



Working Paper Series

Working Paper No. 24-05 • October 2024

Does Ad Hoc Language Training Improve the Economic Integration of Refugees?

Evidence from Germany's Response to the Syrian Refugee Crisis

Moritz Marbach, Ehsan Vallizadeh, Niklas Harder, Dominik Hangartner, and Jens Hainmueller

Does Ad Hoc Language Training Improve the Economic Integration of Refugees? Evidence from Germany's Response to the Syrian Refugee Crisis

Moritz Marbach,^{1,8} Ehsan Vallizadeh,^{3,9} Niklas Harder,^{1,4}
Dominik Hangartner^{1,2,6} Jens Hainmueller,^{1,5,7}

¹Immigration Policy Lab, Stanford University, Stanford, USA, and ETH Zurich, Zurich, Switzerland

²Center for Comparative and International Studies, ETH Zurich, Zurich, Switzerland

³Institute for Employment Research (IAB), Nürnberg, Germany

⁴German Centre for Integration and Migration Research (DeZIM) , Berlin, Germany

⁵Department of Political Science, Stanford University, Stanford, USA

⁶London School of Economics and Political Science, London, United Kingdom

⁷Graduate School of Business, Stanford University, Stanford, USA

⁸Department of Political Science, University College London, London, United Kingdom

⁹University of Bamberg, Bamberg, Germany

(forthcoming in Journal of the Royal Statistical Society: Series A)

*To whom correspondence should be addressed: Jens Hainmueller, Department of Political Science, Stanford University,

Encina Hall West, Room 440, 616 Serra Street, Stanford, CA 94305-6044, USA. E-mail: jhain@stanford.edu

Abstract

Given the global displacement crisis, the integration of refugees has emerged as a critical policy issue for many host countries. A key challenge involves supporting refugees in learning the language of their host country. While several European nations have instituted publicly funded language training for asylum seekers and refugees soon after their arrival, evidence on the efficacy of these early language programs in promoting economic integration remains limited. This study examines the impact of a pioneering, large-scale ad hoc program introduced by German policymakers, which provided basic language training to over 230,000 refugees arriving in 2015-16. Utilizing register data on the population of asylum seekers and exploiting a cutoff date in program eligibility, we assess the program's effectiveness using a regression discontinuity design. Our findings reveal no discernible effect on refugee employment over the subsequent two years. To explore whether language programs are generally ineffective during refugee crises, we contrast these results with the impacts of a more comprehensive, preexisting, yet smaller-scale program. Using a variety of difference-in-differences estimators, we find that this program considerably increased refugee employment. These contrasting findings offer important insights for policymakers on designing effective language training programs for refugees.

Keywords: immigrant integration, refugee language training program, regression discontinuity design

1 Introduction

Over the last decades, policymakers in Europe and across the globe have grappled with major displacement crises that have resulted in a significant increase in the number of refugees fleeing conflict and persecution. Western destination countries face a significant challenge in designing policies and programs that facilitate the integration of refugees into the host country’s economy and society (OECD, 2016; UNHCR, 2013; Council of the European Union, 2014; The Economist, 2015; Degler and Liebig, 2017).

Given that refugees are often unfamiliar with the host country’s language, one of the key policy challenges involves the provision of language training programs (Scholten et al., 2017; Liebig, 2007; Dumont et al., 2016). Language acquisition is often the first step in the successful integration of refugees and serves as a vehicle for finding employment (Dustmann and Fabbri, 2003; Chiswick and Miller, 1995; Cortes, 2004; Dribe and Lundh, 2011; Schönwälder et al., 2005). Proficiency in the host country’s language can facilitate labor market integration through at least three channels. First, language is a fundamental element of human capital (van Tubergen et al., 2004; Dustmann and Fabbri, 2003; Chiswick and Miller, 1995). Second, language complements and enhances the transferability of skills obtained abroad in the host country labor market (Berman et al., 2003; Chiswick and Miller, 2001). Third, employers and refugees in Europe often state that the lack of sufficient language skills is one of the main obstacles to employability, even for low-skilled jobs (Degler and Liebig, 2017; Fasani et al., 2018).

These findings suggest that early investments in language courses for refugees shortly after arrival may subsequently yield substantial returns in terms of improved economic integration (Bach et al., 2017). By improving economic integration, early language training might generate significant economic benefits, not only for refugees, who could more quickly find jobs that match their skill set but also for the host economy, in terms of higher tax contributions from employed refugees and lower welfare expenditures for unemployed refugees.

Between 2015 and 2017, nearly all OECD countries offered publicly financed language programs for refugees. Considerable variety existed, however, regarding the content of the programs and the number of instruction hours. As Figure 1 shows, the maximum course hours ranged from 70 in Croatia to 4,800 or more in Scandinavian countries (Denmark, Norway, and Sweden). Programs also varied on several other dimensions, such as how soon after arrival refugees could enroll in the programs, whether instruction followed a standardized curriculum, whether certificates were offered, and whether there were national standards for course providers (OECD, 2016; Konle-Seidl, 2018; Robert Bosch Stiftung, 2015; Schönwälder et al., 2005).

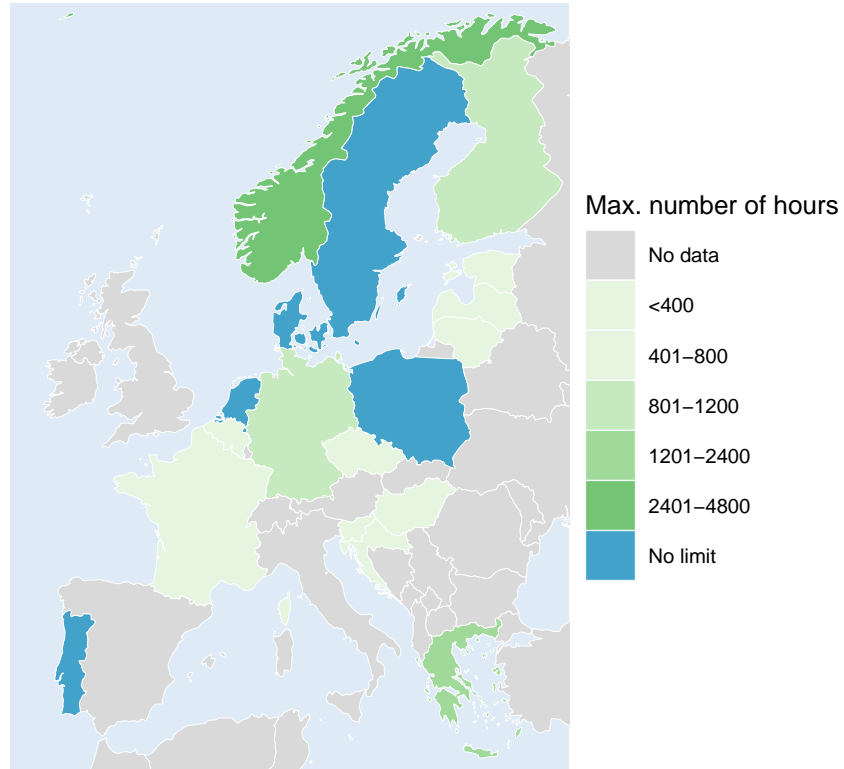


Figure 1: Maximum number of hours of public language courses in European countries around 2017. There is considerable heterogeneity in the maximum number of hours of language instruction refugees receive, ranging from 70 hours in Croatia to an unlimited number of hours in countries such as Denmark and Sweden. Data according to authors' communication with UNHCR and OECD (2016).

But do early language training programs for refugees actually improve their subsequent economic integration? And if so, what quantity and quality of instruction are required to achieve measurable improvements in economic integration? A sizable literature has documented strong links between language skills and improved labor market prospects for immigrants more generally (Chiswick, 1978; Lubotsky, 2007; Cohen and Haberfeld, 2007; Berman et al., 2003; Bleakley and Chin, 2004; Delander et al., 2005; Dustmann and Fabbri, 2003; Hayfron, 2001; Lochmann et al., 2019; Chiswick and Miller, 2001). However, there exists much less evidence on the impact of language training programs for refugees in particular.

Perhaps the studies closest to ours are Foged et al. (2022) and Foged and Van der Werf (2023), which both consider language programs for refugees that arrived in Denmark. Foged et al. (2022) evaluate a policy reform that increased the number of hours of language training and made training mandatory for refugees who arrived in Denmark after January 1, 1999. The policy change was accompanied by some other features introduced with the same cutoff, including a temporary reduction in welfare benefits and a new scheme for placing refugees in municipalities. Using a regression discontinuity design the study finds that the policy as a whole increased the employment probabilities by around 4 percentage points and led to higher earnings. At the cutoff refugees accumulated around 200 hours of additional language training. Foged and Van der Werf (2023) leverages variation in the accessibility of local language training centers in Denmark to examine the impact of language classes for refugees on language fluency. The study finds that an additional 100 hours of language instruction improved refugees Danish language fluency by around 8-9 percent, looking at refugees who arrived from 2003 to 2013. The study also finds that language classes reduced the probability that refugees leave the localities in which they are initially placed.

