keeping a job, all the variables reflecting training history are interacted with the previous month's employment status E_{t-1}.

Consider next the training equation modeling the transition into and out of training. It is estimated simultaneously with the employment equation if the individual is not employed in the respective month and has not yet left a training program. Since participation can only occur during non-employment the two equation system reduces to a single equation for observations for which the employment status is equal to one. Then, the treatment equation is switched off. The vector of observed regressors, $\boldsymbol{X}_{Q,t}$, includes variables driving the decision to enter and to stay in a program. The covariate vector contains a dummy indicating whether the individual was enrolled in training in the previous month, Q_{t-1}, a variable for the elapsed months in the program $\tau_{Q,t}$, and a polynomial of the time until the planned end (in case the planned duration is not yet exhausted). These variables are equal to zero if the individual has not yet started training ($Q_{t-1} = 0$). Furthermore, the vector of independent variables includes variables summarizing the current unemployment experience: a dummy variables indicating whether the current month is the inflow month, as well as whether a repeated transition from employment to non-employment has occurred, and a polynomial of the elapsed unemployment duration in months ($\tau_{E,t}$). Moreover, information on age, schooling, vocational training, the last job, number of months in employment during the last three years before the inflow, health, children, remaining entitlement to unemployment benefits, season, and year are incorporated. Some individual characteristics are interacted with the elapsed duration in training and whether the individual was in training in the previous month. Moreover, we add the unemployment rate in the county as well as an interaction of this unemployment rate with the elapsed duration in training. Finally, the independent variables include the stock of training participants (and of job creation schemes, respectively) in the previous month in the labor office divided by the number of unemployed, as well as the number of entries into training divided by the number of unemployed in the previous year.

4.5 MCMC Estimation

We estimate the bivariate random effects probit model for employment and training transitions using Bayesian Markov Chain Monte Carlo (MCMC) techniques.[26] The draws of the parameters along the MCMC iterations allow us to estimate the posterior distribution of the parameters and of functions thereof. From a classical perspective, the mean of the posterior distribution converges to the point estimator from a maximum likelihood estimation and the variance of the posterior distribution

[26]See Chib (2001) for a survey of MCMC techniques. See Chib and Hamilton (2002) and Chib and Jacobi (2007) for applications of MCMC methods for the estimation of treatment effects. These papers analyze binary treatments and allow for heterogeneous treatment effects in terms of unobservables.

converges to the asymptotic variance of the point estimator in a maximum likelihood estimation. Thus, the standard deviation of the draws can be interpreted as standard errors from the classical perspective (see Train, 2003, for an overview over important properties of MCMC estimators). To obtain a sample from the posterior distribution, we use the Gibbs sampler. To simplify the sampling from a complex joint distribution, the Gibbs sampler forms blocks of model parameters and samples recursively from the distribution of one block conditional on the current values of the remaining parameters. The resulting sequence of simulated parameters is a Markov Chain whose invariant distribution is the desired posterior distribution. After convergence, the draws are samples from this posterior distribution.

MCMC estimation of probit models augments the data by simulating the continuous latent dependent variable as one step of the Gibbs sampler in each iteration. Then a standard linear regression on the observed covariates is performed to update the coefficients on the observed covariates (Albert and Chib, 1993). In a further step, the random effects (α_E, α_Q) are sampled conditional on the data and the remaining model parameters (Zeger and Karim, 1991). In this way, we obtain the posterior distribution of all model parameters as well as that of the individual specific effects. In particular, we can compute the expected value of the random effects for a given individual given the data and calculate the correlation between the unobservables in the training and employment equations.

We specify the following prior densities. For the coefficients involving the treatment effects parameters, we take diffuse, independent normal priors centered around prior expected values of zero. The prior for the variance of the random effects is taken to be an Inverse Wishart distribution with the scale matrix set to the individual level variances, obtained by separately estimating the two equations by Maximum Likelihood, on the diagonal and zeroes off-diagonal. The degrees of freedom parameter is chosen to be small such that the prior is diffuse.

The results reported below are obtained from running 50,000 iterations of the algorithm. We monitor convergence to the posterior distribution of the estimated parameters by investigating the sequence of the parameters and of the implied value of the likelihood function conditional on the values of the random effects along the MCMC iterations. Furthermore, we compare the means at different stages of the chains. Based on this evidence, we decided to discard the first 5,000 iterations as the burn-in phase, because the sequence after the 5,000 iterations appears compatible with draws from a time-invariant distribution. Therefore, our results are based on 45,000 draws. We implement the Gibbs sampler in Stata and Mata. The algorithm is described in more details in Appendix B.

4.6 Estimation of Treatment Effects

The raw coefficient estimates are difficult to interpret because of the complex dynamic structure of the model involving many interaction effects. Therefore, we directly analyze the posterior distribution of two treatment effects of interest defined formally in Section 3.2. These are the *Classical ATT* comparing training against no training (Δ^C, eq. (12)) and the *ATT of Changing Planned Training Duration* ($\Delta^P(p, p')$, eq. (13)), both evaluated in the sample of training participants. For $\Delta^P(p, p')$, we consider planned training durations of three, nine and twelve months and compare them to a planned duration of six months, about the median in the data. To estimate the treatment effects, we simulate 1,500 draws from their posterior distribution based on the parameter values from every 30th MCMC iteration after the burn-in phase. Then we obtain the point estimates of our treatment effects by taking the mean across the 1,500 draws. We use the standard deviation across draws to construct confidence intervals. We also investigate the sensitivity of our inference to an alternative method to assess estimation uncertainty. For this purpose, we implement an adapted version of the weighted confidence interval (WCI) bootstrap approach suggested recently by Ham and Woutersen (2013).[27]

A major advantage of the MCMC technique is that it provides for each individual an estimate of the posterior distribution of the individual specific effects. This allows us to account explicitly for the dynamic selection on unobservables when calculating the posterior distribution of the treatment effects of interest. EHL and Ham et al. (2010) among others also estimate treatment effects based on a nonlinear dynamic panel or hazard rate model and account for the estimated distribution of unobserved heterogeneity. The key difference of our approach compared with the existing litera-

[27] Ham and Woutersen (2013) emphasize that confidence intervals for an estimand that is a function of the original parameters may not have full coverage if the function is nondifferentiable (as in our case of simulating zero-one outcomes), has zero or unbounded derivatives. They suggest to use a confidence set around the original parameter estimates, e.g. based on a Mahalanobis distance and a linear approximation of the function. Then, the full range of the image of this confidence set provides a confidence interval of the estimand of interest.

ture is that the other studies average across the estimated unconditional distribution of unobserved heterogeneity. In contrast, we use the posterior distribution of the unobserved heterogeneity of the treated who are, in terms of their unobservables, a dynamically selected subset of the initial population.[28]

To simulate the posterior distribution of the *Classical ATT* we perform the following steps for each of the $1,500$ MCMC iterations:

Step 1. For each participant i, we use the observed training start s_i to simulate the sequence of actual employment outcomes, $\{E(q_i)\}$. Beginning with the first period after program participation, we predict the realized employment status based on the respective vector of covariates, the respective draws from the vector of coefficients attached to the covariates, the individual specific effects ($\alpha_{E,i}$) and draws of the idiosyncratic error term ($\varepsilon_{E,it}$) from a standard normal distribution. The dynamic covariates that involve lags of employment status are updated according to the predictions for the previous periods.

Step 2. For each participant, we simulate the sequence of counterfactual employment outcomes resulting if the participant did not participate in training at any time, $\{E(0)_{it}\}_t$, $t > s_i$. We begin with the first period after program start and predict the employment status for each period based on the same draws of the coefficient vector, the $\alpha_{E,i}$, and an $\varepsilon_{E,it}$ as in Step 1. We adapt the covariates to a situation with no training participation and update them while moving from one period to the next.

Step 3. To obtain a draw from the posterior distribution of the average treatment effect on the treated in period θ since training start, Δ_θ^C, we average the difference of the two employment outcomes from steps 1 and 2 over all treated individuals (N_1), i.e. $\frac{1}{N_1}\sum_{i=1}^{N_1}[E(q_i)_{i,\theta} - E(0)_{i,\theta}]$.

To simulate the posterior distribution of the *ATT of Changing Planned Training Duration* we proceed as follows. First, we simulate the employment and participation

[28]A further difference is that we focus on the probability to be employed in a given time period rather than on expected employment or unemployment durations as do Ham et al. (2010). In order to calculate treatment effects on expected employment or unemployment durations, one would have to make assumptions about how the exit rate from employment/unemployment evolves at large elapsed durations that are beyond the time horizon of the sample at hand. In contrast, we calculate the treatment effects on employment probabilities only for time periods within the support of our data.

status for the benchmark case in which all participants are assigned to a planned length of six months.[29] In the month in which a participant i starts the program, $t = s_i$, we have that $E(q(6))_{it} = 0$ and $Q(6)_{it} = 1$ by definition. From the next month onwards, $t \geq s_i+1$, $E(q(6))_{it}$ and $Q(6)_{it}$ are simulated in turn for each period. The covariates that involve lags of employment or participation status are adapted dynamically. Again, the values of the coefficients attached to observed covariates and the individual specific effects of the corresponding draw are used. The idiosyncratic error terms are drawn from independent standard normal distributions. In next step, we simulate training and employment states for the alternative scenarios in which the planned program duration is set to three months ($E(q(3))_{it}$ and $Q(3)_{it}$), nine months ($E(q(9))_{it}$ and $Q(9)_{it}$), and twelve months ($E(q(12))_{it}$ and $Q(12)_{it}$), respectively, in the same way, using the same draws for the idiosyncratic errors. Then we calculate the differential treatment effects for a planned duration of three, nine, and twelve months, respectively, against six months, i.e. $\frac{1}{N_1}\sum_{i=1}^{N_1}[E(q(\tilde{p}))_{i,\theta} - E(q(6))_{i,\theta}]$, with $\tilde{p} = 3, 9, 12$ and $\theta = 1, \ldots, 30$.

5 Estimation Results

We estimate the impact of incidence and duration of training on the transition probabilities between employment and unemployment using the MCMC estimation approach described in the previous section. Our empirical model accounts for selection into training based on observables and unobservables. Estimation is carried out separately for West German males, West German females, East German males, and East German females. The detailed estimation results are given in table 2 in Appendix C. The first column for each sample refers to the mean of the coefficients and the second to their standard deviation over MCMC iterations after the burn-in phase. We interpret them in an analogous way as the point estimates and standard errors of the coefficients obtained by a frequentist approach. Next, we briefly discuss the overall fit of the model and the individual level variances of the error terms. Because of the complexity of the model (it comprises 188 parameters), we refrain from further discussing single parameters. Rather, we assess the estimated model in general and discuss the results for different treatment effects of interest.

[29] In terms of the model specification, this means that the explanatory variables in the participation equation involving the planned end date (i.e. *months until planned end if enough duration left* and *months until planned end if enough duration left squared*) are adapted to this setting.

5.1 Model Fit and Selection on Unobservables

Evidence on the fit of the model is provided in figure 5 for the treated individuals from the start of the program onwards.[30] Actual and predicted employment rates of the trainees match closely in all four samples. Thus, our rich model specification does a good job in replicating the employment dynamics found in the data. This suggests that our model is not grossly misspecified.

— Insert figure 5 about here. —

Conceptually, our framework for identification of dynamic treatment effects assumes randomization conditional on unobserved heterogeneity. The last panel of table 2 in Appendix C provides an empirical account of the importance of selection on unobservables. It displays the estimated variances and covariances of the error terms of the employment and training equation. The share of the variance that is due to the individual specific effects varies between 19% and 31% for the employment equation and between 5% and 24% for the training equation. Thus, despite the richness of our data and the flexibility of the econometric model, it seems empirically important to take account of the dynamic selection on unobservables in addition to conditioning on employment and training histories. However, the correlation between the two random effects is small and insignificant for all four groups. This suggests that, given our rich data and flexible econometric model, there is no significant joint selection on unobservables that warrants a joint model for employment and training. Nevertheless, our results imply that it is important to account for unobserved heterogeneity in the employment equation as well for the employment and training history when estimating the causal effect on employment because eligibility for treatment hinges on not being employed. This finding is analogous to EHL.

5.2 Classical ATT

Figure 6 shows the average effect of training versus no training for participants on the probability to be employed in a given month.[31] More precisely, we compare the

[30] The exact numbers are given in table 3 in Appendix C along with information on the number of observations available in each month.

[31] The corresponding numbers are given in table 4 in Appendix C.

average of the actual employment outcomes of trainees with the expected counterfactual outcome obtained by setting the lags of training status in the employment equation to zero. We obtain actual as well as counterfactual outcomes through simulation using the estimated distribution of the model parameters from the sequence of MCMC iterations. In figure 6, the average difference in the monthly employment rates is depicted on the vertical axis, while months since program start are measured on the horizontal axis. The dashed lines around the estimated treatment effects are 95 percent confidence bands. The treatment effect for a particular month is statistically significant if the confidence band does not contain zero. A participation in training reduces the employment probability during the nine to 14 months after program start. During the first six months ($\theta = t - s \leq 6$) the employment probabilities of participants decline between nine (East German females) and 15 (East German males) percentage points (ppoints) compared to the situation of no participation. This lock-in effect peaks about three months later in the East German samples compared to the West German ones. Roughly one year after program start, the difference in employment rates turns positive and continues to increase until the end of the observation window. 30 months after program start ($\theta = t - s = 30$), West German females have a 14 ppoints higher employment probability than in the absence of training. The effects are somewhat smaller for East German females (10 ppoints) and East German males (11 ppoints) and much smaller for West German males (6 ppoints).

— Insert figure 6 about here. —

5.3 ATT of Changing Planned Training Duration

Next, we use our model estimates to analyze how treatment effects vary with the planned program duration. Typically, participation in training decreases the exit rates from unemployment between the start and the end date of a program compared to a situation of no participation (lock-in effect). The size of the reduction in job finding probabilities may change over the course of training. If job finding efforts increase towards the end of the program, the time until the scheduled end date has a negative effect on exits from unemployment. In a mechanical sense, a shorter planned enrollment length should therefore be associated with a shorter and less pronounced lock-in period. However, it is unclear whether such an advantage in the short run persists over time. If, by administering shorter programs, one could decrease the lock-in effect without reducing the long-run employment gains, policymakers would be advised to shorten planned durations.

Similarly as above, we simulate the training and employment histories of the subsample of training participants that result after fixing the planned program duration to a prespecified value. Specifically, we consider planned program durations of three, six, nine, and twelve months. We then evaluate the effect of participating in a program scheduled over three, nine, and twelve months, respectively, as opposed to six months, the median of planned duration. Tables 5 to 8 in Appendix C show the simulated participation and employment rates associated with different planned program durations. Note that the simulated realized program duration can be shorter or longer than the planned one. However, the tables suggest that there is a strong positive correlation between planned and realized program durations.

— Insert figure 7 about here. —

— Insert figure 8 about here. —

— Insert figure 9 about here. —

Figure 7 displays the treatment effects associated with a planned duration of three versus six months.[32] There are considerable gains of a shorter participation (up to five ppoints) but they are only transient. These gains reach their maximum in the sixth month ($\theta = t - s = 6$) after program start and about three months later they have already vanished. In the medium and long run, those attending programs with a scheduled length of six months fare better, exhibiting employment rates that are consistently higher by four to six ppoints. A similar pattern emerges when comparing programs with a scheduled length of nine and twelve months, respectively, with those planned to last six months, cf. figures 8 and 9.[33] Trainees attending longer programs are worse off around the time of the later scheduled end of their program but a few months later, they have consistently higher employment rates than compared to the benchmark case of a six-month program. Indeed, compared to a planned duration of

[32]The corresponding numbers are given in table 9 in Appendix C.