Our study differs from previous work in that we examine the impact of language training programs on refugees labour market integration in Germany in the context of a major displacement crisis characterized by mass arrivals. In particular, we leverage register population data to provide evidence on the impact of a large-scale language training program that the German government rapidly developed and implemented in response to the large increase in refugees in 2015. We find no discernible effect of this program on refugee employment over the following 22 months. To shed light on the question whether language programs are generally ineffective during refugee crisis, we contrast this null finding with the impacts of a more comprehensive, preexisting yet smaller-scale program. Using a variety of difference-in-differences estimators, we find that this preexisting program increased refugee employment by about between 4-5 percentage points 12 months after enrollment. Our contrasting findings illuminate the nuanced impacts of early language training programs during a period when a

large number of refugees arrived. Understanding these impacts is crucial for informing the design of language policies in Germany and beyond, and for preparing for future crises.

2 Language Training and Refugee Integration

Research has shown that refugees experience significant employment and wage gaps compared to other immigrant groups and the native population, even many years after arrival (Fasani et al., 2018; Cheng et al., 2021). Studying early language training programs for refugees during times of crisis is important, as refugees face many additional challenges compared to immigrants who migrate voluntarily through mechanisms like family reunification or for study and work (Becker and Ferrara, 2019; Liebau and Schacht, 2016; Robert Bosch Stiftung, 2015; Beiser and Hou, 2000).

First, refugee crises are marked by large numbers of arrivals, leading to prolonged asylum processing times. This results in extended uncertainty about the prospects of staying in the host country, overcrowded accommodations, overburdened support programs, higher unemployment rates among refugees, and often, political backlash against new arrivals (Hainmueller et al., 2016; Fasani et al., 2018; Hangartner et al., 2019; Pecoraro et al., 2022).

Second, many refugees have endured traumatic events in their home countries or during their journey to safety. The psychological impact of these experiences can impede their ability to focus on learning a new language, and the stress of resettlement and adjusting to a new culture can further hinder language acquisition (Bjertrup et al., 2018; Schlaudt et al., 2020).

Third, the language of the host country may be vastly different from the refugees' native languages, presenting a significant learning barrier (Asfar et al., 2019; Cheng et al., 2021).

Fourth, refugees often have limited opportunities to practice the language outside formal learning settings. They may reside in reception centers where their native language predominates, or they may face discrimination or social isolation, which can further impede their engagement with the new language (Van Tubergen, 2010).

These factors can significantly obstruct both language acquisition and labor market integration for refugees, underscoring the importance of studying the returns to language training for this group. Early language training might be especially critical given the existing evidence of a formative early window for integration. Studies have demonstrated that interventions shortly after arrival—such as early access to labor markets, voting rights, shorter asylum wait times, or better matching between refugee characteristics and host communities—can durably improve immigrants' subsequent integration trajectory (Hainmueller et al., 2016; Marbach et al., 2018; Ferwerda et al., 2020; Bansak et al., 2018).

3 Setting

In 2015/2016, Germany received over one million asylum seekers, a stark contrast to an average of 71,000 asylum applications per year during the period from 1995 to 2014 (BAMF, 2018). A significant portion of these refugees were escaping conflict and persecution in countries such as Syria, Afghanistan, and Iraq. It is important to note that for the purposes of this study, we use the terms “refugees” and “asylum seekers” interchangeably. Within our estimation sample, most individuals initially enter as asylum seekers but later receive some form of humanitarian protection, thereby transitioning into refugee status.

Prior to the 2015 crisis, Germany relied on a language program offered by the Federal Office for Migration and Refugees (BAMF). This program, called Integration Course (*Integrationskurs*) has been in operation since 2005 and is rather comprehensive in scope, covering up to 600 hours of instruction with a standardized curriculum. Between January 2015 and January 2017, approximately 500,000 individuals participated in the BAMF courses (Deutscher Bundestag, 2016). However, in the context of the mass arrivals of refugees during the same time period, this preexisting program struggled to meet the demand during the refugee crisis. Since scaling up the preexisting program proved challenging, policymakers established a novel, ad hoc language training program called Introductory German Language Course (*Einstiegskurs zur Deutschförderung*), administered by the Federal Employment Agency (BA).

This program was ambitious in scale and designed to promptly accommodate a large number of eligible participants from Syria, Iraq, Iran, and Eritrea. It took a less intensive approach, covering only 320 hours of instruction and lacking a standardized curriculum, and was rolled out rapidly. The first course began less than four weeks after announcement, in late October 2015. In total, about 230,000 refugees, approximately 38% of the eligible arrival population in 2015, enrolled in this program (Bundesrechnungshof, 2017), see Figure 2. The cost amounted to about 400 million Euros, or approximately 4.8 Euros per participant per training hour. Additional details on the two language training programs are provided in the Supplementary Materials (SM).

Comparing the ad hoc program to the distribution of course hours in other European countries (see Figure 1), we observe that the ad hoc program, with 320 hours, ranks towards the bottom of the distribution. In contrast, the preexisting course falls within the modal category of 600 hours. Despite being less intensive than the preexisting language course, the ad hoc program shared the goal of integrating refugees into the labor market (BA Presseteam, 2015; BAMF, 2015). However, there is scant evidence that this shift in design and delivery of language programs succeeded.

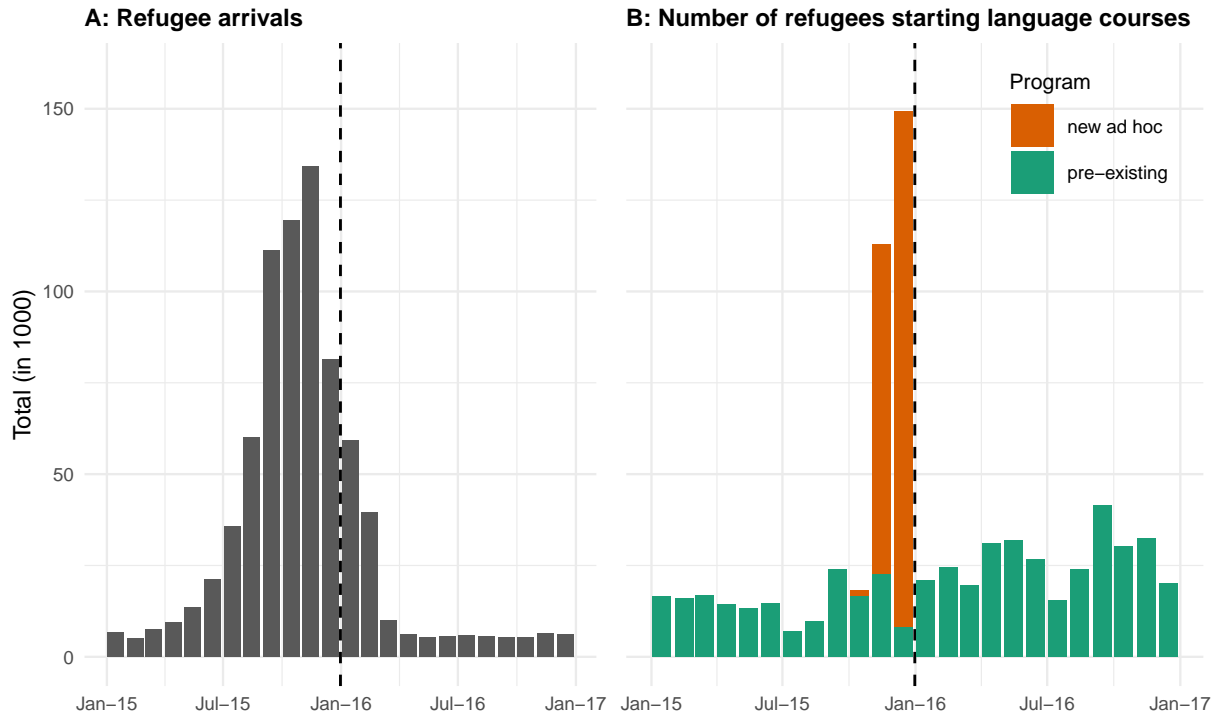


Figure 2: The left panel displays monthly registrations of refugee arrivals from Syria, Iraq, Iran, and Eritrea via the EASY-System (*Erstverteilung der Asylbegehrenden-System*) between January 2015 and January 2017. The right panel shows the number of refugees participating in the ad hoc language training program (orange bars). The vertical dashed line marks the cutoff date (December 31, 2015) for ad hoc program eligibility. For comparison, total enrollment in the preexisting language program is also displayed (green bars), with the last six months estimated based on the total number of courses starting. Sources: Federal Office for Migration and Refugees (BAMF) and Federal Employment Agency (BA).