[33]The corresponding numbers are given in tables 10 and 11 in Appendix C.

six months, the employment rates associated with attending a nine-month program are three to five ppoints higher and those associated with attending a one-year program are five to eight ppoints higher.

5.4 Sensitivity Analyses

The large panel data set we use allows us to specify the dependence of training and employment probabilities on past employment and training states in a very flexible way. Furthermore, we account for potentially correlated unobserved heterogeneity in the employment and training equations. In this section, we investigate how the estimated classical treatment effects change if we use alternative more parsimonious specifications. First, we consider a pooled probit model for the employment equation that does not allow for selection on unobservables. Second, we consider a simple specification that models the dependence of the employment probability on lagged employment and training in a restrictive way. Third, we estimate a model that imposes both no selection on unobservables and no complex dynamic effects.

In the first sensitivity analysis, we use our benchmark specification of the employment equation but estimate a pooled probit model using standard Maximum Likelihood methods. This model does not account for unobserved heterogeneity. A simplified version of our simulation procedure is then employed to obtain the classical treatment effects. Table 12 in Appendix C provides the results of this exercise. Overall, the magnitude of the treatment effects differs from our benchmark estimates, but these differences are not big. For example, the long-run effects based on the pooled probit model are two to four ppoints lower than those obtained from our benchmark model. The most important difference occurs for males in West Germany: our benchmark specification suggests a positive treatment effect of 6 ppoints after 2.5 years while the pooled probit estimation leads to an effect of only 3 ppoints.

In the second sensitivity analysis, we use our benchmark two-equation model but apply a much less flexible specification of the employment dynamics and the treatment history in the employment equation. In this exercise the employment dynamics are modeled using only the first employment lag E_{t-1}, the elapsed duration in the current employment state $\tau_{E,t}$, and the time since inflow t. The treatment history is specified using only the dummy if the individual was participating in training in the previous month Q_{t-1} and *treat*. *treat* is a dummy variable set to 1 if at least one of the Q_{t-j}, $j > 1$ is 1, thus capturing incidence of training in the past. There are no changes in the qualification equation. Table 13 shows that most results are fairly similar to our benchmark treatment effects with the treatment effects for East German men involving the largest changes. While our benchmark results suggest a treatment effect of 11 ppoints at the end of our observation period, the simple specification suggests an effect of only 8 ppoints.

Finally, we combine the restrictions of the first two sensitivity analyses and estimate the employment equation with the simple specification using pooled probit. With this specification we find no positive treatment effects at all for males in West Germany and males and females in East Germany and much smaller effects than in our benchmark specification for females in West Germany (table 14). We conclude from this exercise that – in our application – it is crucial to either account for unobserved heterogeneity or to specify the employment and training dynamics in a flexible way. When both aspects are neglected, completely different results are obtained. When one of these aspects is neglected, some of the treatment effect estimates change but the overall picture is similar to a model considering both, unobserved heterogeneity and a flexible specification of the dynamics.

In a further sensitivity analysis, we investigate the sensitivity of our inference to an alternative method to assess estimation uncertainty. For the sample of East German females, we implement an adapted version of the weighted confidence interval (WCI) bootstrap approach suggested recently by Ham and Woutersen (2013). The idea of the WCI approach is to construct confidence intervals for a non-smooth function of the estimated parameters based on the closure of the image of the confidence interval of the weighted deviations. The confidence intervals obtained with the WCI approach prove to be much larger than those based on the MCMC estimates. However, the weighted confidence intervals lead to an extreme overcoverage, with a nominal 90% WCI covering almost the entire empirical distribution. Based on a smoothed version of the WCI ensuring correct empirical coverage, we obtain confidence intervals that are very similar to those based on the MCMC estimates.[34]

[34]The detailed results are available from the authors on request.

6 Conclusions

Training programs are an important part of active labor market policies of many advanced countries. Yet, their effectiveness is discussed controversial. One reason might be that labor market policy pursues heterogeneous, partly conflicting goals. Substantive skill development requires longer-term programs that may initially prolong unemployment. Thus, quick reintegration does not seem to be a viable goal of training that rather aim at integration into high quality jobs. A second key issue that complicates the evaluation of training programs is methodological. Standard statistical models for treatment evaluation are static. This paper examines the dynamic effects of training incidence and duration on labor market transitions in discrete time. Building on Robins (1997), we devise an evaluation framework in discrete time that takes the dynamics of program start and duration into account. First, we show how conditionally on lagged endogenous variables, time-varying observed covariates and time-constant unobserved heterogeneity causal effects can be identified under no-anticipation, sequential randomization, and support conditions. In a next step, we identify the distribution of the unobserved heterogeneity relying on results for bivariate dynamic discrete choice models, especially Heckman and Navarro (2007).

We estimate a dynamic random effects probit model including an employment and a participation equation based on large administrative data for Germany. The participation equation models the start of training as well as its end accounting for endogenous dropout. We control for selection on unobservables by allowing the random effects of both equations to be correlated. We account for time and duration dependence as well as for various forms of effect heterogeneity in a flexible way. Using Bayesian Markov Chain Monte Carlo (MCMC) methods, we estimate the posterior distribution of the model parameters, including the individual random effects. The analysis is implemented separately for West and East Germany and for males and females. We simulate different treatment effects of interest using the estimated distribution of the parameters and individual specific effects from the sequence of MCMC iterations.

Our findings imply positive effects of training on the employment probability emerging nine to twelve months after program start in all four subsamples considered. 30 months after program start, the effect of treatment on unconditional employment rates for the treated individuals lies between 6 and 14 percentage points. The effects are higher for women than for men and and the initial lock-in period is shorter in West Germany than in East Germany. Further, we use our estimates to analyze how training effects vary with planned training duration. Longer planned enrollment lengths of 9 and 12 months as opposed to just 6 months months lead to an increase in employment rates by 3-5 percentage points and 5-8 percentage points, respectively, 2.5 years after program start. This suggests that, on average, the higher costs of longer training programs translate into higher long-run employment gains.

References

Abbring, J.H. and J.J. Heckman (2007). "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation", in J.J. Heckman and E.E. Leamer (eds), *Handbook of Econometrics*, Vol. 6B. Amsterdam: Elsevier Science, ch. 72, 5145-5303.

Abbring, J.H. and G.J. van den Berg (2003). "The Nonparametric Identification of Treatment Effects in Duration Models", *Econometrica*, 71, 1491-1517.

Albert, J.H. and S. Chib (1993). "Bayesian Analysis of Binary and Polychotomous Response Data", *Journal of the American Statistical Association*, 88, 669-679.

Biewen, M., B. Fitzenberger, A. Osikominu, and M. Paul (2014). "The Effectiveness of Public Sponsored Training Revisited: The Importance of Data and Methodological Choices", *Journal of Labor Economics*, 32, forthcoming.

Buchinsky, M., D. Fougère, F. Kramarz, and R. Tchernis (2010). "Interfirm Mobility, Wages, and the Returns to Seniority and Experience in the U.S.", *Review of Economic Studies*, 77, 972-1001.

Bundesanstalt für Arbeit (2001). *Arbeitsstatistik 2000 – Jahreszahlen*, Nürnberg.

Bundesagentur für Arbeit (2006). *Arbeitsstatistik 2005 – Jahreszahlen*, Nürnberg.

Bundesagentur für Arbeit (2007, 2008). *Arbeitsmarkt 2006, 2007*, Nürnberg.

Card, David, Jochen Kluve, and Andrea Weber (2010). "Active Labour Market Policy Evaluations: A Meta-Analysis", *Economic Journal*, 120, F452-F477.

Chib, S. (2001). "Markov Chain Monte Carlo Methods: Computation and Inference", in J.J. Heckman and E.E. Leamer (eds.): *Handbook of Econometrics*, Vol. 5, Amsterdam: Elsevier Science, ch. 57, 3569-3649.

Chib, S. and B.H. Hamilton (2002). "Semiparametric Bayes Analysis of Longitudinal Data Treatment Models", *Journal of Econometrics*, 110, 67-89.

Chib, S. and L. Jacobi (2007). "Modeling and Calculating the Effect of Treatment at Baseline from Panel Outcomes", *Journal of Econometrics*, 140, 781-801.

Dyke, A., C.J. Heinrich, P.R. Mueser, K.R. Troske, and K.-S. Jeon (2006). "The Effects of Welfare-to-Work Program Activities on Labor Market Outcomes", *Journal of Labor Economics*, 24, 567-607.

Eberwein, C., J.C. Ham, and R.J. Lalonde (1997). "The Impact of Being Offered and Receiving Classroom Training on the Employment Histories of Disadvantaged Women: Evidence from Experimental Data", *Review of Economic Studies*, 64, 655-682.

Fredriksson, P. and P. Johansson (2008). "Dynamic Treatment Assignment: The Consequences for Evaluations Using Observational Data", *Journal of Business and Economic Statistics*, 26, 435-445.

Ham, J.C. and R.J. LaLonde (1996). "The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training", *Econometrica*, 64, 175-205.

Ham, J.C., Li, X. and L. Shore-Sheppard (2010). "Seam Bias, Multiple-State, Multiple-Spell Duration Models and the Employment Dynamics of Disadvantaged Women", *Unpublished Manuscript*, University of Maryland.

Ham, J.C. and T. Woutersen (2013). "Calculating Confidence Intervals for Continuous and Discontinuous Functions of Parameters", *Unpublished Manuscript*, University of Maryland.

Heckman, J.J., R.J. LaLonde, and J.A. Smith (1999). "The Economics and Econometrics of Active Labor Market Programs", in: O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, Vol. 3 A, Amsterdam: Elsevier Science, ch. 31, 1865-2097.

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

Horny, G., R. Mendes, and G.J. van den Berg (2012). "Job Durations with Worker and Firm Specific Effects: MCMC Estimation with Longitudinal Employer-Employee Data", *Journal of Business and Economic Statistics*, 30, 468-480.

Hotz, V. Joseph, Guido W. Imbens and Jacob A. Klerman (2006). "Evaluating the Differential Effects of Alternative Welfare-to-Work Training Components: A Reanalysis of the California GAIN Program." *Journal of Labor Economics*, 24, 521-566.

Kühnlein, G. and B. Klein (2003). "Bildungsgutscheine – mehr Eigenverantwortung, mehr Markt, mehr Effizienz? – Erfahrungen bei der Neuausrichtung der beruflichen Weiterbildung", Arbeitspapier Band 74, Hans-Böckler-Stiftung.

Lechner, M. and R. Miquel (2010). "Identification of the Effects of Dynamic Treatments by Sequential Conditional Independence Assumptions", *Empirical Economics*, 39, 111-137.

Martin, John P., and David Grubb (2001). "What Works for Whom: A Review of OECD Countries' Experiences with Active Labour Market Policies." *Swedish Economic Policy Review*, 8, 9-56.

Neyman, J. (1923). "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9." *Statistical Science*, 5(4): 465-472.

Osikominu, A. (2013). "Quick Job Entry or Long-Term Human Capital Development? The Dynamic Effects of Alternative Training Schemes", *Review of Economic Studies*, 80, 313-342.

Paul, M. (2014). "Many Dropouts? Never Mind! – Employment Prospects of Dropouts from Training Programs", *Annals of Economics and Statistics*, forthcoming.

Plaßmann, G. (2002). "Der Einfluss der Arbeitslosenversicherung auf die Arbeitslosigkeit in Deutschland." *Beiträge zur Arbeitsmarkt- und Berufsforschung, BeitrAB* 255, Nürnberg.

Robins, J. (1986). "A new Approach to Causal Inference in Mortality Studies with Sustained Exposure Periods: Application to Control of the Healthy Worker Survivor Effect", *Mathematical Modelling*, 7, 1393-1512.

Robins, J. (1997). "Causal Inference from Complex Longitudinal Data", in M. Berkane (ed.), *Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics*, 120, 69-117.

Roy, A.D. (1951). "Some Thoughts on the Distribution of Earnings", *Oxford Economic Papers*, 3, 135-146.

Rubin, D.B. (1974). "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 66, 688-701.

Schneider, H., K. Brenke, D. Hess, L. Kaiser, J. Steinwedel and A. Uhlendorff (2006). "Evaluation der Maßnahmen zur Umsetzung der Vorschläge der Hartz-Kommission - Modul 1b: Förderung beruflicher Weiterbildung und Transferleistungen", IZA Research Report No. 7, Bonn.

Sianesi, B. (2004). "An Evaluation of the Swedish System of Active Labor Market Programs in the 1990s", *Review of Economics and Statistics*, 86, 133-155.

Train, K. (2003). *Discrete Choice Methods with Simulation*, Cambridge: Cambridge University Press.

Troske, K.R., and A. Voicu (2010). "Joint Estimation of Sequential Labor Force and Fertility Decisions Using Markov Chain Monte Carlo Techniques", *Labour Economics*, 17, 150-169.

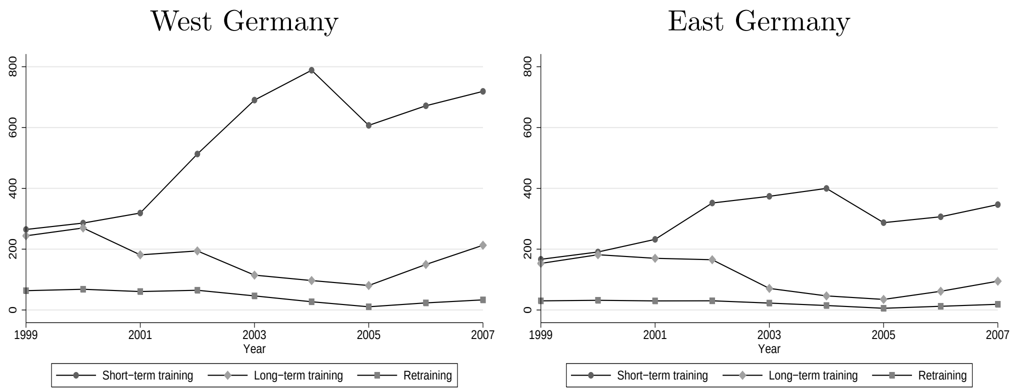
Waller, M. (2008). "On the Importance of Correcting Reported End Dates of Labor Market Programs", *Journal of Applied Social Science Studies*, 128, 213-236.

Zeger, S. and M. Karim (1991). "Generalized linear models with random effects: a Gibbs sampling approach", *Journal of the American Statistical Association*, 86, 79-86.

Zimmermann, R., S. Kaimer, D. Oberschachtsiek (2007). "Dokumentation des Scientific Use Files der Integrierten Erwerbsbiographien (IEBS-SUF V1) Version 1.0." FDZ Datenreport 1/2007, Institut für Arbeitsmarkt- und Berufsforschung (IAB), Nürnberg.

Figures

Figure 1: Entries into Training Programs in West and East Germany (in 1000)



Source: Bundesagentur für Arbeit (2001, 2006, 2007, 2008); own calculations.

Figure 2: Planned and Realized Training Durations

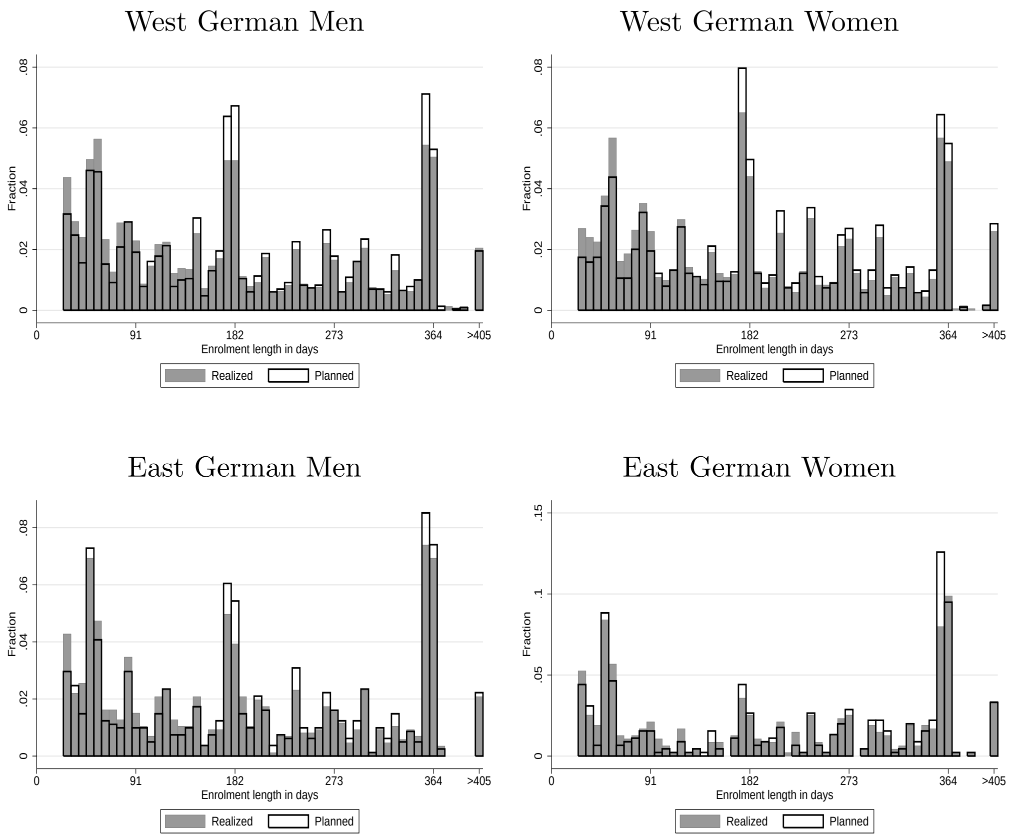


Figure 3: Employment and Participation Rates over Time

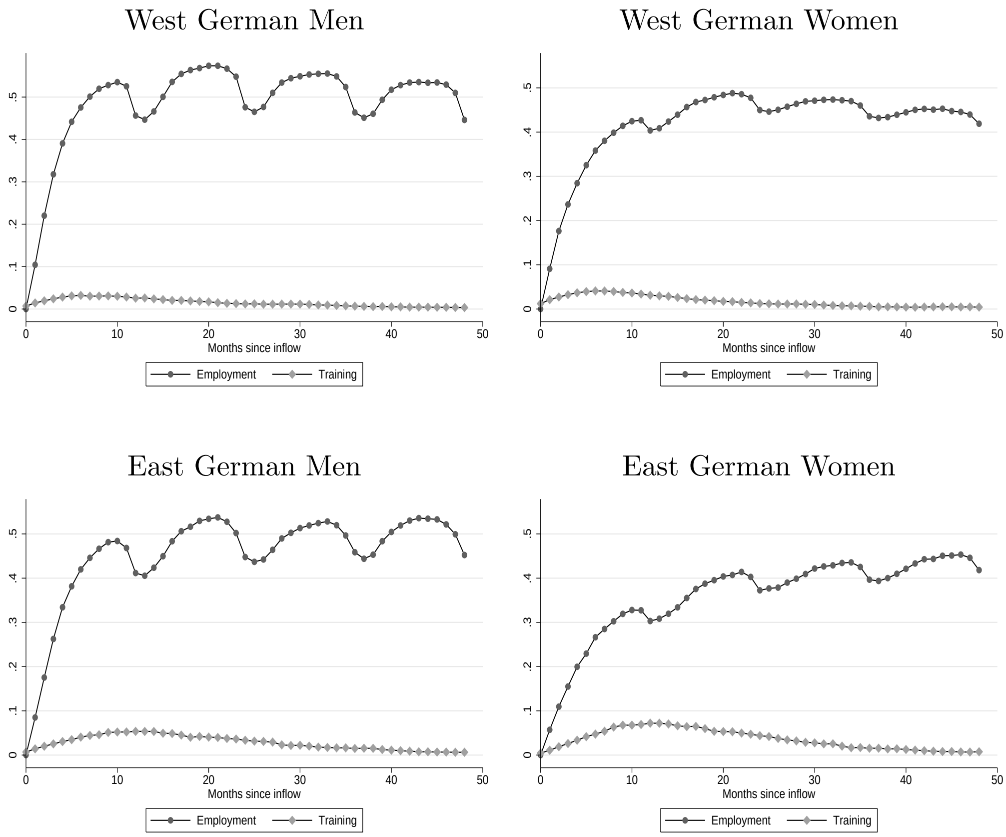
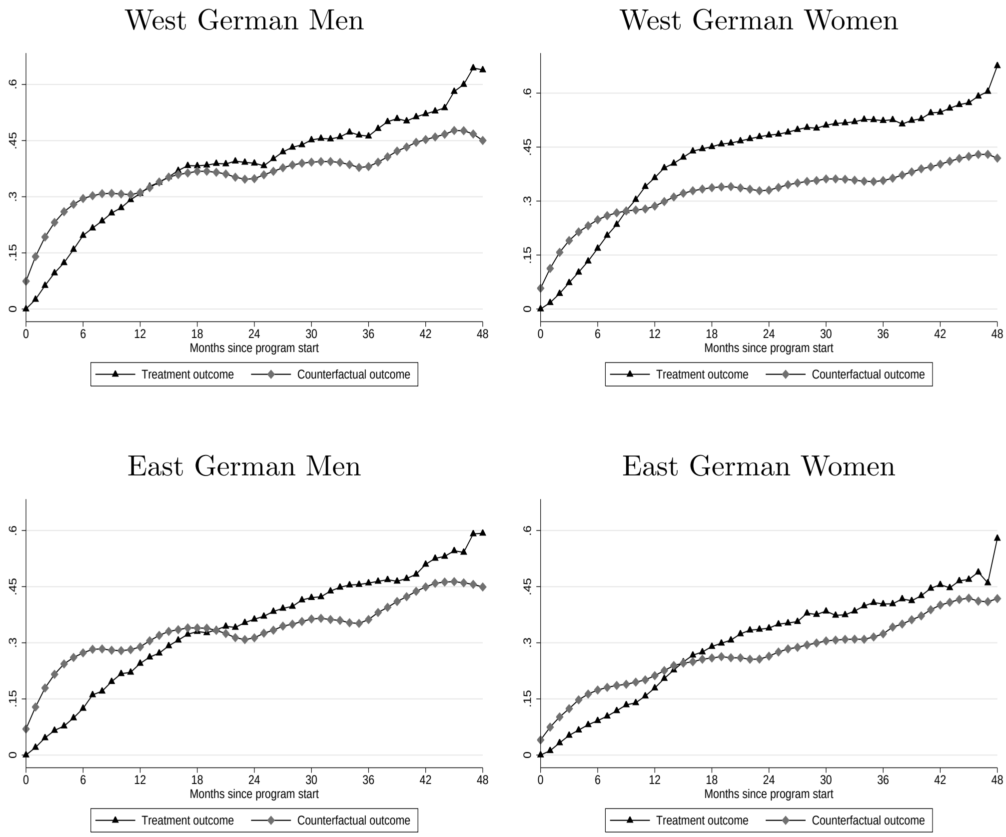
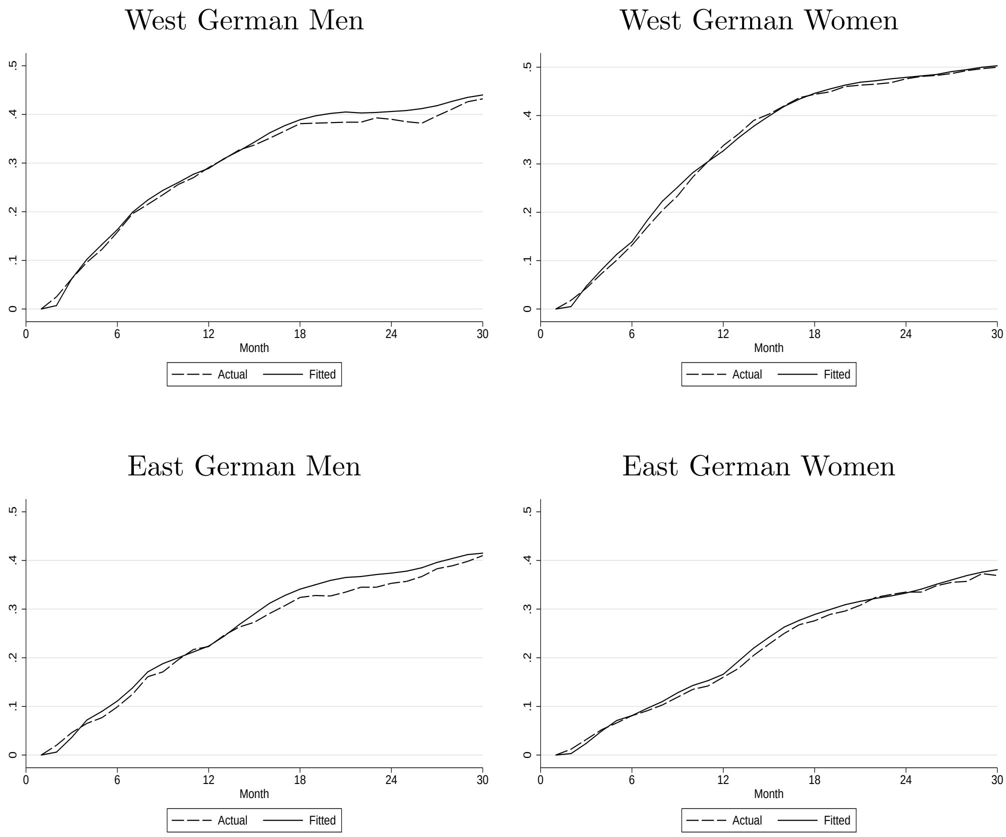


Figure 4: Raw Treatment and Nontreatment Employment Rates



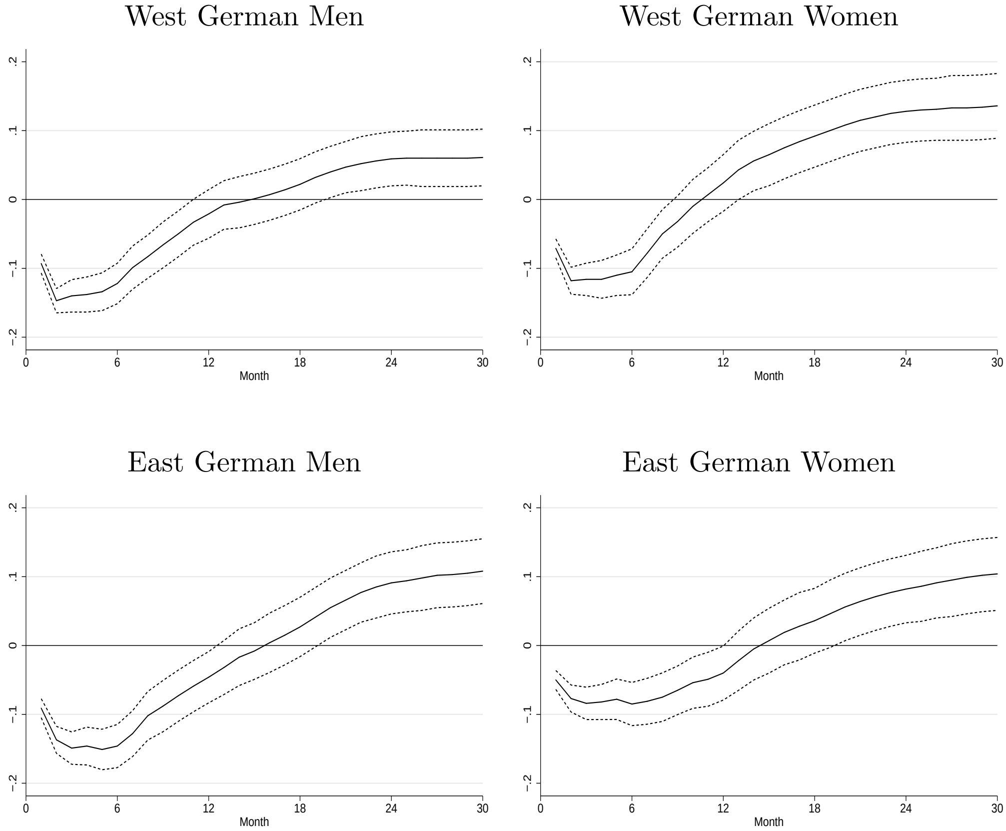
Notes: Raw estimates of the treatment effect on the treated, where treated and controls are aligned in the time dimension only. In particular, treated and nontreated individuals are matched on the calendar month of their first inflow and elapsed unemployment duration in the current spell. No adjustments are made for other potential sources of selection bias.

Figure 5: Actual and Predicted Employment Probabilities of Training Participants



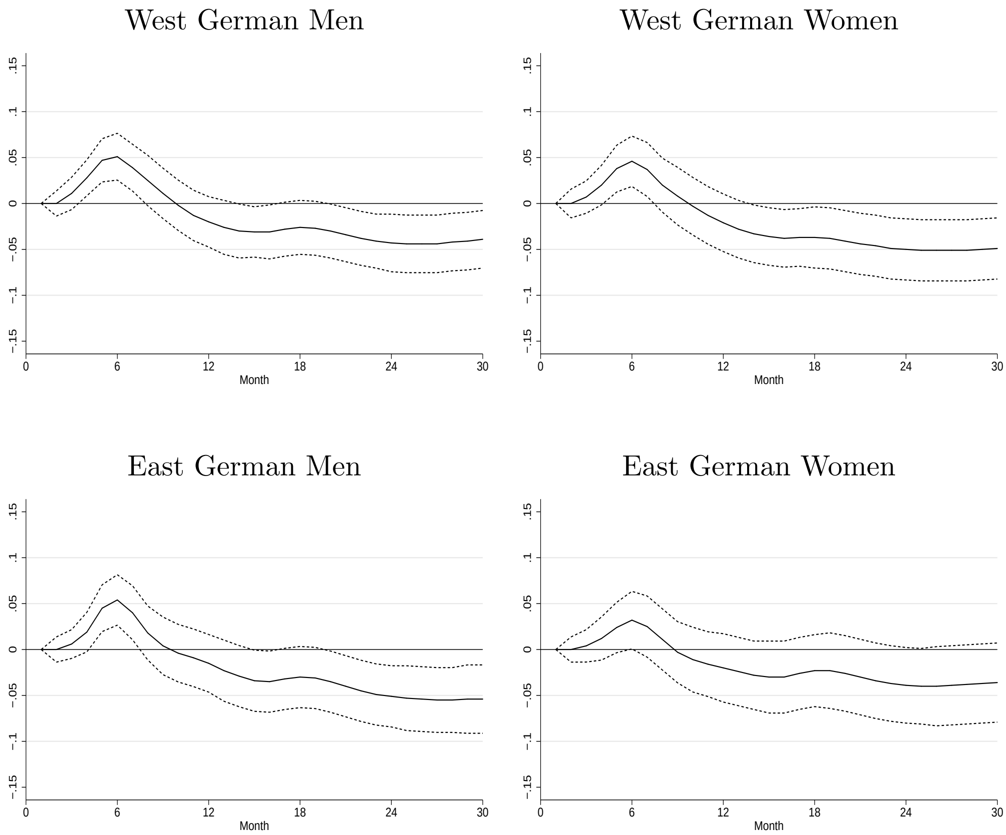
Notes: Actual and predicted employment rates measured on the ordinate, number of months since program start on the abscissa.