4 Data, Measures, and Statistical Analysis

4.1 Data

We use monthly panel data compiled from the Integrated Employment Biographies (IEB) to study the impact of both programs on labour market integration. We combine the IEB data with another administrative individual data *Statushistorik Zuwanderung* (SHZ) that contains detailed information on socio-demographic characteristics and important migration-related information. This data is based on administrative records from the BA and allows us to cover almost all refugees who arrived in Germany.

We focus on refugees from Syria, Iran, Iraq, and Eritrea, with a observed and registered arrival date between June 2015 and June 2016, and ages 18 – 35 at the time of arrival. Note that at the time refugees from these four origin countries who applied for asylum in Germany were considered to have a high probability of getting protection status and therefore eligible to participate in the language training programs and eligible to work.

To examine the effects of the ad hoc language training program we leverage the data on 210,369 refugees which we observe over approximately two years for a total of 5,487,256 person-months. Note that for about 96.8% of all refugees in the data, we have a record up until the last month in the panel, but about 3.2% drop before the end of the panel. In the robustness section below we examine the sensitivity of our results to this panel attrition.

To examine the effects of the comprehensive training program we use a random person sample of the population which includes 50,000 refugees (1,335,073 person-months). To enhance the comparability of estimates between the preexisting and the ad hoc programs we restrict this random sample to individuals who were eligible for the ad hoc program (i.e., individuals who arrived before December 31, 2015). The reason why we use a random sample is to reduce computational burden for the panel models.

4.2 Measures

Our main outcome measure of labor market integration is the binary variable ‘Has job’ which measures employment status. It is coded as one for refugees when they are in full-time or marginal employment in a given month and zero otherwise. Note that marginal employment includes all part-time, low-wage jobs (including low-paying “minijobs,” apprenticeships, and paid internships). We also use two alternative outcomes. ‘Has job (w/o minijob)’ focuses only on full time employment and does not include any marginal employment. ‘Has minijob’ focuses only on marginal employment, respectively.

We also observe a limited set of covariates, including the age at arrival, a binary variable measuring education that takes the value of one if the individual completed more than primary school and zero otherwise, origin, and gender. More details on the data and measures is provided in the SM.

4.3 Descriptive Statistics

Table 1 shows descriptive statistics for the estimation samples. The columns on the left show the full population we use to examine the effects of the ad hoc program. The average refugee is 25 years old, the share of females is 25%, and the nationality shares are 73% Syrian, 16% Iraqi, 6% Iranian, and 5% Eritrean. 60% have some schooling.

The mean of the variable measuring enrollment in the preexisting language training program is only 2%. This low participation rate in the preexisting program demonstrates the limited capacity of this program to serve the large refugee population at the time and was the motivation for creating the large scale ad hoc program.

The mean for the ‘has job’ outcome is only 7% indicating that most refugees are unemployed. The columns on the right show the random person sample of the population eligible for the ad-hoc program we use to examine the effects of the preexisting program. This sample mirrors the full population we use to examine the effects of the ad hoc program with only very limited differences across covariates.

4.4 Identification and Estimation Strategy

4.4.1 RDD for Ad Hoc Program

To identify the effects of the ad hoc program, we leverage a regression discontinuity design (RDD) based on the criterion that only refugees from Syria, Iran, Iraq, and Eritrea who received their registration on or before December 31, 2015, were eligible to participate. This allows us to isolate the effect of program eligibility by comparing otherwise similar refugees who arrived just before this cutoff date to those who arrived just afterward in January 2016 and were ineligible. Note that refugees in our sample all had a strong incentive to register as soon as possible after their arrival in Germany in order to start the asylum application review and get access to rights and benefits. Therefore it seems unlikely that refugees would try to manipulate their arrival date around the cutoff. Indeed, results from placebo tests indicate that the covariates are well balanced at the threshold, consistent with a quasi-random assignment of eligibility around the cutoff date (see SM Table S.2 and Figure S.1).

To estimate the effect of the ad hoc program, we use a local linear regression with a triangular kernel of the form:

$$Y_{it} = \alpha + \beta_1 Z_i + \delta D_i + \beta_2 (D_i * Z_i) + \epsilon_{it}$$

where Y_{it} measures the employment of refugee i in a specific month t after arrival, Z_i is the running variable that measures the distance between the day of arrival and the eligibility cutoff date (December 31, 2015), D_i is a treatment indicator coded one if the arrival was before the cutoff date and zero otherwise, and ϵ_{it} is the error term. In this regression, δ identifies the intention-to-treat (ITT) effect of program eligibility on employment.

Note that participation in the ad hoc courses is not observed in our register data, and therefore our estimation focuses on ITT effects. However, using a secondary dataset, we can also approximate the proportion of compliers close to the cutoff date to examine how the eligibility cutoff affected program uptake (see SM Figure S.2).

We run separate regressions to measure the impact on Y_{it} for months $t = 6, 7, \dots, 22$ after arrival. Since the program duration was about 8 weeks, we would expect effects to emerge at the earliest after this initial lock-in period (although refugees were, in principle, allowed to work even while completing the program). For the main specification, we use the common Mean Squared Error (MSE)-optimal bandwidth based on an automated selection algorithm (Calonico et al., 2014) and a heteroskedasticity-robust plug-in residuals variance estimator without weights.

4.4.2 DID for Preexisting Program

To identify the effects of the preexisting language program, we leverage the panel dimension of our data, as there is no discontinuity in program eligibility. Hence, we employ a difference-in-differences (DID) design, comparing refugees who enrolled in the preexisting program at different points in time.

We leverage two heterogeneous treatment effect robust estimators to identify the average treatment effect on the treated (ATT), including the difference-in-differences imputation estimator with interactive fixed effects proposed by Liu et al. (2024) (also see Borusyak et al. (2024)) and the doubly robust difference-in-differences estimator proposed by Callaway and Sant’Anna (2021). Both estimators address recent concerns regarding the traditional two-way fixed effects linear regression estimator, which does not converge to a convex combination of treatment effects in the presence of dynamic treatment effect heterogeneity (Callaway and Sant’Anna, 2021; De Chaisemartin and d’Haultfoeuille, 2020; Goodman-Bacon, 2021).

Let $D_{i,t}$ be a treatment indicator coded 1 if refugee i has ever enrolled in the preexisting language program at month t , and zero otherwise. Once a refugee enrolls, he or she is considered treated for the rest of the study period, i.e., $D_{i,t} = 1 \Rightarrow D_{i,t+1} = 1$ for $t = 1, 2, \dots, T$

(staggered adoption). Let $Y_{i,t}(1)$ and $Y_{i,t}(0)$ denote the potential outcomes of refugee i under the treatment and control conditions.

The core idea of the difference-in-differences imputation estimator is to estimate a flexible model for the counterfactual outcome $Y_{i,t}(0)$ in the control observations $D_{i,t} = 0$ and then use the fitted model to impute the missing potential outcomes $\hat{Y}_{i,t}(0)$ for each treated observation $D_{i,t} = 1$. One can then compute for the treated observations the individual treatment effects as $\hat{\tau}_{i,t} = Y_{i,t}(1) - \hat{Y}_{i,t}(0)$, the difference between the observed $Y_{i,t}(1)$ and the imputed counterfactual outcomes $\hat{Y}_{i,t}(0)$, and then average these to obtain the ATTs for each period $s > 0$ following the treatment onset (Liu et al., 2024):

$$ATT(s) = E[\tau_{i,t} | D_{i,t-s} = 0, D_{i,t-s+1} = D_{i,t-s+2} = \dots = D_{i,t} = 1, \forall i \in T]$$

This method avoids the non-convex weighting problem because treated observations are never used as controls and allows for heterogeneous treatment effects because missing counterfactual outcomes are imputed for each treated observation. Note that the preexisting language program runs for about six months and enrollment is typically full-time. Therefore, we expect treatment effects to materialize, if at all, after around six months.

To model the counterfactual outcome, we leverage the difference-in-differences imputation estimator with interactive fixed effects proposed by Liu et al. (2024). In this approach, we assume a factor-augmented model for untreated potential outcomes given by

$$Y_{i,t}(0) = \alpha + \gamma_i + \xi_t + \lambda_i' f_t + \epsilon_{it}$$

for all i, t . In this model, γ_i and ξ_t are refugee and calendar month fixed effects that control for time-invariant refugee characteristics (such as education, employment experience acquired in the origin country, or time-invariant cognitive skills and personality traits) and common shocks that vary at the month level. The model also includes an interactive fixed effects component given by $\lambda_i' f_t$, where $f_t = [f_{t,1}, \dots, f_{t,r}]$ is a $(r \times 1)$ vector of unobserved common factors and $\lambda_i = [\lambda_{i,1}, \dots, \lambda_{i,r}]$ is a $(r \times 1)$ vector of unknown factor loadings (Bai, 2009).