Figure 6: Classical Average Treatment Effect on the Treated



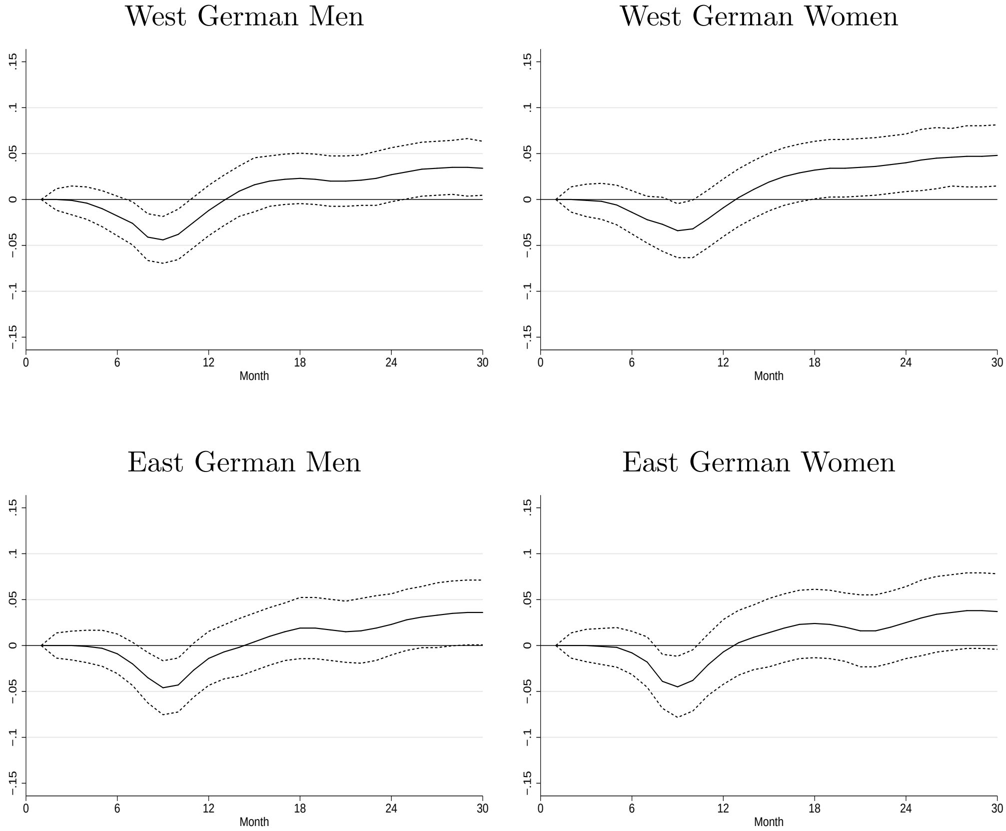
Notes: Difference in employment rates measured on the ordinate, number of months since program start on the abscissa.

Figure 7: ATT of Attending a Program Scheduled for Three versus Six Months



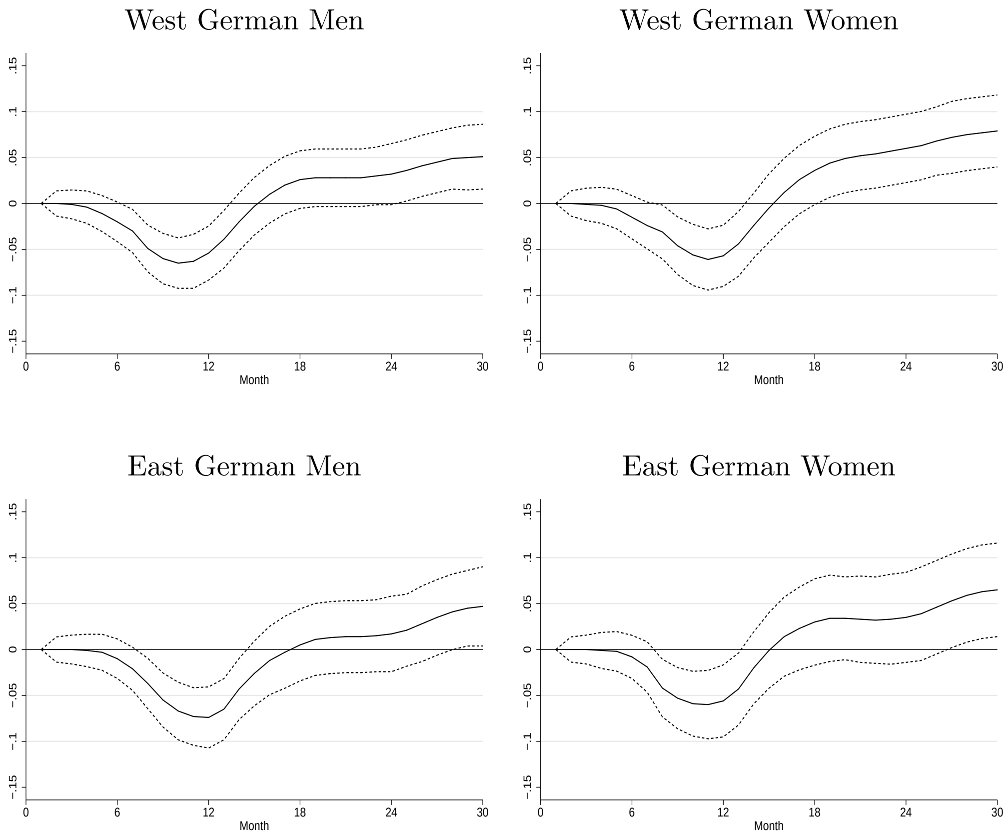
Notes: Difference in employment rates measured on the ordinate, number of months since program start on the abscissa.

Figure 8: ATT of Attending a Program Scheduled for Nine versus Six Months



Notes: Difference in employment rates measured on the ordinate, number of months since program start on the abscissa.

Figure 9: ATT of Attending a Program Scheduled for Twelve versus Six Months



Notes: Difference in employment rates measured on the ordinate, number of months since program start on the abscissa.

Tables

Table 1: Descriptive Statistics

	Male, West	Female, West	Male, East	Female, East
Persons	17,475	12,610	9,207	4,961
Months per Person	43.34	44.95	39.18	39.54
Months Employed p. P.	20.78	18.48	17.35	13.69
Months Unemployed p. P.	22.56	26.47	21.83	25.84
Months in Training p. P.	0.66	0.78	1.12	1.41
Employment Spells p. P.	1.50	1.11	1.33	0.91
Unemployment Spells p. P.	2.15	1.75	2.02	1.61
Training Spells p. P.	0.11	0.13	0.17	0.19

Appendix

A Detailed Information on the Data

This study uses data from the IEBS Version 4.02. A German description of the IEBS Version 3.01 can be found in Zimmermann et al. (2007). Information in English can be found on the website of the Research Data Center of the Federal Employment Agency (http://fdz.iab.de/en.aspx). The website also describes the conditions under which researchers may obtain access to the IEBS.

The first of the four administrative data sources included in the IEBS, the IAB Employment History, consists of social insurance register data for employees subject to contributions to the public social security system. It covers the time period from 1990 to 2004. The main feature of these data is detailed daily information on the employment status of each recorded individual. For each employment spell, in addition to start and end dates, data from the Employment History contain information on personal as well as job and firm characteristics such as wage, industry or occupation.

The IAB Benefit Recipient History, the second data source, includes daily spells of unemployment benefit, unemployment assistance and subsistence allowance payments the individuals received between January 1990 and June 2005. In addition to the sort of the payment and the start and end dates of periods of transfer receipt the spells contain further information like sanctions, periods of disqualification from benefit receipt and personal characteristics. Furthermore, the information in the Employment and the Benefit Recipient History allows one to calculate the individual entitlement periods to unemployment benefits.[35]

The third data source included in the IEBS is the so-called Data on Job Search Originating from the Applicants Pool Database, which contains rich information on individuals searching for jobs. It contains all the records starting January 2000 to June 2005 and partly also those beginning before 2000 if the person in question keeps the same client number throughout. The database includes a rich variety of information on personal characteristics (in particular education, family status and health condition), information related to placement fields (e.g. qualification and experience in the target profession), and regional information.

The Participants–in–Measures Data, the fourth data source, contains diverse information on participation in public sector sponsored labor market programs, for example training programs, job-creation measures, integration subsidies, business start-up allowances covering the period January 2000 to July 2005. Comparing the

[35]For the calculation of the claims, the present study relies on Plaßmann (2002) that contains a summary of the different regulations.

entries into different programs in 1999 with the figures for later years shows that information on programs starting in 1999 seems to be already complete for most active labor market programs. Furthermore, this database allows to distinguish subsidized employment in the context of active labor market policy from regular employment. Similar to the other sources, information comes in the form of spells indicating the start and end dates at the daily level, the type of the program as well as additional information on the program such as the planned end date or if the program ends with a certificate.

B Algorithm for the MCMC Estimation

Collect the observed explanatory variables in two vectors $\boldsymbol{Z}_{E,it} \equiv (E_i^{t-1}, Q_i^{t-1}, \boldsymbol{X}_{E,it})$ and $\boldsymbol{Z}_{Q,it} \equiv (E_i^{t}, Q_i^{t-1}, \boldsymbol{X}_{Q,it})$. We specify the index functions as linear in parameters, i.e. $\psi_j(\boldsymbol{Z}_{j_it}) = \boldsymbol{Z}_{j,it}\boldsymbol{\eta}_j$, $j = E, Q$. The posterior distribution combines the likelihood and the priors. We set the following independent priors: the prior distributions of the coefficients $\boldsymbol{\eta}_E$ are given by independent normal priors with distribution $\mathcal{N}(b_{E,0}, B_{E,0})$ with $b_{E,0} = 0$ and $B_{E,0} = 1000$. $\mathcal{N}(\textbf{.})$ denotes the normal distribution. Setting these very large values for the variance $B_{E,0}$, we use extremely diffuse priors. The same is done for the elements of the coefficient vector $\boldsymbol{\eta}_Q$, whose prior distributions are given by $\mathcal{N}(b_{Q,0}, B_{Q,0})$ with $b_{Q,0} = 0$ and $B_{Q,0} = 1000$. The distribution of the random effects is $\mathcal{N}(\mathbf{0}, \Sigma)$. The matrix Σ follows the prior distribution $\mathcal{W}^{-1}(h_0, H_0)$, where h_0 denotes the degrees of freedom and H_0 is the scale matrix. $\mathcal{W}^{-1}$ denotes the Inverse Wishart distribution. In order to set a diffuse prior, we choose a small value for h_0. In particular, we set $h_0 = 30$. The diagonal elements of H_0 are set to the individual level variances of separate Maximum Likelihood estimations of the two equations and the off-diagonal elements are set to zero.

- Step 0: Set starting values: For the elements of the coefficient vectors $\boldsymbol{\eta}_E$ and $\boldsymbol{\eta}_Q$ we use the estimated coefficients of pooled probit models as starting values. We set the starting values of the random effects ($\alpha_{E,i}$, $\alpha_{Q,i}$) to zeroes, and the starting value for the variance covariance matrix of the random effects Σ to $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

- Step 1a: Sample the latent employment propensity E_{it}^* from truncated $\mathcal{N}(\boldsymbol{Z}_{E,it}\boldsymbol{\eta}_E + \alpha_{E,i}, 1)$ with support $(0, \infty)$ if $E_{it} = 1$ and with support $(-\infty, 0]$ if $E_{it} = 0$.

- Step 1b: Sample the latent training propensity Q_{it}^* from truncated $\mathcal{N}(\boldsymbol{Z}_{Q,it}\boldsymbol{\eta}_Q + \alpha_{Q,i}, 1)$ with support $(0, \infty)$ if $Q_{it} = 1$ and with support $(-\infty, 0]$ if $Q_{it} = 0$ (using only the time periods in which the training equation is to be estimated).

- Step 2: Sample $(\alpha_{E,i}, \alpha_{Q,i})'$ from its bivariate normal conditional posterior distribution $\mathcal{N}(\boldsymbol{\mu}, V_{\alpha_i})$, where $\boldsymbol{\mu} = V_{\alpha_i} \cdot \begin{pmatrix} T_{E,i} & 0 \\ 0 & T_{Q,i} \end{pmatrix} \cdot \begin{pmatrix} (\bar{E}_i^* - \bar{\boldsymbol{Z}}_{E,i}\boldsymbol{\eta}_E) \\ (\bar{Q}_i^* - \bar{\boldsymbol{Z}}_{Q,i}\boldsymbol{\eta}_Q) \end{pmatrix}$ and $V_{\alpha_i} = \left(\Sigma^{-1} + \begin{pmatrix} T_{E,i} & 0 \\ 0 & T_{Q,i} \end{pmatrix}\right)^{-1}$, a bar over a variable denotes its mean across time, $T_{E,i}$ the number of observations for person i, and $T_{Q,i}$ the number of observations for person i for which the training equation is to be estimated.

- Step 3a: Sample the $\boldsymbol{\eta}_E$ vector from its multivariate normal conditional posterior distribution $\mathcal{N}(M_E, V_E)$, where $M_E = V_E(B_{E,0}^{-1} b_{E,0} + \sum_{i=1}^{N} \sum_{t=1}^{T_{E,i}} \boldsymbol{Z}'_{E,it}(E_{it}^* - \alpha_{E,i}))$ and $V_E = (B_{E,0}^{-1} + \sum_{i=1}^{N} \sum_{t=1}^{T_{E,i}} \boldsymbol{Z}'_{E,it}\boldsymbol{Z}_{E,it})^{-1}$. N is the number of persons in the data.

- Step 3b: Using only the time periods in which the training equation is to be estimated, sample the $\boldsymbol{\eta}_Q$ vector from its multivariate normal conditional posterior distribution $\mathcal{N}(M_Q, V_Q)$, where $M_Q = V_Q(B_{Q,0}^{-1} b_{Q,0} + \sum_{i=1}^{N} \sum_{t=1}^{T_{Q,i}} \boldsymbol{Z}'_{Q,it}(Q_{it}^* - \alpha_{Q,i}))$ and $V_Q = (B_{Q,0}^{-1} + \sum_{i=1}^{N} \sum_{t=1}^{T_{Q,i}} \boldsymbol{Z}'_{Q,it}\boldsymbol{Z}_{Q,it})^{-1}$.