The idea of the interactive fixed effects component is to capture time-varying unobserved confounders by allowing a set of refugee-specific fixed effects to interact with time-varying factors. For example, an unobserved common shock is allowed to have a heterogeneous impact on each refugee or alternatively, the effects of unobserved time-invariant refugee confounders can change over time. We allow for two factors ($r = 2$), such that the interactive fixed component is modeled by $\lambda_{i,1} f_{t,1} + \lambda_{i,2} f_{t,2}$. Following Bai (2009), we estimate this model using an iterative algorithm that applies a factor analysis to the residuals from a linear model and then re-estimates the linear model while incorporating the influence from a fixed number of

the most important factors. To obtain uncertainty estimates, we use a nonparametric block bootstrap clustered at the refugee level (Liu et al., 2024).

The key assumption in this approach is that the model for the counterfactual outcome is correctly specified under a strict exogeneity assumption that states that the error is independent of treatment assignment and the unobserved temporal and cross-sectional heterogeneities across all refugees and time periods. This implies a conditional parallel trends assumption conditional on the refugee, month, and interactive fixed effects.

As an alternative approach, we also use the doubly robust CS DR difference-in-differences estimator proposed by Callaway and Sant’Anna (2021). Let $G_{i,g}$ be a cohort indicator, coded as 1 if refugee i is first treated at time g , and zero otherwise. Cohorts are defined as groups of refugees who enroll in the same month. Let C be an indicator for a never-treated group, coded as 1 for refugees who never enroll and zero otherwise. The idea of this approach is to identify cohort-time specific ATTs given by:

$$ATT(g, t) = E[Y_t(g) - Y_t(0) | G_g = 1]$$

for $t \geq g$. As shown in Callaway and Sant’Anna (2021), this estimand is analogous to a doubly robust augmented inverse probability weighting estimand given by:

$$ATT_{DR}(g, t) = E \left\{ \left(\frac{G_g}{E[G_g]} - \frac{\frac{p_{g,t}(X)(1-D_t)}{1-p_{g,t}(X)}}{E \left[\frac{p_{g,t}(X)(1-D_t)}{1-p_{g,t}(X)} \right]} \right) (Y_t - Y_{g-1} - \mu_{g,t}^0(X)) \right\}$$

where $\mu_{g,t}^0(X) = E[Y_t - Y_{g-1} | X, D_t = 0, G_g = 0]$ is the counterfactual based on the comparison group of never-treated units, and $p_{g,t}(X) = P(G_g = 1 | X, G_g + (1 - D_t)(1 - G_g))$ is the propensity score for treatment assignment. We summarize the cohort-specific effects by computing the weighted average treatment effect for groups of refugees who have been exposed to the treatment for a given number of months. We remove the first six calendar months since there are too few observations.

The key assumption for the doubly robust approach is that either the outcome model or the treatment assignment model is correctly specified. For the outcome model, this assumes the conditional parallel trends assumption, stating that the counterfactual evolution of the outcome in the treatment cohorts follows that of the comparison group to estimate the cohort-time average treatment effects. For the treatment assignment model, this assumes that we correctly model the conditional probability of a refugee being in a treatment cohort given their covariates (Callaway and Sant’Anna, 2021; Sant’Anna and Zhao, 2020).

| Variable | Ad-hoc program | | | | Preexisting program | | | |
|-----------------------------|----------------|-------|------|-----|---------------------|-------|-----|-----|
| | Mean | SD | Min | Max | Mean | SD | Min | Max |
| Age | | | | | | | | |
| In years | 25.20 | 4.91 | 18 | 35 | 25.15 | 4.92 | 18 | 35 |
| 18-20 | 0.22 | 0.41 | 0 | 1 | 0.22 | 0.42 | 0 | 1 |
| 21-25 | 0.33 | 0.47 | 0 | 1 | 0.33 | 0.47 | 0 | 1 |
| 26-30 | 0.27 | 0.44 | 0 | 1 | 0.27 | 0.44 | 0 | 1 |
| 31-35 | 0.18 | 0.38 | 0 | 1 | 0.18 | 0.38 | 0 | 1 |
| Gender | | | | | | | | |
| Female | 0.25 | 0.43 | 0 | 1 | 0.23 | 0.42 | 0 | 1 |
| Nationality | | | | | | | | |
| Syrian | 0.73 | 0.44 | 0 | 1 | 0.74 | 0.44 | 0 | 1 |
| Iraq | 0.16 | 0.37 | 0 | 1 | 0.15 | 0.36 | 0 | 1 |
| Iran | 0.06 | 0.24 | 0 | 1 | 0.05 | 0.22 | 0 | 1 |
| Eritrea | 0.05 | 0.22 | 0 | 1 | 0.05 | 0.23 | 0 | 1 |
| Schooling | | | | | | | | |
| No | 0.26 | 0.44 | 0 | 1 | 0.25 | 0.43 | 0 | 1 |
| Yes | 0.59 | 0.49 | 0 | 1 | 0.60 | 0.49 | 0 | 1 |
| (Missing) | 0.16 | 0.36 | 0 | 1 | 0.15 | 0.36 | 0 | 1 |
| Employment | | | | | | | | |
| Has job | 0.07 | 0.26 | 0 | 1 | 0.08 | 0.27 | 0 | 1 |
| Has job (w/o minijob) | 0.04 | 0.20 | 0 | 1 | 0.04 | 0.20 | 0 | 1 |
| Has minijob | 0.04 | 0.19 | 0 | 1 | 0.04 | 0.19 | 0 | 1 |
| Others | | | | | | | | |
| Running Variable | 78.36 | 67.07 | -152 | 212 | 95.18 | 51.55 | 0 | 212 |
| Residency | 12.86 | 7.91 | 0 | 30 | 13.14 | 8.03 | 0 | 30 |
| Enrolled pre-existing prog. | 0.02 | 0.15 | 0 | 1 | 0.03 | 0.16 | 0 | 1 |

Table 1: Descriptive statistics for monthly panel data. The data for the ad-hoc program covers 210,369 persons aged 18-35 arriving from Syria, Eritrea, Iraq or Iran ($N = 5,487,256$). The data for the preexisting program is a random person sample from this data covering 50,000 persons $N = 1,335,073$ that are eligible for the ad-hoc program.

5 Results: Effects of the Ad Hoc Language Program

Figure 3 shows the regression discontinuity design estimates of the effect of the ad hoc program on refugee employment. Panel A of Figure 3 shows the average employment rates as a function of the distance between the arrival date and the program eligibility cutoff of December 31, 2015. If participation in the ad hoc language program led to increased employment, we would expect a higher rate of employment among those refugees who arrived right before the cutoff date and were therefore eligible for the ad hoc courses compared to those refugees who arrived right after that date and were therefore not eligible. However, we find that there is no drop in the average employment rates among those who arrived right after the cutoff date compared to those who arrived right before. This indicates that the ad hoc program had no discernible positive effect on enhancing the labor market participation of refugees for up to 22 months after arrival. Note that this null finding is not due to low program take-up among the eligible refugees. Indeed, as we show in the SM, the estimated program participation rate drops sharply by about 21 percentage points at the cutoff (see Figure S.2).

Panel B in Figure 3 shows the estimation results from the local linear regression, which identifies the employment effects of the ad hoc program at the cutoff date, measured from six to up to 22 months after arrival. The estimated optimal bandwidth is on average 28 days around the cutoff date (the underlying estimates are also presented in Table S.3 in the SM). Consistent with the graphs in Panel A, we find that the program had no discernible positive effect on average employment rates. The point estimate for the effect 7 months after arrival is 0.03 with a narrow 95% confidence interval ranging from -0.34 to 0.40. For 12 months after arrival, the effect estimate is -0.00 with a 95% confidence interval of -0.75 to 0.75. For longer follow-up periods, the effects, if anything, turn more negative but also less precise. For month 17, the estimate is -0.92 with a 95% confidence interval from -2.17 to 0.33, and for month 22, the estimate is -2.01 with a confidence interval from -4.14 to 0.12. Note that for two of the months, 19 and 20, the effect estimates are negative and significant at the 95% confidence level, but the confidence intervals for those two months overlap with those of the null effect estimates for the other months, and this negative effect is also not robust across alternative specifications (see below). Given the totality of the evidence we interpret this as a null finding and clear evidence against a positive effect of the program, but we refrain from interpreting this as a negative effect.