- Step 4: Sample Σ from its conditional posterior distribution $\mathcal{W}^{-1}\left(N + h_0, \begin{pmatrix} \sum_{i=1}^{N} \alpha_{E,i}^2 & \sum_{i=1}^{N} \alpha_{E,i}\alpha_{Q,i} \\ \sum_{i=1}^{N} \alpha_{E,i}\alpha_{Q,i} & \sum_{i=1}^{N} \alpha_{Q,i}^2 \end{pmatrix} + H_0\right)$. Go to Step 1. Always use the current parameter values.

C Detailed Estimation Results

Table 2: Means and Standard Deviations of Parameters from MCMC Estimation

Name	Male West Mean	SD	Female West Mean	SD	Male East Mean	SD	Female East Mean	SD
Employment Equation								
Q_{t-1}	-0.422	0.057	-0.242	0.061	-0.400	0.062	-0.241	0.081
$[Q_{t-1} = 1] \times \ \ldots$								
$\ldots [Q_{t-2} = 1] \times [Q_{t-3} = 0]$	-0.094	0.079	-0.311	0.095	-0.286	0.108	-0.341	0.156
$\ldots [Q_{t-3} = 1] \times [Q_{t-4} = 0]$	-0.106	0.083	-0.236	0.096	-0.142	0.103	-0.126	0.142
$\ldots [Q_{t-4} = 1] \times [Q_{t-5} = 0]$	-0.167	0.091	-0.203	0.100	-0.359	0.129	-0.170	0.150
$\ldots [Q_{t-5} = 1] \times [Q_{t-6} = 0]$	-0.039	0.091	-0.165	0.103	-0.170	0.125	-0.299	0.174
$\ldots [Q_{t-6} = 1] \times [Q_{t-7} = 0]$	0.145	0.089	0.101	0.095	0.152	0.107	-0.088	0.154
$\ldots \sum_{j=48}^{7} [Q_{t-j} = 1 \times Q_{t-7} = 0]$	-0.075	0.111	0.236	0.102	0.189	0.119	-0.454	0.222
	-0.069	0.029	-0.011	0.019	-0.009	0.050	-0.008	0.046
$\ldots \tau_{Q,t} \times$ unskilled	0.064	0.010	0.047	0.011	0.036	0.021	0.015	0.031
$\ldots \tau_{Q,t} \times$ high school	0.027	0.010	0.038	0.011	0.013	0.013	0.023	0.015

<continued on next page>

Means and Standard deviations of Parameters from MCMC Estimation <continued>

Name	Male West Mean	SD	Female West Mean	SD	Male East Mean	SD	Female East Mean	SD
$\ldots \tau_{Q,t} \times$ health probl.	0.024	0.016	0.034	0.018	0.018	0.035	-0.001	0.040
$\ldots \tau_{Q,t} \times$ age ≥ 50	0.044	0.018	-0.019	0.021	-0.003	0.019	-0.005	0.021
$[Q_{t-1} = 0] \times [E_{t-1} = 0] \times \ \ldots$								
$\ldots D_t$	-0.129	0.058	-0.019	0.072	-0.073	0.063	-0.025	0.090
$\ldots D_t \times$ 1 m. passed since end	0.207	0.067	0.022	0.079	-0.063	0.082	-0.090	0.115
$\ldots D_t \times$ 2 or 3 m. passed since end	0.006	0.059	-0.097	0.070	-0.249	0.073	-0.224	0.104
$\ldots D_t \times$ 4 or 5 m. passed since end	0.033	0.059	-0.094	0.071	-0.313	0.075	-0.418	0.112
$\ldots D_t \times$ 6 to 11 m. passed since end	-0.095	0.043	-0.112	0.051	-0.187	0.050	-0.249	0.071
$\ldots D_t \times \tau_{Q,t}$	0.074	0.013	0.083	0.015	0.101	0.015	0.100	0.022
$\ldots D_t \times \tau^2_{Q,t}$	-0.003	0.001	-0.002	0.001	-0.003	0.001	-0.002	0.001
$\ldots D_t \times (t-e) \times \tau_{Q,t}$	-0.002	0.000	-0.002	0.000	-0.002	0.000	-0.002	0.001
$\ldots D_t \times$ unskilled	-0.037	0.065	0.094	0.085	0.043	0.116	0.104	0.191
$\ldots D_t \times$ high school	-0.080	0.096	0.039	0.113	0.018	0.131	0.160	0.168
$\ldots D_t \times$ health problems	0.002	0.102	-0.050	0.139	-0.070	0.186	0.098	0.272
$\ldots D_t \times$ age ≥ 50	0.231	0.114	-0.118	0.121	0.027	0.117	-0.312	0.165
$\ldots D_t \times \tau_{Q,t} \times$ unskilled	0.030	0.010	0.010	0.012	-0.000	0.017	-0.005	0.025
$\ldots D_t \times \tau_{Q,t} \times$ high school	0.020	0.012	0.019	0.014	-0.009	0.014	-0.016	0.018
$\ldots D_t \times \tau_{Q,t} \times$ health problems	-0.015	0.017	0.003	0.020	-0.012	0.023	-0.028	0.035
$\ldots D_t \times \tau_{Q,t} \times$ age ≥ 50	-0.048	0.019	-0.014	0.018	-0.018	0.014	0.013	0.018
$[Q_{t-1} = 0] \times [E_{t-1} = 1] \times \ \ldots$								
$\ldots D_t$	0.074	0.061	0.102	0.077	0.053	0.070	0.220	0.119
$\ldots D_t \times$ 1 m. passed since end	-0.322	0.105	-0.163	0.138	-0.329	0.151	0.299	0.315
$\ldots D_t \times$ 2 or 3 m. passed since end	-0.007	0.081	-0.011	0.101	-0.079	0.107	0.043	0.167
$\ldots D_t \times$ 4 or 5 m. passed since end	-0.009	0.075	0.060	0.097	-0.137	0.096	0.106	0.156
$\ldots D_t \times$ 6 to 11 m. passed since end	-0.005	0.049	0.042	0.057	0.019	0.063	0.060	0.096
$\ldots D_t \times \tau_{Q,t}$	0.034	0.015	0.027	0.017	0.063	0.019	-0.019	0.036
$\ldots D_t \times \tau^2_{Q,t}$	-0.001	0.001	-0.000	0.001	-0.003	0.001	0.002	0.002
$\ldots D_t \times (t-e) \times \tau_{Q,t}$	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
$\ldots D_t \times$ unskilled	-0.115	0.073	-0.066	0.099	0.331	0.167	0.265	0.274
$\ldots D_t \times$ high school	0.103	0.111	0.176	0.127	-0.023	0.138	-0.041	0.194
$\ldots D_t \times$ health problems	-0.123	0.147	0.066	0.189	0.067	0.303	0.250	0.329
$\ldots D_t \times$ age ≥ 50	-0.095	0.133	-0.208	0.140	-0.154	0.144	-0.201	0.220
$\ldots D_t \times \tau_{Q,t} \times$ unskilled	0.003	0.012	0.020	0.014	-0.046	0.026	0.018	0.042
$\ldots D_t \times \tau_{Q,t} \times$ high school	-0.030	0.014	-0.022	0.016	0.009	0.016	-0.001	0.022
$\ldots D_t \times \tau_{Q,t} \times$ health prob.	-0.008	0.026	0.003	0.026	-0.012	0.041	-0.013	0.044
$\ldots D_t \times \tau_{Q,t} \times$ age ≥ 50	0.019	0.024	0.015	0.022	-0.006	0.020	0.016	0.025
E_{t-1}	2.254	0.077	2.682	0.095	2.461	0.128	3.182	0.183
E_{t-2}	-0.071	0.016	-0.052	0.022	-0.089	0.023	-0.111	0.042
E_{t-3}	-0.037	0.016	-0.072	0.022	-0.037	0.023	-0.131	0.042
E_{t-4}	-0.032	0.014	-0.040	0.020	-0.029	0.021	-0.074	0.037
E_{t-5}	-0.023	0.015	-0.051	0.021	-0.052	0.021	-0.086	0.038
E_{t-6}	-0.139	0.014	-0.120	0.020	-0.126	0.020	-0.276	0.035

<continued on next page>

Means and Standard deviations of Parameters from MCMC Estimation <continued>

Name	Male West Mean	SD	Female West Mean	SD	Male East Mean	SD	Female East Mean	SD
$\sum_{j=7}^{11} E_{t-j}$	-0.076	0.006	-0.081	0.009	-0.076	0.009	-0.054	0.016
E_{t-12}	0.448	0.025	0.306	0.035	0.375	0.036	0.269	0.065
$\sum_{j=13}^{18} E_{t-j}$	-0.139	0.002	-0.142	0.003	-0.108	0.003	-0.180	0.006
$\sum_{j=19}^{24} E_{t-j}$	0.020	0.002	0.014	0.003	0.021	0.003	0.035	0.006
$\sum_{j=25}^{49} E_{t-j}$	-0.029	0.001	-0.037	0.002	-0.023	0.002	-0.041	0.004
$t > 1$	0.317	0.020	0.221	0.025	0.245	0.029	0.166	0.048
$t > 2$	0.144	0.020	0.047	0.025	0.146	0.028	0.107	0.048
$t > 3$	0.049	0.020	0.076	0.026	0.059	0.029	0.159	0.047
$t > 4$	0.019	0.021	0.070	0.028	-0.011	0.030	-0.027	0.048
$t > 5$	-0.004	0.022	0.057	0.028	0.042	0.031	0.168	0.049
$t > 6$	0.056	0.023	-0.004	0.030	0.017	0.033	-0.063	0.051
$t > 7$	-0.047	0.019	0.016	0.024	0.001	0.027	0.017	0.042
$t > 12$	-0.030	0.019	0.032	0.025	0.029	0.027	0.052	0.045
$t > 13$	0.335	0.019	0.223	0.025	0.254	0.027	0.237	0.045
$t > 19$	0.013	0.012	0.030	0.016	-0.009	0.018	0.056	0.028
$t > 25$	0.171	0.013	0.080	0.016	0.119	0.018	0.077	0.029
$E_{t-1} \times t$	0.070	0.033	-0.086	0.045	-0.104	0.048	-0.169	0.085
$\sum_{j=2}^{3} E_{t-j} \times t$	-0.079	0.021	0.037	0.029	-0.060	0.030	0.058	0.054
$\sum_{j=4}^{6} E_{t-j} \times t$	0.060	0.014	0.024	0.019	0.062	0.020	0.099	0.034
$\sum_{j=7}^{11} E_{t-j} \times t$	0.103	0.010	0.094	0.014	0.101	0.014	0.065	0.024
$E_{t-12} \times t$	0.191	0.038	0.280	0.055	0.080	0.055	0.459	0.098
$\tau_{E,t} \times [E_{t-1} = 0]$	-0.024	0.004	-0.074	0.004	-0.034	0.007	-0.067	0.008
$\tau_{E,t} \times [E_{t-1} = 1]$	0.037	0.005	0.009	0.005	0.038	0.008	-0.001	0.010
$\tau_{E,t}^2 \times [E_{t-1} = 0]$	0.001	0.000	0.001	0.000	0.001	0.000	0.001	0.000
$\tau_{E,t}^2 \times [E_{t-1} = 1]$	0.000	0.000	-0.000	0.000	-0.000	0.000	-0.000	0.000
last job: assisting workers	-0.041	0.015	0.026	0.028	-0.087	0.021	-0.020	0.053
last job: jobs in service	-0.065	0.022	0.001	0.023	-0.045	0.037	-0.092	0.042
last job: office or business job	-0.053	0.023	0.010	0.026	-0.096	0.043	-0.036	0.047
last job: technician or related	-0.028	0.024	-0.022	0.030	-0.013	0.037	-0.032	0.055
last job: academic or managers	-0.027	0.027	-0.012	0.031	-0.058	0.043	-0.018	0.060
share last wages censored	0.578	0.084	-0.011	0.124	0.605	0.168	0.481	0.264
log last average real wage	0.106	0.027	0.101	0.035	0.050	0.089	0.157	0.085
log last average real wage squared	0.014	0.004	-0.011	0.005	0.023	0.013	-0.002	0.012
last job: whitecollar job	-0.057	0.019	-0.029	0.021	-0.080	0.031	0.007	0.039

<continued on next page>

Means and Standard deviations of Parameters from MCMC Estimation <continued>

Name	Male West Mean	Male West SD	Female West Mean	Female West SD	Male East Mean	Male East SD	Female East Mean	Female East SD
last job: seasonal worker	0.123	0.019	0.079	0.025	0.116	0.026	0.212	0.041
last job: parttime worker	-0.045	0.023	-0.044	0.019	-0.001	0.045	0.035	0.037
months employed last 3 years	0.023	0.003	0.021	0.004	0.007	0.005	0.021	0.008
months employed last 3 years squared	-0.000	0.000	-0.000	0.000	0.000	0.000	-0.000	0.000
region with bad labor market conditions	0.003	0.078	0.213	0.099	-0.451	0.337	0.058	0.314
urban region with many unemployed	-0.115	0.017	-0.049	0.023	-0.514	0.337	0.044	0.315
age/100	-0.698	0.085	0.099	0.109	-0.613	0.117	-1.276	0.221
low skilled	-0.175	0.014	-0.204	0.020	-0.177	0.029	-0.212	0.053
no schooling degree	-0.202	0.015	-0.481	0.024	-0.211	0.030	-0.611	0.064
high school (Abitur)	-0.181	0.023	-0.025	0.025	-0.064	0.037	-0.093	0.052
health problems	-0.642	0.029	-0.697	0.039	-0.666	0.050	-0.611	0.083
at least one child	-0.326	0.015	-0.515	0.021	-0.399	0.020	-0.648	0.037
winter (Jan.–Mar.)	0.078	0.009	0.053	0.012	-0.009	0.014	0.024	0.022
spring (Apr.–Jun.)	0.386	0.009	0.171	0.011	0.375	0.013	0.232	0.020
summer (Jul.–Sept.)	0.287	0.008	0.118	0.010	0.302	0.011	0.160	0.018
year 1999 or 2000	0.437	0.033	0.106	0.043	0.264	0.050	-0.038	0.078
year 2001	0.287	0.026	0.087	0.035	0.145	0.040	0.004	0.063
year 2002	0.122	0.019	0.034	0.025	0.007	0.029	-0.027	0.046
year 2003	0.099	0.014	0.011	0.018	0.052	0.022	0.017	0.034
age/100 × E_{t-1}	0.155	0.099	-0.444	0.131	0.346	0.145	-0.842	0.246
low skilled × E_{t-1}	0.072	0.016	0.012	0.023	0.120	0.033	0.167	0.057
high school (Abitur) × E_{t-1}	0.281	0.027	0.085	0.029	0.277	0.044	0.285	0.059
health problems × E_{t-1}	0.116	0.035	0.054	0.049	0.100	0.068	-0.177	0.105
share last wages censored × E_{t-1}	-0.136	0.092	-0.655	0.129	-0.742	0.137	-0.903	0.278
log last average real wage × E_{t-1}	-0.039	0.017	-0.089	0.019	-0.065	0.029	-0.142	0.040
age/100 × $\tau_{E,t}$	-0.056	0.006	-0.017	0.006	-0.042	0.008	-0.016	0.010
low skilled × $\tau_{E,t}$	0.002	0.001	0.002	0.001	0.003	0.002	0.002	0.002
high skilled × $\tau_{E,t}$	0.001	0.001	-0.003	0.002	0.001	0.002	0.003	0.003
health problems × $\tau_{E,t}$	0.009	0.001	0.007	0.002	0.004	0.003	0.004	0.004
share last wages censored × $\tau_{E,t}$	-0.033	0.005	-0.008	0.006	-0.019	0.007	-0.007	0.011
log last average real wage × $\tau_{E,t}$	-0.007	0.001	-0.003	0.001	-0.004	0.001	-0.002	0.001
age/100 × $\tau_{E,t}$ ×E_{t-1}	0.037	0.008	0.056	0.009	0.014	0.011	0.048	0.017
low skilled × $\tau_{E,t}$ ×E_{t-1}	-0.001	0.001	-0.002	0.002	-0.000	0.003	-0.005	0.004
high school (Abitur) × $\tau_{E,t}$ ×E_{t-1}	0.002	0.002	-0.000	0.002	0.002	0.003	-0.014	0.004
health problems × $\tau_{E,t}$ ×E_{t-1}	-0.013	0.003	-0.016	0.004	-0.016	0.007	-0.002	0.010
share last wages cens. × $\tau_{E,t}$ ×E_{t-1}	0.029	0.007	0.031	0.010	0.030	0.010	0.068	0.021
log last av. real wage × $\tau_{E,t}$ ×E_{t-1}	0.004	0.001	0.005	0.001	0.002	0.002	0.009	0.003
remaining UB claim less than six months	-0.039	0.014	0.016	0.018	-0.006	0.020	0.021	0.033
remaining UB claim less than four months	0.037	0.017	0.083	0.022	0.017	0.025	0.071	0.040
remaining UB claim less than two months	0.075	0.015	0.128	0.019	0.026	0.022	0.136	0.033
unemployment rate in county last year	-0.002	0.002	-0.001	0.003	0.005	0.002	0.002	0.004
nr job creat. / nr unemp. last year in office	0.358	0.464	-2.253	0.655	0.308	0.175	0.204	0.317