To gauge whether these null effects are due to a lack of statistical precision of our regression discontinuity design, we conduct a series of inferiority tests (one-sided t-tests of the null hypothesis that an effect at least as large as a pre-specified threshold can be rejected). For

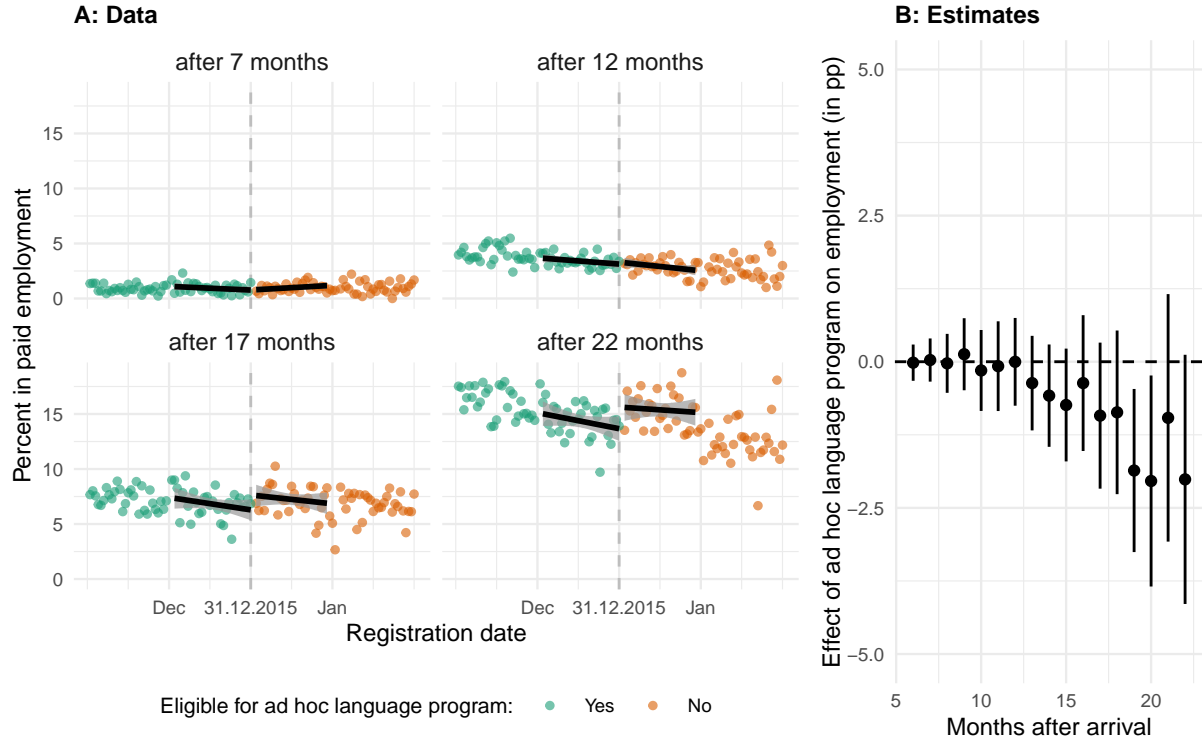


Figure 3: Estimates from regression discontinuity design show that eligibility for participation in the ad hoc language program had no discernible effect on employment outcomes six to 22 months after arrival. Left panel shows the average employment rates for various months after arrival for refugees conditional on their day of arrival. The dashed vertical line is the eligibility cut-off date (Dec 31, 2015). Green dots are daily averages for refugees who arrived before the cut-off and were thus eligible for the ad hoc program, orange dots are daily averages for refugees who arrived after the cut-off and were thus not eligible for the ad hoc program. Black solid lines are the fitted local linear regressions using a symmetric 28 day bandwidth around the cut-off date. Left panel shows point estimates and 95% confidence intervals from the local linear regressions with (MSE)-optimal bandwidth that estimate the intention-to-treat effect of program eligibility at the cut-off date.

all months except 21 after arrival, we can reject effect sizes of 1.5 percentage points or larger with p -values < 0.05 . Together, these tests suggest that we can rule out all but very small benefits of the program.

In the SM, we present results from various robustness checks that corroborate the main findings. First, we replicate the models to estimate the effects by employment type: all jobs, marginal employment, and jobs excluding marginal employment. We find no discernible effects across all employment types (Figure S.4). Second, we replicate the models but also adjust for covariates (age, nationality, schooling, and sex), and the estimates are similar, as expected given that they are balanced across the eligibility threshold (Figure S.5). Third, we check if the estimates are affected by attrition and replicate the main models while imputing a zero for all years post-arrival without an observed employment outcome (e.g. due to emigration out of Germany or death). The idea behind this imputation strategy is that any form of formal employment should be observed in our register data. We again find that the null effects are robust to this adjustment (Figure S.6). Fourth, we find that the null findings are not the result of a specific bandwidth or model specification (Figure S.7). Fifth, we find that the effects are robust when we use a (one-sided) donut regression discontinuity design that excludes observations to the left of the threshold (Figure S.8). Sixth, we check whether the null findings might be driven by the fact that those who are ineligible for the ad hoc program may subsequently enroll in the preexisting comprehensive training program. To do so, we replicate the models while excluding all individuals that at some point during our study period (also) enroll in the pre-existing program. Again, we find no discernible effects for the ad hoc program (Figure S.9). We also find no significant effect of the eligibility for the ad hoc program on the probability of enrolling in the pre-existing program (Figure S.10). Seventh, we estimate the results separately for men and women, and the results are similar, with slightly more variability for men (Figure S.11). Eighth, we estimate the ad hoc program broken down by refugees' state of residence at arrival. We find that there is some heterogeneity, but the distribution of effects is centered on zero and the null effects are robust across the large majority of states (Figure S.12). Lastly, we replicate the models broken down by refugees' education levels; we find some limited heterogeneity but the distribution of estimates centered on zero (Figure S.13).

The results so far have shown that the ad hoc, large-scale language training program rapidly established by the German government in response to the surge in refugees did not have any discernible impacts on the labor market integration of refugees on average. This raises two important questions: Are these null effects a consequence of specific features of this ad hoc program? Or do these findings indicate more general and systemic limitations of language programs' potential to foster economic integration during a large scale displacement

crisis? To begin addressing these questions, we contrast the estimates from the ad hoc program with the effects of the much smaller scale preexisting language program during the same time period and in the same context.

6 Results: Effects of Preexisting Language Program

Figure 4 displays the effect estimates for two estimators: the difference-in-differences imputation estimator with interactive fixed effect (top panel) and the doubly robust CS DR difference-in-differences estimator (bottom panel). The effects are shown for three outcomes: all jobs (left), all jobs but minijobs (middle), and minijobs (right). The underlying estimates are also presented in Table S.4 and S.5 in the SM.

The results indicate that enrolling in the preexisting language course significantly increased employment following the six-month course duration. This result holds across both estimators. For instance, at month 12 following enrollment in the course, the estimated effect from the imputation estimator on the ‘all jobs’ outcome is an increase of 4.8 percentage points with a standard error of 0.9. The corresponding estimate from the CS DR estimator is an increase of 4.4 percentage points with a standard error of 0.9. The results in the middle and left panels suggest that the increase in employment in later periods is not primarily driven by minijobs. While the point estimates are similar up until month 14 (imputation estimator) and 15 (CS DR estimator) for the minijob outcome and the all jobs but minijobs outcome, the effects for the all jobs but minijob outcome continues to grow while it declines for the minijob outcome.

Consistent with the parallel trends assumption, there are no large differences in employment rates in the months prior to enrollment for all three outcomes. In Table S.6 in the SM, we present F-tests for three selected pre-treatment periods (3 to 1 month before enrollment, 9 to 1 month, and 29 to 1 month) and all three employment outcomes. We observe one statistically significant difference in pre-treatment estimates across the three employment outcomes but the substantive magnitude of this difference is small, as shown in Figure 4.

In the SM, we present several robustness checks that corroborate the main results. First, we re-estimate the CS DR estimator using the not-yet and the never-treated units as controls, and the estimates are very similar (Figure S.14). Second, we examine whether the results are affected by panel attrition by re-estimating the models while imputing any missing outcomes as zero, and the results are again similar (Figure S.15). Third, we re-estimate the models separately for male and female refugees. The effects are roughly similar across both males and females, with the estimates for women being somewhat elevated in magnitude but with larger confidence intervals due to the reduced estimation sample size (Figure S.16).