<continued on next page>

Means and Standard deviations of Parameters from MCMC Estimation <continued>

Name	Male West Mean	SD	Female West Mean	SD	Male East Mean	SD	Female East Mean	SD
entries train. / nr unemp. last year in office	0.600	0.119	0.470	0.157	0.816	0.157	0.048	0.256
constant	-2.540	0.092	-1.855	0.115	-2.025	0.386	-2.014	0.384
Participation Equation								
Q_{t-1}	2.044	0.096	1.862	0.115	1.626	0.113	1.436	0.146
$\tau_{Q,t}$	0.006	0.016	-0.002	0.019	0.020	0.030	0.096	0.034
months to pl. end if enough dur. left	0.557	0.016	0.620	0.019	0.836	0.024	0.767	0.024
months to pl. end if . . . squared	-0.019	0.001	-0.019	0.001	-0.036	0.001	-0.028	0.001
$\tau_{E,t}$	0.024	0.004	0.001	0.005	0.028	0.005	0.008	0.006
$\tau_{E,t}^2$	-0.001	0.000	-0.000	0.000	-0.001	0.000	-0.000	0.000
inflow month	0.009	0.056	0.073	0.063	-0.192	0.069	-0.365	0.102
days inflow to end month if $t = 1$	-0.022	0.004	-0.030	0.005	-0.022	0.005	-0.030	0.012
repeated inflow	0.001	0.030	0.013	0.042	-0.006	0.037	-0.095	0.051
younger than 30	0.060	0.033	-0.156	0.044	-0.006	0.042	-0.060	0.055
30–34 years old	-0.017	0.029	-0.121	0.037	-0.030	0.037	-0.019	0.044
40–44 years old	0.016	0.030	0.031	0.038	-0.027	0.036	0.004	0.043
45–49 years old	-0.057	0.036	0.010	0.043	-0.022	0.039	-0.063	0.046
50 years or more	-0.213	0.043	-0.251	0.053	-0.121	0.046	-0.117	0.051
no schooling degree	-0.155	0.029	-0.372	0.046	-0.203	0.051	-0.274	0.066
high school (Abitur)	0.180	0.030	0.008	0.034	0.130	0.040	0.058	0.037
no vocational degree	-0.045	0.025	-0.109	0.032	-0.152	0.043	-0.161	0.054
last job: office or business jobs	0.153	0.037	0.327	0.032	0.116	0.054	0.071	0.031
last job: technician or related	0.122	0.037	0.051	0.041	0.063	0.048	0.086	0.043
last job: whitecollar job	0.095	0.028	0.045	0.037	0.135	0.038	0.173	0.038
last job: seasonal worker	-0.173	0.039	-0.097	0.045	-0.151	0.048	-0.089	0.046
last job: parttime worker	0.032	0.041	-0.023	0.036	-0.046	0.067	0.089	0.037
log last average real wage	0.048	0.016	-0.010	0.020	0.022	0.019	0.062	0.024
health problems	0.031	0.034	-0.008	0.043	-0.140	0.048	-0.050	0.056
at least one child	0.038	0.025	-0.002	0.032	0.029	0.027	0.026	0.030
months employed last 3 years	-0.002	0.006	-0.004	0.007	-0.009	0.007	-0.014	0.008
months employed last 3 years squared	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
winter (Jan.–Mar.)	0.168	0.026	0.215	0.030	0.192	0.032	0.263	0.039
spring (Apr.–Jun.)	0.135	0.025	0.150	0.030	0.172	0.031	0.299	0.038
summer (Jul.–Sept.)	0.085	0.025	0.105	0.029	0.111	0.031	0.158	0.039
year 1999 or 2000	0.542	0.071	0.410	0.080	0.361	0.085	0.254	0.112
year 2001	0.362	0.065	0.205	0.074	0.299	0.078	0.268	0.106
year 2002	0.335	0.064	0.139	0.073	0.257	0.072	0.256	0.098
year 2003	0.094	0.060	-0.112	0.064	-0.020	0.074	-0.043	0.096
entitl. unempl. compens. in months	0.004	0.008	-0.006	0.008	-0.025	0.009	-0.010	0.012
entitl. unempl. comp. in m's squared	-0.001	0.000	0.000	0.000	0.001	0.000	-0.001	0.001
still entitled to unempl. compens.	0.066	0.037	0.065	0.043	0.140	0.043	0.141	0.055
younger than 30 × $\tau_{Q,t}$	0.032	0.018	0.016	0.023	-0.001	0.024	-0.027	0.027
30–34 years old × $\tau_{Q,t}$	0.029	0.014	-0.016	0.018	-0.011	0.019	-0.037	0.024

<continued on next page>

Means and Standard deviations of Parameters from MCMC Estimation <continued>

Name	Male West		Female West		Male East		Female East	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
40–44 years old × $\tau_{Q,t}$	-0.042	0.018	-0.019	0.019	-0.056	0.020	-0.032	0.023
45–49 years old × $\tau_{Q,t}$	-0.014	0.017	-0.011	0.020	-0.045	0.021	-0.010	0.024
50 years or more × $\tau_{Q,t}$	-0.059	0.025	-0.023	0.028	-0.078	0.023	-0.049	0.027
no vocat. degree × $\tau_{Q,t}$	0.027	0.011	0.030	0.013	0.028	0.019	0.027	0.023
younger than 30 × Q_{t-1}	-0.262	0.119	-0.122	0.159	-0.139	0.166	0.505	0.233
30–34 years old × Q_{t-1}	-0.216	0.105	0.043	0.135	0.026	0.145	0.448	0.195
40–44 years old × Q_{t-1}	0.246	0.124	0.054	0.134	0.347	0.154	0.431	0.191
45–49 years old × Q_{t-1}	0.117	0.128	-0.001	0.144	0.299	0.165	0.158	0.203
50 years or more × Q_{t-1}	0.534	0.163	0.078	0.179	0.670	0.195	0.625	0.231
no vocat. degree × Q_{t-1}	-0.056	0.075	0.037	0.091	0.075	0.135	0.235	0.174
unemployment rate in county last month	0.004	0.004	0.006	0.005	0.005	0.004	-0.000	0.005
unempl. rate in county last month × $\tau_{Q,t}$	0.001	0.001	0.003	0.001	0.000	0.001	-0.003	0.001
nr training / nr unemp. last month in office	0.044	0.307	-0.489	0.415	0.151	0.388	0.707	0.452
nr job creat. / nr unemp. last month in office	-0.307	0.489	-0.788	0.616	0.203	0.296	0.249	0.357
entries train. / nr unemp. last year in office	0.555	0.318	2.091	0.391	0.572	0.331	1.011	0.377
constant	-3.632	0.135	-3.281	0.179	-3.161	0.165	-3.242	0.201
Individual Level Variances and Covariances								
$\text{Var}(\alpha_E)$	0.254	0.013	0.332	0.019	0.230	0.017	0.458	0.042
$\text{Var}(\alpha_Q)$	0.204	0.034	0.318	0.052	0.127	0.028	0.049	0.012
$\text{Cov}(\alpha_E, \alpha_Q)$	0.002	0.013	0.010	0.020	-0.008	0.015	0.003	0.019
$\text{Var}(\alpha_E)/(\text{Var}(\alpha_E)+1)$	0.202	0.008	0.249	0.011	0.187	0.011	0.314	0.020
$\text{Var}(\alpha_Q)/(\text{Var}(\alpha_Q)+1)$	0.169	0.024	0.240	0.030	0.112	0.022	0.046	0.011
$\text{Corr}(\alpha_E+\epsilon_{E,t}, \alpha_Q+\epsilon_{Q,t})$	0.001	0.011	0.007	0.015	-0.006	0.012	0.003	0.015
$\text{Corr}(\alpha_E, \alpha_Q)$	0.008	0.059	0.031	0.062	-0.042	0.085	0.024	0.127

Notes: $t = 1, \ldots, 50$ denote the month since the inflow into unemployment. E_t indicates the employment status and Q_t the training status in period t. $\tau_{E,t}$ and $\tau_{Q,t}$ indicate the elapsed duration in employment/unemployment and training, respectively. D_t is a dummy equal to one if a participation in training occurred during any previous quarter since the inflow. $(t-e)$ denotes the elapsed time since the end of program participation. α_E (α_Q) denotes the individual specific effect in the employment (qualification) equation, $\epsilon_{E,t}$ ($\epsilon_{Q,t}$) the idiosyncratic error term in the employment (qualification) equation.

Table 3: Employment Rate and Number of Participants Still Observed Aligned to Start of Program

	Male West			Female West			Male East			Female East		
$t-s$	$\bar{E}_{t-s}$	$\hat{E}_{t-s}$	N_{t-s}	$\bar{E}_{t-s}$	$\hat{E}_{t-s}$	N_{t-s}	$\bar{E}_{t-s}$	$\hat{E}_{t-s}$	N_{t-s}	$\bar{E}_{t-s}$	$\hat{E}_{t-s}$	N_{t-s}
1	0.000	0.000	1850	0.000	0.000	1483	0.000	0.000	1385	0.000	0.000	865
2	0.025	0.007	1850	0.018	0.005	1483	0.020	0.006	1385	0.012	0.003	865
3	0.062	0.062	1840	0.043	0.047	1476	0.046	0.036	1379	0.032	0.024	864
4	0.096	0.102	1836	0.073	0.081	1473	0.065	0.072	1376	0.052	0.049	860
5	0.123	0.133	1832	0.101	0.113	1469	0.077	0.090	1372	0.066	0.071	860
6	0.158	0.163	1825	0.132	0.139	1462	0.099	0.111	1368	0.081	0.081	856
7	0.196	0.199	1818	0.169	0.183	1452	0.125	0.138	1365	0.091	0.096	855
8	0.215	0.224	1810	0.204	0.223	1448	0.161	0.171	1357	0.103	0.110	852
9	0.235	0.244	1799	0.234	0.252	1437	0.171	0.188	1354	0.119	0.128	849
10	0.256	0.260	1790	0.273	0.282	1429	0.196	0.200	1349	0.135	0.143	847
11	0.270	0.277	1780	0.305	0.305	1422	0.217	0.212	1345	0.142	0.153	845
12	0.291	0.289	1774	0.338	0.327	1412	0.223	0.224	1340	0.160	0.166	843
13	0.308	0.309	1759	0.362	0.354	1409	0.246	0.244	1335	0.178	0.193	837
14	0.327	0.325	1741	0.390	0.378	1396	0.263	0.268	1321	0.205	0.220	835
15	0.337	0.343	1728	0.403	0.399	1389	0.273	0.290	1304	0.228	0.242	830
16	0.351	0.362	1714	0.420	0.419	1377	0.291	0.312	1294	0.250	0.263	828
17	0.366	0.377	1700	0.437	0.434	1370	0.307	0.328	1281	0.268	0.277	821
18	0.381	0.389	1684	0.444	0.446	1365	0.324	0.341	1273	0.276	0.289	813
19	0.382	0.397	1671	0.449	0.455	1356	0.328	0.350	1264	0.289	0.299	810
20	0.383	0.402	1660	0.460	0.463	1349	0.327	0.359	1248	0.296	0.309	796
21	0.384	0.405	1634	0.463	0.469	1338	0.335	0.365	1237	0.308	0.316	791
22	0.384	0.403	1616	0.465	0.472	1327	0.345	0.367	1216	0.324	0.322	778
23	0.393	0.404	1591	0.468	0.476	1314	0.345	0.371	1199	0.330	0.327	772
24	0.390	0.406	1571	0.476	0.479	1305	0.353	0.374	1181	0.335	0.333	762
25	0.385	0.408	1547	0.481	0.482	1287	0.357	0.378	1162	0.335	0.341	753
26	0.382	0.412	1520	0.483	0.485	1278	0.367	0.385	1141	0.348	0.351	744
27	0.397	0.418	1492	0.487	0.491	1259	0.383	0.396	1118	0.355	0.360	730
28	0.411	0.427	1464	0.493	0.495	1249	0.389	0.404	1092	0.357	0.369	719
29	0.426	0.435	1426	0.497	0.500	1227	0.398	0.412	1059	0.373	0.376	711
30	0.432	0.440	1399	0.500	0.503	1212	0.410	0.415	1033	0.369	0.381	697

Notes: $t-s$ denotes the months elapsed since program start, and $\bar{E}_{t-s}$ the sample mean of the employment dummy in month $t-s$. $\hat{E}_{t-s}$ is the mean of the employment dummy as predicted using the simulation strategy (prediction of treatment outcomes).