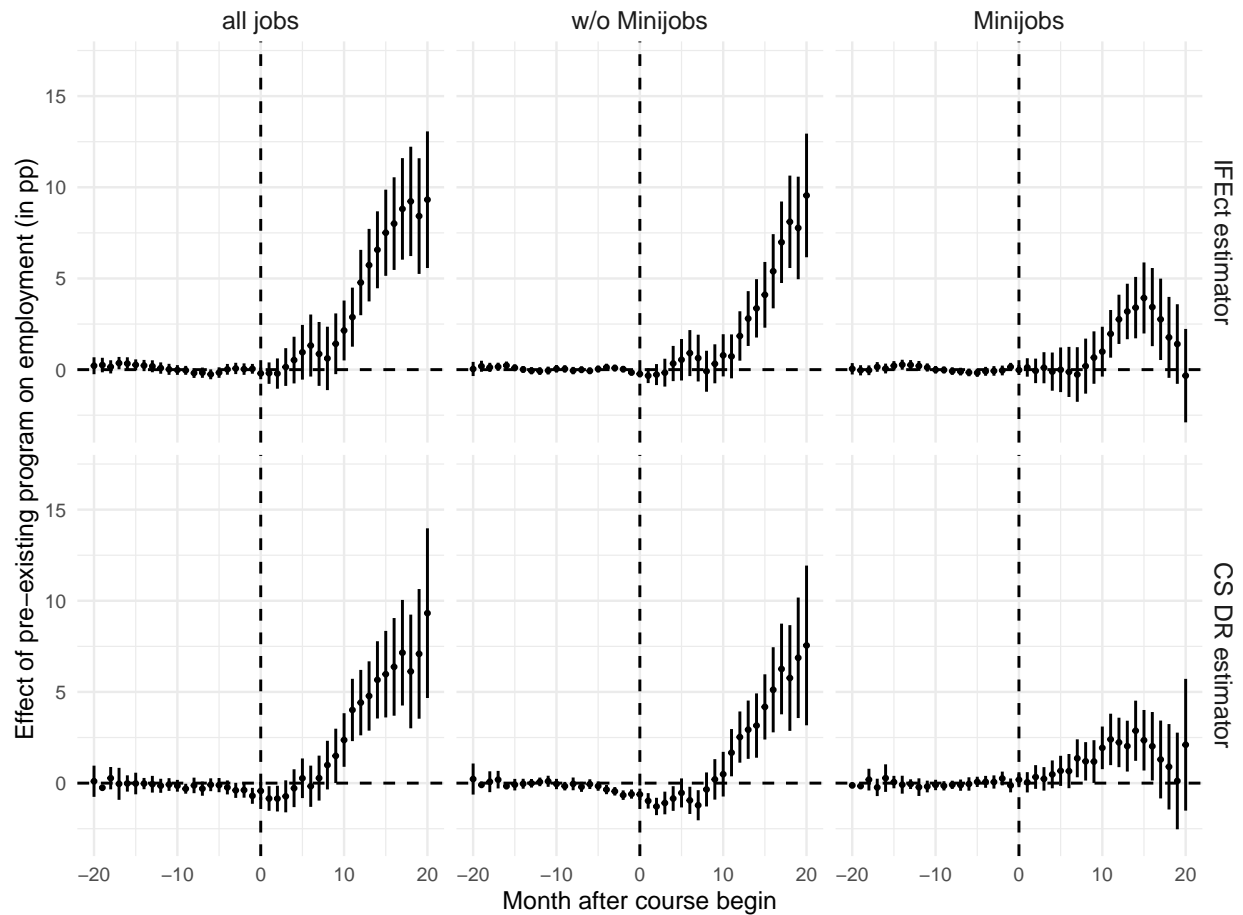


Figure 4: Top panel: Estimates of the effects of the preexisting language training program from difference-in-differences imputation estimator with interactive fixed effects for three outcomes: all jobs (left), all jobs but minijobs (middle), minijobs (right). Bottom panel: Estimates from doubly robust CS DR difference-in-differences estimator. Overall the estimates suggest that the preexisting language program increased employment following the six months course duration.

7 Discussion

The results indicate that while the large-scale ad hoc program had no discernible effects on employment, the smaller-scale preexisting program led to considerable increases in employment. This latter finding refutes the hypothesis that language courses are generally ineffective in facilitating economic integration during times of large-scale displacement crises. However, the contrasting results raise a new question: Which program features account for their differential effectiveness? Without randomizing the different features of the programs, it is difficult to disentangle their individual contributions to the overall effect. Nonetheless, the two programs can be distinguished along several dimensions that past research has shown to be important for effectiveness, including the duration of the course (Rolstad et al., 2005), the provision of skill certificates (Desiderio, 2016; Pecoraro and Wanner, 2019; Tani, 2017; Brücker et al., 2021), and the classroom composition (Sprietsma and Pfeil, 2015).

While the maximum class size was fairly similar between the two programs, the preexisting program was much longer, with 600 hours compared to only 320 hours in the ad hoc program. In addition, the preexisting program had a consistent curriculum, set standards for course providers, and provided successful course participants with a certificate they can show to potential employers. The ad hoc program lacked these features (see SM Table S.1 for details).

Given these differences, it appears likely that the ad hoc program lacked the quality and quantity of instruction necessary for participants to acquire sufficient German proficiency that would translate into more success in the labor market. An alternative interpretation is that the program did lead to improved language skills, but the lack of certification meant that participants could not credibly signal those skills to potential employers. We hope that future research will be able to shed light on the relative impact of these and other program features.

It is also important to recognize that selection of refugees into the two programs may have contributed to the heterogeneous effects we find. When comparing the characteristics of the participants in the preexisting program with the rest of the sample, we observe that the participants in the preexisting program had a slightly higher share of Syrians and refugees with some education (SM Table S.5). To the extent that the courses had a greater effect on refugees who are positively selected, this could partly explain why the preexisting program had more positive effects.

8 Conclusion

Our findings carry important implications for the design of language policies and programs during periods of mass refugee displacement. Over the past decades, millions of displaced individuals have arrived in Europe and been granted refugee status and subsidiary protection. Mastering the host country’s language is typically the first, and potentially most crucial, step toward successful integration into the host country’s economy and society. While virtually all European host countries offer language programs, it remains unclear how best to expand their capacity during crises and periods of high demand.

Our results suggest that the ambitious initiative by policymakers in Germany to establish a large-scale, ad hoc language program meant that many refugees received basic language training shortly after arrival. However, ultimately, this program proved ineffective in improving their employment prospects. To investigate whether these null effects stem from particular features of the ad hoc program or arise from the general challenge of delivering effective language programs during refugee crises, we contrast our analysis with a preexisting, more comprehensive language course. We find that this program nearly doubled refugees’ chances of employment. This indicates that more comprehensive programs can successfully facilitate labor market integration even during periods of large refugee arrivals and associated pressures on the asylum system and host communities.

Our findings regarding the divergent effects of the two programs highlight a significant trade-off in designing an effective response to refugee crises. While the preexisting program yielded high returns, it served far fewer refugees, whereas the ad hoc program reached more refugees but failed to generate discernible employment benefits. Ultimately, the null finding from the ad hoc program suggests that it prioritized quantity over quality in an effort to swiftly assist the maximum number of individuals. It appears that a more promising approach would involve investing in scaling up comprehensive programs that provide at least some employment benefits to participants, even if it means not everyone can be served immediately. Determining the optimal level of comprehensiveness for such investments remains an open and crucial question for future research.

Last but not least, it is important to recognize two limitations of the study. First, since there was no variation in the participation in the different components of the two programs, we could only evaluate the effects of the programs as they were rolled out in reality. While this evidence is of first-order policy importance, it does not quantify the relative returns to the different components of each program (e.g., course duration, certification, quality of instruction). Future research on this question would be important to help design optimal programs.

Second, given our data limitations, we could only examine effects on employment outcomes. It is well understood that immigrant integration is a multi-dimensional concept that goes beyond economic success. To better inform policy debates, future research should examine the impacts on other important dimensions of refugee integration, such as linguistic, psychological, political, or social integration (Harder et al., 2018).

References

- Asfar, D., M. P. Born, J. K. Oostrom, and M. van Vugt (2019). Psychological Individual Differences as Predictors of Refugees’ Local Language Proficiency. *European Journal of Social Psychology* 49(7), 1385–1400.
- BA Presseteam (02.10.2015). Sprachförderung als Basis für Integration in Arbeit.
- Bach, S., H. Brücker, P. Haan, A. Romiti, K. van Deuverden, and E. Weber (2017). Refugee Integration: A Worthwhile Investment. *DIW Economic Bulletin* 7(3/4), 33–43.
- Bai, J. (2009). Panel Data Models With Interactive Fixed Effects. *Econometrica* 77(4), 1229–1279.
- BAMF (2015). Konzept für einen bundesweiten Integrationskurs: Überarbeitete Neuauflage - April 2015.
- BAMF (2018). *Das Bundesamt in Zahlen 2017: Asyl*. Nürnberg: Bundesamt für Migration und Flüchtlinge.
- Bansak, K., J. Ferwerda, J. Hainmueller, A. Dillon, D. Hangartner, D. Lawrence, and J. Weinstein (2018). Improving Refugee Integration through Data-driven Algorithmic Assignment. *Science* 359(6373), 325–329.
- Becker, S. O. and A. Ferrara (2019). Consequences of Forced Migration: A Survey of Recent Findings. *Labour Economics* 59, 1–16.
- Beiser, M. and F. Hou (2000). Gender Differences in Language Acquisition and Employment Consequences among Southeast Asian Refugees in Canada. *Canadian Public Policy / Analyse de Politiques* 26(3), 311.
- Berman, E., K. Lang, and E. Siniver (2003). Language-Skill Complementarity: Returns to Immigrant Language Acquisition. *Labour Economics* 10(3), 265–290.
- Bjertrup, P. J., M. Bouhenia, P. Mayaud, C. Perrin, J. B. Farhat, and K. Blanchet (2018). A Life in Waiting: Refugees’ Mental Health and Narratives of Social Suffering After European Union Border Closures in March 2016. *Social Science & Medicine* 215, 53–60.
- Bleakley, H. and A. Chin (2004). Language Skills and Earnings: Evidence from Childhood Immigrants. *Review of Economics and Statistics* 86(2), 481–496.
- Borusyak, K., X. Jaravel, and J. Spiess (2024). Revisiting event study designs: Robust and efficient estimation. *Review of Economic Studies*, rdae007.