Table 4: Classical ATT Aligned to Program Start

	Male West		Female West		Male East		Female East	
$t-s$	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	-0.093	0.007	-0.071	0.007	-0.091	0.007	-0.050	0.007
2	-0.147	0.009	-0.118	0.010	-0.137	0.010	-0.077	0.010
3	-0.140	0.012	-0.116	0.012	-0.149	0.012	-0.084	0.012
4	-0.138	0.013	-0.116	0.014	-0.146	0.014	-0.082	0.013
5	-0.134	0.014	-0.110	0.015	-0.151	0.015	-0.078	0.015
6	-0.122	0.015	-0.105	0.017	-0.146	0.016	-0.085	0.016
7	-0.099	0.016	-0.078	0.018	-0.128	0.017	-0.081	0.017
8	-0.083	0.016	-0.050	0.018	-0.102	0.018	-0.075	0.018
9	-0.066	0.017	-0.032	0.019	-0.088	0.019	-0.065	0.018
10	-0.050	0.017	-0.010	0.020	-0.073	0.019	-0.054	0.019
11	-0.033	0.017	0.007	0.020	-0.059	0.019	-0.049	0.020
12	-0.021	0.018	0.024	0.021	-0.046	0.019	-0.040	0.020
13	-0.008	0.018	0.043	0.022	-0.032	0.020	-0.022	0.022
14	-0.004	0.019	0.056	0.022	-0.017	0.021	-0.005	0.023
15	0.001	0.019	0.065	0.023	-0.008	0.021	0.007	0.024
16	0.007	0.019	0.075	0.023	0.004	0.022	0.019	0.024
17	0.014	0.019	0.084	0.023	0.015	0.022	0.028	0.025
18	0.022	0.019	0.092	0.023	0.027	0.022	0.036	0.024
19	0.032	0.019	0.100	0.023	0.041	0.022	0.046	0.025
20	0.040	0.019	0.108	0.023	0.055	0.022	0.056	0.025
21	0.047	0.019	0.115	0.023	0.066	0.022	0.064	0.025
22	0.052	0.020	0.120	0.023	0.077	0.022	0.071	0.025
23	0.056	0.020	0.125	0.023	0.085	0.023	0.077	0.025
24	0.059	0.020	0.128	0.023	0.091	0.023	0.082	0.025
25	0.060	0.020	0.130	0.023	0.094	0.023	0.086	0.026
26	0.060	0.021	0.131	0.023	0.098	0.024	0.091	0.026
27	0.060	0.021	0.133	0.024	0.102	0.024	0.095	0.027
28	0.060	0.021	0.133	0.024	0.103	0.024	0.099	0.027
29	0.060	0.021	0.134	0.024	0.105	0.024	0.102	0.027
30	0.061	0.021	0.136	0.024	0.108	0.024	0.104	0.027

Notes: $t-s$ denotes the months elapsed since program start.

Table 5: Predicted Participation and Employment Rates for Different Planned Program Durations: Male, West

	Three Months		Six Months		Nine Months		Twelve Months	
$t-s$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$
1	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
2	0.763	0.043	0.934	0.043	0.954	0.043	0.957	0.043
3	0.502	0.086	0.858	0.075	0.915	0.074	0.920	0.074
4	0.258	0.136	0.751	0.107	0.873	0.104	0.886	0.103
5	0.145	0.183	0.605	0.135	0.830	0.126	0.857	0.125
6	0.086	0.222	0.415	0.171	0.766	0.153	0.822	0.150
7	0.053	0.253	0.219	0.214	0.667	0.188	0.776	0.184
8	0.034	0.276	0.126	0.251	0.541	0.210	0.737	0.201
9	0.023	0.290	0.076	0.279	0.374	0.235	0.684	0.219
10	0.016	0.296	0.048	0.298	0.203	0.260	0.607	0.233
11	0.011	0.300	0.032	0.313	0.118	0.288	0.493	0.250
12	0.008	0.303	0.022	0.324	0.072	0.312	0.343	0.270
13	0.006	0.305	0.015	0.331	0.047	0.331	0.189	0.293
14	0.004	0.312	0.011	0.342	0.032	0.351	0.112	0.322
15	0.003	0.324	0.008	0.355	0.022	0.371	0.071	0.352
16	0.003	0.339	0.006	0.369	0.016	0.390	0.048	0.380
17	0.002	0.352	0.005	0.380	0.012	0.403	0.034	0.400
18	0.002	0.363	0.004	0.390	0.010	0.413	0.025	0.416
19	0.001	0.371	0.003	0.398	0.008	0.421	0.019	0.427
20	0.001	0.375	0.003	0.405	0.006	0.425	0.015	0.434
21	0.001	0.376	0.002	0.410	0.005	0.430	0.012	0.438
22	0.001	0.373	0.002	0.411	0.005	0.431	0.010	0.439
23	0.001	0.370	0.002	0.411	0.004	0.435	0.009	0.441
24	0.001	0.369	0.002	0.412	0.004	0.439	0.007	0.445
25	0.001	0.367	0.002	0.412	0.003	0.441	0.006	0.447
26	0.001	0.370	0.001	0.414	0.003	0.447	0.006	0.455
27	0.001	0.374	0.001	0.418	0.003	0.452	0.005	0.463
28	0.001	0.383	0.001	0.425	0.002	0.460	0.005	0.473
29	0.000	0.391	0.001	0.432	0.002	0.467	0.004	0.482
30	0.000	0.397	0.001	0.436	0.002	0.470	0.004	0.487

Notes: $t-s$ denotes the months elapsed since program start. $\hat{Q}_{t-s}$ and $\hat{E}_{t-s}$ are the simulated means of the participation and employment probability, respectively.

Table 6: Predicted Participation and Employment Rates for Different Planned Program Durations: Female, West

	Three Months		Six Months		Nine Months		Twelve Months	
$t-s$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$
1	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
2	0.780	0.046	0.940	0.047	0.953	0.047	0.953	0.046
3	0.522	0.074	0.885	0.067	0.927	0.067	0.929	0.067
4	0.263	0.112	0.791	0.092	0.896	0.090	0.903	0.090
5	0.147	0.156	0.645	0.119	0.859	0.112	0.875	0.112
6	0.088	0.195	0.445	0.150	0.807	0.135	0.848	0.134
7	0.055	0.231	0.228	0.194	0.712	0.172	0.805	0.170
8	0.035	0.260	0.125	0.240	0.567	0.213	0.752	0.209
9	0.024	0.283	0.076	0.275	0.389	0.241	0.704	0.229
10	0.017	0.302	0.048	0.305	0.202	0.273	0.632	0.249
11	0.012	0.316	0.032	0.329	0.115	0.308	0.517	0.268
12	0.009	0.328	0.022	0.349	0.071	0.340	0.359	0.292
13	0.007	0.338	0.016	0.366	0.046	0.367	0.190	0.322
14	0.005	0.351	0.012	0.384	0.031	0.395	0.110	0.360
15	0.004	0.364	0.009	0.400	0.022	0.419	0.069	0.396
16	0.003	0.378	0.007	0.416	0.016	0.441	0.045	0.428
17	0.003	0.390	0.005	0.427	0.012	0.456	0.031	0.453
18	0.002	0.400	0.004	0.437	0.010	0.469	0.023	0.473
19	0.002	0.407	0.004	0.445	0.008	0.479	0.017	0.489
20	0.002	0.413	0.003	0.454	0.006	0.488	0.013	0.502
21	0.001	0.417	0.003	0.461	0.005	0.496	0.010	0.513
22	0.001	0.420	0.002	0.465	0.004	0.501	0.008	0.520
23	0.001	0.420	0.002	0.468	0.004	0.506	0.007	0.525
24	0.001	0.421	0.002	0.471	0.003	0.511	0.006	0.531
25	0.001	0.422	0.001	0.474	0.003	0.517	0.005	0.537
26	0.001	0.424	0.001	0.475	0.002	0.520	0.004	0.543
27	0.001	0.428	0.001	0.479	0.002	0.526	0.004	0.551
28	0.001	0.432	0.001	0.482	0.002	0.530	0.003	0.557
29	0.001	0.436	0.001	0.486	0.002	0.533	0.003	0.563
30	0.001	0.439	0.001	0.489	0.002	0.536	0.003	0.568

Notes: $t-s$ denotes the months elapsed since program start. $\hat{Q}_{t-s}$ and $\hat{E}_{t-s}$ are the simulated means of the participation and employment probability, respectively.

Table 7: Predicted Participation and Employment Rates for Different Planned Program Durations: Male, East

	Three Months		Six Months		Nine Months		Twelve Months	
$t-s$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$
1	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
2	0.840	0.036	0.961	0.036	0.964	0.036	0.964	0.036
3	0.571	0.058	0.931	0.052	0.944	0.052	0.945	0.052
4	0.241	0.094	0.871	0.075	0.918	0.074	0.919	0.074
5	0.113	0.131	0.757	0.086	0.902	0.084	0.905	0.083
6	0.057	0.161	0.523	0.107	0.875	0.097	0.887	0.097
7	0.030	0.184	0.220	0.143	0.813	0.123	0.856	0.122
8	0.016	0.201	0.100	0.183	0.690	0.148	0.825	0.146
9	0.009	0.214	0.049	0.210	0.469	0.164	0.799	0.155
10	0.005	0.225	0.026	0.229	0.197	0.186	0.753	0.162
11	0.003	0.234	0.014	0.243	0.091	0.216	0.644	0.170
12	0.002	0.242	0.008	0.257	0.045	0.243	0.435	0.183
13	0.001	0.247	0.005	0.270	0.024	0.263	0.181	0.206
14	0.001	0.258	0.003	0.287	0.014	0.285	0.085	0.244
15	0.001	0.271	0.002	0.305	0.008	0.309	0.043	0.279
16	0.000	0.287	0.001	0.322	0.005	0.332	0.023	0.309
17	0.000	0.301	0.001	0.333	0.003	0.349	0.014	0.330
18	0.000	0.312	0.001	0.342	0.002	0.361	0.008	0.347
19	0.000	0.319	0.000	0.350	0.002	0.370	0.006	0.361
20	0.000	0.325	0.000	0.360	0.001	0.377	0.004	0.373
21	0.000	0.327	0.000	0.367	0.001	0.382	0.003	0.381
22	0.000	0.326	0.000	0.371	0.001	0.386	0.002	0.385
23	0.000	0.325	0.000	0.374	0.001	0.393	0.002	0.389
24	0.000	0.326	0.000	0.377	0.000	0.401	0.001	0.394
25	0.000	0.327	0.000	0.380	0.000	0.408	0.001	0.402
26	0.000	0.331	0.000	0.385	0.000	0.416	0.001	0.414
27	0.000	0.339	0.000	0.394	0.000	0.427	0.001	0.429
28	0.000	0.346	0.000	0.401	0.000	0.436	0.001	0.442
29	0.000	0.352	0.000	0.407	0.000	0.443	0.000	0.452
30	0.000	0.356	0.000	0.409	0.000	0.446	0.000	0.456

Notes: $t-s$ denotes the months elapsed since program start. $\hat{Q}_{t-s}$ and $\hat{E}_{t-s}$ are the simulated means of the participation and employment probability, respectively.

Table 8: Predicted Participation and Employment Rates for Different Planned Program Durations: Female, East

	Three Months		Six Months		Nine Months		Twelve Months	
$t-s$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$	$\hat{Q}_{t-s}$	$\hat{E}_{t-s}$
1	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000
2	0.817	0.027	0.969	0.027	0.974	0.026	0.974	0.026
3	0.530	0.043	0.940	0.039	0.960	0.038	0.961	0.038
4	0.217	0.070	0.873	0.058	0.939	0.057	0.940	0.057
5	0.097	0.098	0.737	0.074	0.918	0.072	0.924	0.072
6	0.047	0.122	0.499	0.090	0.894	0.081	0.911	0.081
7	0.023	0.140	0.211	0.115	0.836	0.096	0.893	0.096
8	0.012	0.153	0.096	0.142	0.716	0.103	0.882	0.099
9	0.006	0.164	0.047	0.166	0.483	0.122	0.854	0.113
10	0.004	0.173	0.024	0.184	0.211	0.146	0.802	0.125
11	0.002	0.182	0.013	0.198	0.099	0.176	0.687	0.138
12	0.001	0.190	0.007	0.210	0.050	0.203	0.475	0.154
13	0.001	0.198	0.004	0.222	0.027	0.225	0.215	0.179
14	0.001	0.207	0.003	0.235	0.015	0.244	0.107	0.215
15	0.000	0.217	0.002	0.248	0.009	0.262	0.057	0.247
16	0.000	0.230	0.001	0.260	0.006	0.278	0.032	0.273
17	0.000	0.243	0.001	0.269	0.004	0.292	0.020	0.293
18	0.000	0.256	0.001	0.279	0.003	0.303	0.013	0.308
19	0.000	0.265	0.000	0.287	0.002	0.310	0.008	0.321
20	0.000	0.272	0.000	0.298	0.001	0.318	0.006	0.332
21	0.000	0.277	0.000	0.307	0.001	0.324	0.004	0.340
22	0.000	0.280	0.000	0.314	0.001	0.331	0.003	0.346
23	0.000	0.282	0.000	0.319	0.001	0.339	0.002	0.351
24	0.000	0.286	0.000	0.325	0.001	0.350	0.002	0.360
25	0.000	0.290	0.000	0.330	0.000	0.360	0.002	0.369
26	0.000	0.296	0.000	0.336	0.000	0.369	0.001	0.382
27	0.000	0.302	0.000	0.341	0.000	0.377	0.001	0.394
28	0.000	0.308	0.000	0.346	0.000	0.384	0.001	0.405
29	0.000	0.313	0.000	0.350	0.000	0.387	0.001	0.413
30	0.000	0.318	0.000	0.354	0.000	0.391	0.001	0.419

Notes: $t-s$ denotes the months elapsed since program start. $\hat{Q}_{t-s}$ and $\hat{E}_{t-s}$ are the simulated means of the participation and employment probability, respectively.

Table 9: ATT of Planned Program Duration of Three Months versus Six Months

	Male West		Female West		Male East		Female East	
$t-s$	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.007	-0.000	0.008	-0.000	0.007	0.000	0.007
3	0.011	0.009	0.007	0.009	0.006	0.008	0.004	0.009
4	0.028	0.010	0.020	0.011	0.019	0.011	0.012	0.012
5	0.047	0.012	0.038	0.013	0.045	0.013	0.024	0.014
6	0.051	0.013	0.046	0.014	0.054	0.014	0.032	0.016
7	0.039	0.013	0.037	0.015	0.040	0.015	0.025	0.017
8	0.025	0.014	0.020	0.015	0.018	0.015	0.011	0.017
9	0.011	0.014	0.008	0.016	0.004	0.016	-0.003	0.017
10	-0.002	0.014	-0.003	0.016	-0.004	0.016	-0.011	0.018
11	-0.013	0.014	-0.013	0.016	-0.009	0.016	-0.016	0.018
12	-0.020	0.014	-0.021	0.016	-0.015	0.016	-0.020	0.019
13	-0.026	0.015	-0.028	0.016	-0.023	0.017	-0.024	0.019
14	-0.030	0.015	-0.033	0.016	-0.029	0.017	-0.028	0.019
15	-0.031	0.014	-0.036	0.016	-0.034	0.017	-0.030	0.020
16	-0.031	0.015	-0.038	0.016	-0.035	0.017	-0.030	0.020
17	-0.028	0.015	-0.037	0.016	-0.032	0.017	-0.026	0.020
18	-0.026	0.015	-0.037	0.017	-0.030	0.017	-0.023	0.020
19	-0.027	0.015	-0.038	0.017	-0.031	0.017	-0.023	0.021
20	-0.030	0.015	-0.041	0.017	-0.035	0.017	-0.026	0.021
21	-0.034	0.015	-0.044	0.017	-0.040	0.017	-0.030	0.021
22	-0.038	0.015	-0.046	0.017	-0.045	0.017	-0.034	0.021
23	-0.041	0.015	-0.049	0.017	-0.049	0.017	-0.037	0.021
24	-0.043	0.016	-0.050	0.017	-0.051	0.017	-0.039	0.021
25	-0.044	0.016	-0.051	0.017	-0.053	0.018	-0.040	0.021
26	-0.044	0.016	-0.051	0.017	-0.054	0.018	-0.040	0.022
27	-0.044	0.016	-0.051	0.017	-0.055	0.018	-0.039	0.022
28	-0.042	0.016	-0.051	0.017	-0.055	0.018	-0.038	0.022
29	-0.041	0.016	-0.050	0.017	-0.054	0.019	-0.037	0.022
30	-0.039	0.016	-0.049	0.017	-0.054	0.019	-0.036	0.022

Notes: $t-s$ denotes the months elapsed since program start.