- Brücker, H., A. Glitz, A. Lerche, and A. Romiti (2021). Occupational Recognition and Immigrant Labor Market Outcomes. *Journal of Labor Economics* 39(2), 497–525.
- Bundesrechnungshof (2017). Abschließende Mitteilung an den Vorstand der Bundesagentur für Arbeit über die Prüfung von Sprachkursen nach § 421 SGB III (Einstiegsurse).
- Callaway, B. and P. H. Sant’Anna (2021). Difference-In-Differences With Multiple Time Periods. *Journal of Econometrics* 225(2), 200–230.
- Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica* 82(6), 2295–2326.
- Cheng, Z., B. Z. Wang, and L. Taksa (2021). Labour Force Participation and Employment of Humanitarian Migrants: Evidence From the Building a New Life in Australia Longitudinal Data. *Journal of Business Ethics* 168, 697–720.
- Chiswick, B. R. (1978). The Effect of Americanization on the Earnings of Foreign-born Men. *Journal of Political Economy* 86(5), 897–921.
- Chiswick, B. R. and P. W. Miller (1995). The Endogeneity Between Language and Earnings: International Analyses. *Journal of Labor Economics* 13(2), 246–288.
- Chiswick, B. R. and P. W. Miller (2001). A Model of Destination-Language Acquisition: Application to Male Immigrants in Canada. *Demography* 38(3), 391.
- Cohen, Y. and Y. Haberfeld (2007). Self-Selection and Earnings Assimilation: Immigrants From the Former Soviet Union in Israel and the United States. *Demography* 44(3), 649–668.
- Cortes, K. E. (2004). Are Refugees Different From Economic Immigrants? Some Empirical Evidence on the Heterogeneity of Immigrant Groups in the United States. *Review of Economics and Statistics* 86(2), 465–480.
- Council of the European Union (2014, June). Council Conclusions of the Council and the Representatives of the Governments of the Member States on the Integration of Third-Country Nationals Legally Residing in the EU. <http://www.consilium.europa.eu/en/workarea/downloadAsset.aspx?id=15904>.
- De Chaisemartin, C. and X. d’Haultfoeuille (2020). Two-way Fixed Effects Estimators With Heterogeneous Treatment Effects. *American Economic Review* 110(9), 2964–2996.

- Degler, E. and T. Liebig (2017). Finding Their Way. Labour Market Integration of Refugees in Germany. *OECD*. <http://www.oecd.org/migration..>
- Delander, L., M. Hammarstedt, J. Månsson, and E. Nyberg (2005). Integration of Immigrants: The Role of Language Proficiency and Experience. *Evaluation review* 29(1), 24–41.
- Desiderio, M. V. (2016). *Integrating Refugees Into Host Country Labor Markets: Challenges and Policy Options*. Migration Policy Institute.
- Deutscher Bundestag (2016). Integrationskurse. Antwort der Bundesregierung auf die Kleine Anfrage. Drucksache 18/10438, Deutscher Bundestag 18. Wahlperiode.
- Dribe, M. and C. Lundh (2011). Cultural Dissimilarity and Intermarriage. A Longitudinal Study of Immigrants in Sweden 1990-2005. *International Migration Review* 45(2), 297–324.
- Dumont, J.-C., T. Liebig, J. Peschner, F. Tanay, and T. Xenogiani (2016). How Are Refugees Faring on the Labour Market in Europe? A First Evaluation Based on the 2014 EU Labour Force Survey Ad Hoc Module. *European Commission DG Employment Working Paper* (1).
- Dustmann, C. and F. Fabbri (2003). Language Proficiency and Labour Market Performance of Immigrants in the UK. *The Economic Journal* 113(489), 695–717.
- Fasani, F., T. Frattini, and L. Minale (2018). (The Struggle for) Refugee Integration into the Labour Market: Evidence from Europe. Discussion Paper DP12718, CEPR.
- Ferwerda, J., H. Finseraas, and J. Bergh (2020). Voting Rights and Immigrant Incorporation: Evidence From Norway. *British Journal of Political Science* 50(2), 713–730.
- Foged, M., L. Hasager, G. Peri, J. N. Arendt, and I. Bolvig (2022). Language Training and Refugees’ Integration. *Review of Economics and Statistics* 106(4), 1–41.
- Foged, M. and C. Van der Werf (2023). Access to Language Training and the Local Integration of Refugees. *Labour Economics* 84, 102366.
- Goodman-Bacon, A. (2021). Difference-In-Differences With Variation in Treatment Timing. *Journal of Econometrics* 225(2), 254–277.
- Hainmueller, J., D. Hangartner, and D. Lawrence (2016). When Lives Are Put on Hold: Lengthy Asylum Processes Decrease Employment Among Refugees. *Science Advances* 2(8), e1600432.

- Hangartner, D., E. Dinas, M. Marbach, K. Matakos, and D. Xefteris (2019). Does Exposure to the Refugee Crisis Make Natives More Hostile? *American Political Science Review* 113(2), 442–455.
- Harder, N., L. Figueroa, R. M. Gillum, D. Hangartner, D. D. Laitin, and J. Hainmueller (2018). Multidimensional measure of immigrant integration. *Proceedings of the National Academy of Sciences* 115(45), 11483–11488.
- Hayfron, J. E. (2001). Language Training, Language Proficiency and Earnings of Immigrants in Norway. *Applied Economics* 33(15), 1971–1979.
- Konle-Seidl, R. (2018). *Integration of Refugees in Austria, Germany and Sweden: Comparative Analysis*. European Union.
- Liebau, E. and D. Schacht (2016). Language Acquisition: Refugees Nearly Achieve Proficiency Level of Other Migrants. *DIW Economic Bulletin* 6(34/35), 400–406.
- Liebig, T. (2007). *The Labour Market Integration of Immigrants in Denmark*, Volume 50 of *OECD Social, Employment and Migration Working Papers*. Paris: OECD Publishing.
- Liu, L., Y. Wang, and Y. Xu (2024). A Practical Guide to Counterfactual Estimators for Causal Inference With Time-Series Cross-Sectional Data. *American Journal of Political Science* 68(1), 160–176.
- Lochmann, A., H. Rapoport, and B. Speciale (2019). The Effect of Language Training on Immigrants’ Economic Integration: Empirical Evidence From France. *European Economic Review* 113, 265–296.
- Lubotsky, D. (2007). Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings. *Journal of Political Economy* 115(5), 820–867.
- Marbach, M., J. Hainmueller, and D. Hangartner (2018). The Long-Term Impact of Employment Bans on the Economic Integration of Refugees. *Science Advances* 4, eaap9519.
- OECD (2016). *Making Integration Work: Humanitarian Migrants*. OECD Publishing.
- Pecoraro, M., A. Manatschal, E. G. Green, and P. Wanner (2022). How Effective Are Integration Policy Reforms? The Case of Asylum-Related Migrants. *International Migration* 60(6), 95–110.
- Pecoraro, M. and P. Wanner (2019). Does the Recognition of Foreign Credentials Decrease the Risk for Immigrants of Being Mismatched in Education or Skills? In I. Steiner and

- P. Wanner (Eds.), *Migrants and Expats: The Swiss Migration and Mobility Nexus*, pp. 161–186. Cham: Springer.
- Robert Bosch Stiftung (2015). Themendossier Sprachvermittlung und Spracherwerb für Flüchtlinge: Praxis und außerschulische Angebote.
- Rolstad, K., K. Mahoney, and G. V. Glass (2005). The Big Picture: A Meta-Analysis of Program Effectiveness Research on English Language Learners. *Educational Policy* 19(4), 572–594.
- Sant’Anna, P. H. and J. Zhao (2020). Doubly Robust Difference-In-Differences Estimators. *Journal of Econometrics* 219(1), 101–122.
- Schlaudt, V. A., R. Bosson, M. T. Williams, B. German, L. M. Hooper, V. Frazier, R. Carrico, and J. Ramirez (2020). Traumatic Experiences and Mental Health Risk for Refugees. *International Journal of Environmental Research and Public Health* 17(6), 1943.
- Scholten, P., F. Baggerman, L. Dellouche, V. Kampen, J. Wolf, and R. Ympa (2017). *Policy Innovation in Refugee Integration: A Comparative Analysis of Innovative Policy Strategies Toward Refugee Integration in Europe*. Rotterdam: Erasmus Migration & Diversity Institute.
- Schönwälder, K., J. Söhn, and I. Michalowski (2005). *Sprach- und Integrationskurse für MigrantInnen: Erkenntnisse über ihre Wirkungen aus den Niederlanden, Schweden und Deutschland*, Volume 3.
- Sprietsma, M. and L. Pfeil (2015). Peer Effects in Language Training for Migrants. Discussion Paper 15-033, ZEW - Centre for European Economic Research.
- Tani, M. (2017). Local Signals and the Returns to Foreign Education. *Economics of Education Review* 61, 174–190.
- The Economist (2015). Getting the New Arrivals to Work. *The Economist*.
- UNHCR (2013). *A New Beginning: Refugee Integration in Europe: Outcome of an EU funded project on Refugee Integration Capacity and Evaluation (RICE)*. United Nations High Commissioner for Refugees, Bureau for Europe.
- Van Tubergen, F. (2010). Determinants of Second Language Proficiency Among Refugees in the Netherlands. *Social Forces* 89(2), 515–534.

van Tubergen, F., I. Maas, and H. Flap (2004). The Economic Incorporation of Immigrants in 18 Western Societies: Origin, Destination, and Community Effects. *American Sociological Review* 69(5), 704–727.