Table 10: ATT of Planned Program Duration of Nine Months versus Six Months

	Male West		Female West		Male East		Female East	
$t-s$	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.006	0.000	0.007	-0.000	0.007	-0.000	0.007
3	-0.001	0.008	-0.001	0.009	-0.000	0.008	-0.000	0.009
4	-0.004	0.009	-0.002	0.010	-0.001	0.009	-0.001	0.010
5	-0.010	0.010	-0.006	0.011	-0.003	0.010	-0.002	0.011
6	-0.018	0.011	-0.014	0.012	-0.009	0.011	-0.008	0.012
7	-0.026	0.012	-0.022	0.013	-0.020	0.012	-0.018	0.014
8	-0.041	0.013	-0.027	0.015	-0.035	0.014	-0.039	0.015
9	-0.044	0.013	-0.034	0.015	-0.046	0.015	-0.045	0.017
10	-0.038	0.014	-0.032	0.016	-0.043	0.015	-0.038	0.017
11	-0.025	0.014	-0.021	0.016	-0.027	0.015	-0.021	0.017
12	-0.012	0.014	-0.009	0.016	-0.014	0.015	-0.007	0.018
13	-0.001	0.014	0.002	0.016	-0.007	0.015	0.003	0.018
14	0.009	0.014	0.011	0.016	-0.002	0.016	0.009	0.018
15	0.016	0.015	0.019	0.016	0.004	0.016	0.014	0.019
16	0.020	0.014	0.025	0.016	0.010	0.016	0.019	0.019
17	0.022	0.014	0.029	0.016	0.015	0.016	0.023	0.019
18	0.023	0.014	0.032	0.016	0.019	0.017	0.024	0.019
19	0.022	0.014	0.034	0.016	0.019	0.017	0.023	0.019
20	0.020	0.014	0.034	0.016	0.017	0.017	0.020	0.019
21	0.020	0.014	0.035	0.016	0.015	0.017	0.016	0.020
22	0.021	0.014	0.036	0.016	0.016	0.018	0.016	0.020
23	0.023	0.015	0.038	0.016	0.019	0.018	0.020	0.020
24	0.027	0.015	0.040	0.016	0.023	0.017	0.025	0.020
25	0.030	0.015	0.043	0.017	0.028	0.017	0.030	0.021
26	0.033	0.015	0.045	0.017	0.031	0.017	0.034	0.021
27	0.034	0.015	0.046	0.016	0.033	0.018	0.036	0.021
28	0.035	0.015	0.047	0.017	0.035	0.018	0.038	0.021
29	0.035	0.016	0.047	0.017	0.036	0.018	0.038	0.021
30	0.034	0.015	0.048	0.017	0.036	0.018	0.037	0.021

Notes: $t-s$ denotes the months elapsed since program start.

Table 11: ATT of Planned Program Duration of Twelve Months versus Six Months

	Male West		Female West		Male East		Female East	
$t-s$	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.007	-0.000	0.007	-0.000	0.007	-0.000	0.007
3	-0.001	0.008	-0.001	0.009	-0.000	0.008	-0.000	0.008
4	-0.004	0.009	-0.002	0.010	-0.001	0.009	-0.001	0.010
5	-0.011	0.010	-0.006	0.011	-0.003	0.010	-0.002	0.011
6	-0.020	0.011	-0.015	0.012	-0.010	0.011	-0.008	0.012
7	-0.030	0.012	-0.024	0.013	-0.021	0.012	-0.019	0.014
8	-0.049	0.013	-0.031	0.015	-0.037	0.014	-0.042	0.016
9	-0.060	0.014	-0.046	0.016	-0.055	0.015	-0.053	0.017
10	-0.065	0.014	-0.056	0.017	-0.067	0.016	-0.059	0.018
11	-0.063	0.015	-0.061	0.017	-0.073	0.016	-0.060	0.019
12	-0.054	0.015	-0.057	0.017	-0.074	0.017	-0.056	0.020
13	-0.039	0.016	-0.044	0.018	-0.065	0.017	-0.043	0.020
14	-0.020	0.016	-0.024	0.018	-0.043	0.017	-0.020	0.020
15	-0.003	0.016	-0.005	0.019	-0.026	0.018	-0.001	0.021
16	0.010	0.016	0.012	0.019	-0.012	0.019	0.014	0.022
17	0.020	0.016	0.026	0.019	-0.003	0.020	0.023	0.023
18	0.026	0.016	0.036	0.019	0.005	0.020	0.030	0.024
19	0.028	0.016	0.044	0.019	0.011	0.020	0.034	0.024
20	0.028	0.016	0.049	0.019	0.013	0.020	0.034	0.023
21	0.028	0.016	0.052	0.019	0.014	0.020	0.033	0.024
22	0.028	0.016	0.054	0.019	0.014	0.020	0.032	0.024
23	0.030	0.016	0.057	0.019	0.015	0.020	0.033	0.025
24	0.032	0.017	0.060	0.019	0.017	0.021	0.035	0.025
25	0.036	0.017	0.063	0.019	0.021	0.020	0.039	0.026
26	0.041	0.017	0.068	0.019	0.028	0.021	0.046	0.026
27	0.045	0.017	0.072	0.020	0.035	0.021	0.053	0.026
28	0.049	0.017	0.075	0.020	0.041	0.021	0.059	0.026
29	0.050	0.018	0.077	0.020	0.045	0.021	0.063	0.026
30	0.051	0.018	0.079	0.020	0.047	0.022	0.065	0.026

Notes: $t-s$ denotes the months elapsed since program start.

Table 12: Classical ATT Aligned to Program Start *(based on Pooled Probit)*

	Male West		Female West		Male East		Female East	
$t-s$	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	-0.098	0.006	-0.074	0.006	-0.097	0.007	-0.055	0.007
2	-0.158	0.008	-0.125	0.008	-0.149	0.009	-0.088	0.009
3	-0.154	0.010	-0.120	0.010	-0.163	0.011	-0.096	0.010
4	-0.151	0.011	-0.119	0.012	-0.161	0.011	-0.094	0.012
5	-0.147	0.011	-0.112	0.013	-0.167	0.011	-0.090	0.014
6	-0.137	0.012	-0.108	0.013	-0.166	0.012	-0.095	0.014
7	-0.116	0.012	-0.082	0.012	-0.150	0.012	-0.090	0.014
8	-0.099	0.013	-0.055	0.013	-0.125	0.013	-0.084	0.014
9	-0.084	0.012	-0.039	0.014	-0.112	0.014	-0.075	0.015
10	-0.070	0.012	-0.019	0.015	-0.099	0.015	-0.067	0.015
11	-0.054	0.013	-0.002	0.015	-0.087	0.015	-0.060	0.016
12	-0.043	0.012	0.013	0.017	-0.073	0.015	-0.054	0.018
13	-0.034	0.013	0.030	0.018	-0.062	0.016	-0.038	0.017
14	-0.030	0.015	0.042	0.017	-0.048	0.016	-0.022	0.017
15	-0.025	0.015	0.050	0.016	-0.038	0.017	-0.012	0.018
16	-0.018	0.015	0.060	0.015	-0.026	0.017	0.000	0.019
17	-0.010	0.016	0.069	0.015	-0.016	0.018	0.009	0.019
18	-0.003	0.016	0.076	0.014	-0.004	0.020	0.018	0.019
19	0.007	0.016	0.082	0.014	0.011	0.018	0.028	0.020
20	0.016	0.016	0.089	0.015	0.024	0.019	0.037	0.020
21	0.024	0.016	0.096	0.016	0.036	0.020	0.045	0.020
22	0.029	0.016	0.102	0.017	0.046	0.020	0.051	0.020
23	0.033	0.017	0.107	0.017	0.054	0.020	0.055	0.020
24	0.036	0.016	0.110	0.018	0.058	0.018	0.061	0.021
25	0.037	0.015	0.112	0.019	0.062	0.018	0.065	0.022
26	0.035	0.016	0.114	0.020	0.065	0.019	0.070	0.023
27	0.034	0.016	0.116	0.019	0.066	0.019	0.075	0.024
28	0.033	0.017	0.116	0.018	0.066	0.018	0.081	0.024
29	0.033	0.016	0.116	0.017	0.067	0.020	0.082	0.022
30	0.033	0.016	0.117	0.017	0.070	0.020	0.083	0.023

Notes: $t-s$ denotes the months elapsed since program start.

Table 13: Classical ATT Aligned to Program Start *(Simple Specification)*

	Male West		Female West		Male East		Female East	
$t-s$	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	-0.099	0.006	-0.072	0.006	-0.095	0.007	-0.052	0.007
2	-0.162	0.009	-0.126	0.009	-0.151	0.009	-0.083	0.010
3	-0.162	0.011	-0.123	0.012	-0.161	0.011	-0.087	0.012
4	-0.160	0.012	-0.121	0.013	-0.158	0.013	-0.083	0.014
5	-0.154	0.013	-0.114	0.014	-0.159	0.014	-0.077	0.015
6	-0.143	0.013	-0.108	0.015	-0.151	0.014	-0.079	0.015
7	-0.125	0.014	-0.085	0.016	-0.135	0.015	-0.073	0.016
8	-0.108	0.014	-0.063	0.017	-0.110	0.016	-0.064	0.017
9	-0.089	0.015	-0.043	0.018	-0.095	0.016	-0.054	0.018
10	-0.070	0.015	-0.020	0.019	-0.079	0.017	-0.042	0.018
11	-0.051	0.016	-0.001	0.019	-0.064	0.017	-0.036	0.019
12	-0.036	0.016	0.018	0.019	-0.050	0.018	-0.025	0.019
13	-0.012	0.016	0.041	0.019	-0.027	0.019	-0.004	0.020
14	0.003	0.016	0.058	0.019	-0.009	0.019	0.015	0.021
15	0.015	0.016	0.072	0.020	0.006	0.019	0.030	0.022
16	0.026	0.016	0.084	0.020	0.019	0.019	0.044	0.022
17	0.034	0.016	0.094	0.020	0.030	0.019	0.054	0.023
18	0.040	0.016	0.102	0.020	0.038	0.019	0.063	0.023
19	0.045	0.017	0.108	0.020	0.046	0.019	0.071	0.023
20	0.049	0.017	0.114	0.020	0.052	0.019	0.078	0.023
21	0.053	0.017	0.119	0.020	0.056	0.019	0.083	0.023
22	0.055	0.017	0.124	0.020	0.060	0.019	0.088	0.023
23	0.058	0.017	0.127	0.020	0.064	0.019	0.093	0.023
24	0.061	0.017	0.130	0.020	0.067	0.019	0.097	0.023
25	0.063	0.017	0.134	0.020	0.069	0.020	0.101	0.023
26	0.065	0.017	0.136	0.020	0.071	0.020	0.103	0.024
27	0.066	0.017	0.139	0.020	0.074	0.020	0.107	0.024
28	0.067	0.017	0.141	0.020	0.075	0.020	0.108	0.024
29	0.068	0.018	0.142	0.020	0.076	0.020	0.110	0.024
30	0.069	0.018	0.144	0.021	0.077	0.020	0.111	0.024

Notes: $t-s$ denotes the months elapsed since program start. In the simple specification the employment dynamics and the treatment history are modeled using only: E_{t-1}, $\tau_{E,t}$, t, Q_{t-1}, and *treat*. *treat* is a dummy variable set to 1 if at least one of the Q_{t-j}, $j > 1$ is 1.

Table 14: Classical ATT Aligned to Program Start *(Simple Specification and Pooled Probit Combined)*

	Male West		Female West		Male East		Female East	
$t-s$	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	-0.100	0.007	-0.071	0.005	-0.098	0.008	-0.057	0.007
2	-0.167	0.009	-0.126	0.008	-0.158	0.010	-0.097	0.010
3	-0.172	0.010	-0.120	0.010	-0.176	0.012	-0.107	0.012
4	-0.179	0.012	-0.123	0.011	-0.181	0.011	-0.113	0.014
5	-0.182	0.012	-0.122	0.012	-0.192	0.011	-0.116	0.015
6	-0.181	0.012	-0.123	0.012	-0.194	0.013	-0.125	0.016
7	-0.173	0.013	-0.111	0.014	-0.188	0.014	-0.123	0.016
8	-0.165	0.013	-0.099	0.015	-0.173	0.013	-0.122	0.016
9	-0.155	0.014	-0.087	0.014	-0.166	0.014	-0.118	0.017
10	-0.142	0.015	-0.074	0.016	-0.157	0.016	-0.113	0.016
11	-0.131	0.014	-0.062	0.016	-0.148	0.017	-0.112	0.018
12	-0.120	0.015	-0.051	0.016	-0.139	0.014	-0.108	0.018
13	-0.101	0.015	-0.033	0.017	-0.122	0.014	-0.094	0.019
14	-0.091	0.015	-0.024	0.017	-0.109	0.014	-0.084	0.020
15	-0.079	0.015	-0.014	0.016	-0.096	0.015	-0.076	0.020
16	-0.070	0.015	-0.005	0.016	-0.085	0.016	-0.069	0.020
17	-0.062	0.016	0.003	0.018	-0.075	0.016	-0.062	0.020
18	-0.055	0.017	0.010	0.018	-0.066	0.017	-0.055	0.020
19	-0.048	0.018	0.016	0.017	-0.057	0.017	-0.048	0.020
20	-0.042	0.017	0.022	0.018	-0.050	0.018	-0.042	0.021
21	-0.037	0.017	0.029	0.018	-0.043	0.019	-0.038	0.022
22	-0.033	0.017	0.034	0.020	-0.036	0.019	-0.033	0.021
23	-0.028	0.017	0.040	0.020	-0.030	0.017	-0.028	0.023
24	-0.023	0.018	0.045	0.019	-0.025	0.016	-0.025	0.024
25	-0.018	0.018	0.050	0.019	-0.020	0.017	-0.022	0.025
26	-0.016	0.018	0.055	0.018	-0.017	0.016	-0.018	0.027
27	-0.013	0.018	0.059	0.019	-0.012	0.017	-0.015	0.025
28	-0.012	0.020	0.062	0.019	-0.008	0.019	-0.012	0.025
29	-0.008	0.020	0.065	0.020	-0.007	0.019	-0.009	0.026
30	-0.006	0.020	0.069	0.019	-0.003	0.020	-0.005	0.027

Notes: $t-s$ denotes the months elapsed since program start. In the simple specification the employment dynamics and the treatment history are modeled using only: E_{t-1}, $\tau_{E,t}$, t, Q_{t-1}, and *treat*. *treat* is a dummy variable set to 1 if at least one of the Q_{t-j}, $j>1$ is 1.