Acknowledgments

Acknowledgments The authors thank Joelle Pianzola and Jonathan Homola for helpful comments on earlier versions of this draft. We also thank the participant at the IAB-Colloquium in Nuremberg. Hangartner acknowledges funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (Grant No. 804307), the Swiss Network for International Studies, and the Leverhulme Trust. Hainmueller and Hangartner acknowledge funding from the U.S. NSF (grant no. 1627339). The funders had no role in the design, data collection, analysis, decision to publish, or preparation of the manuscript. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders.

Author contributions M.M., E.V., N.H., J.H., and D.H. designed and performed research; M.M. and E.V. analyzed data; J.H., M.M., and D.H. wrote the paper with input from E.V. and N.H.

Data availability Our analysis builds on two datasets: the Integrated Employment Biographies (IEB V14.00.00-190927) and Status History Migration (SHZ V03.03.00-201904). The data is provided by the Institute for Employment Research (IAB), Regensburger Str. 104, 90478 Nuremberg, Germany, homepage: www.iab.de, email: iab@iab.de. The data source is social security data with administrative origin, which are processed and kept by IAB according to Social Code Book III. Due to sensitive information, the access to this data is subject to the confidentiality regulations according to §35 of the Germany Social Code Book I and requires an application granted by the Germany Federal Ministry of Labor and Social Affairs. Therefore, for replication purposes these data can only be accessed on-site at the IAB upon application and approval. For further information regarding the data and replication purposes please contact the authors.

Code availability Replication code will be made available at the Harvard Dataverse: <https://doi.org/10.7910/DVN/ICCFHJ>.

Supplementary Materials for

Does Ad Hoc Language Training Improve the Economic Integration of Refugees? Evidence from Germany's Response to the Syrian Refugee Crisis

Contents

| | |
|---|----|
| Background on the Data | 2 |
| Background on Language Training Programs | 3 |
| Detailed Results: Ad Hoc Language Training Program | 5 |
| Detailed Results: Preexisting Language Training Program | 21 |

Background on the Data

Our dataset is an individual-level monthly panel data of refugees and asylum-seekers. This dataset was constructed from the Integrated Employment Biographies (IEB) data of the Institute for Employment Research (IAB). The IEB draws on administrative sources of the German Federal Employment Agency (BA) and includes all individuals in Germany who either have an employment status subject to social security, a marginal part-time employment status, or are benefit recipients, officially registered as unemployed, or participants in active labor market programs. The data excludes employees of the state and self-employed individuals but since refugees are largely not allowed to become self-employed or unlikely to be employed by the state, this is of little consequence for our study. Based on an additional dataset, the *Statushistorik Zuwanderung-SHZ*, we also have information on the immigration status of all individuals in our data. Since states are reimbursed for some of the costs that they incur for hosting asylum-seekers, the Federal Employment Agency strives for a high data quality also for the immigration-related data that we use [3].

Our analysis dataset consists of asylum seekers from Syria, Iran, Iraq, and Eritrea who arrived (and registered) between June 2015 and June 2016 and were 18-35 years old at the time of arrival. Since the date of arrival is only recorded for individuals that live outside of *Optionskommunen*, we exclude everyone living in any of these 104 counties and focus on individuals living in the remaining 297 counties.¹ We also remove all refugees arriving on the first day of a month as the records of refugees whose exact arrival date was unknown were typically assigned the first of the month by default. Our main outcome measure of labor market integration is employment status (including marginal employment, apprenticeships and paid internships) in a given month. Marginal employment includes all part-time, low-wage jobs (*Geringfügig entlohnte Beschäftigte*, *Kurzfristig Beschäftigte* and *Niedriglohn in Gleitzone*, 400-800 Euro).

¹*Optionskommunen* are counties that decided to take full responsibility and supervision of certain benefit recipients after the labor-market reforms of 2004.

Background on Language Training Programs

At the core of the analysis are two public language training programs for refugees. The first language program refers to the ad hoc language course provided by the BA. After the federal government passed the legal amendment of the Article 421 of Social Code III (§421 *Sozialgesetzbuch III*) to allow the BA to provide quick and unbureaucratic language support for refugees, the BA introduced a temporary, basic language program for refugees (called *Einstiegskurs zur Deutschförderung*) [2].

Access to the ad hoc language program was limited to asylum-seekers with valid proof of registration at course start (e.g., *BüMA*, *Ankunftsnachweis* or *Aufenthaltsgestattung*)² and limited to persons from Syria, Eritrea, Iraq and Iraq (because they had a high probability of getting protection status). By law, course providers had to start the course before December 31, 2015 to receive reimbursement.

The second language training program refers to the preexisting and more comprehensive language course (*Integrationskurs*) offered by the Federal Office for Migration and Refugees (BAMF). The comprehensive language program was introduced with the Immigration Act (*Zuwanderungsgesetz*) in 2005 and includes language training (600 teaching hours) and training of German cultural and values (100 teaching hours). Refugees from Eritrea, Iran, Iraq, Syria or Somalia are allowed to participate in these courses while their asylum application is reviewed. After receiving some form of protection, refugees are typically required to participate in these courses if they have not already done so.

Table S.1 details the content and the structure of both the ad hoc program and the preexisting language program. There are several differences between the two courses. The ad hoc program offered 320 hours of instruction compared to 600 hours in the preexisting program. The preexisting program also had a standardized curriculum developed by the Goethe-Institute and a final examination according to the Common European Framework of Reference for Languages (CEFR). Successful participants obtained a language certificate for B1 language competency in writing, reading, speaking and conversation according to CEFR. In contrast, the ad hoc courses had no standardized curriculum and offered no certification.

²Typically, refugees and asylum-seekers receive a proof of arrival (*Ankunftsnachweis*) or “Asylumseeker Registration Certificate” (*Bescheinigung über die Meldung als Asylsuchender - BüMA*) at reception facilities and arrival centers.

| Program | Ad Hoc | Pre-existing |
|--------------------|--|--|
| Responsible agency | Federal Employment Agency | Federal Office for Migration and Refugees |
| Teaching hours | 320 | 600 |
| Duration | 8 weeks | Full-time (part-time allowed) |
| Class size | 25 participants | max 20 participants |
| Content | Basic German language skills; no standardized framework curriculum | Standardized framework curriculum according to Goethe-Institute; various modules; training in writing, reading, speaking, conversation |
| Certification | no certificate | Certificates according to Common European Framework of Reference for Languages (CEFR) |
| Program provider | Decentralized; no specific requirements | Decentralized; point-based requirement system |

Table S.1: Content and Structure of Ad Hoc and Pre-existing Language Program

Detailed Results: Ad Hoc Language Training Program

- Balance Tests for the RDD: Table S.2 presents a series of placebo estimates using our main specification (local linear regression with an MSE-optimal bandwidth). There are no discernible differences in the covariates of individuals who have been eligible and those who have not been eligible for the ad hoc language program at the cutoff. The only exception is that those eligible for the ad hoc program tend to be about 4 percentage points less likely to have no schooling. This balance of the covariates at the cutoff supports the identification assumption of the RDD.
- Density of Running Variable: Figure S.1 displays a histogram of the running variable. Positive values correspond to individuals arriving before December 31 (the cut-off date), while negative values correspond to individuals arriving after. Since we drop persons arriving on the first of every month, there are some days for which there are zero observations. Using the statistical test suggested in [6], we find no strong evidence for a discontinuity of the density at the cutoff (conventional t-statistic and p-value: $T = -1.3$, $p(> |T|) = 0.19$; bias-corrected: $T = -2.1$, $p(> |T|) = 0.03$).
- Main Effects for the RDD: Table S.3 reports the main estimates as displayed in Figure 3 in the main text. Sample sizes differ because for refugees arriving late, not all 22 months of post-arrival employment outcomes are observed in our data.
- Participation: Participation in the language courses is only observed in our data for the preexisting language program, but not for the ad hoc language program. However, for the latter program, we received additional data based on paper records from the BA. This data includes the number of participants per course, information about when each course started, and which of the 40 local branches of the BA managed the paperwork. Yet, this data does not provide information on the arrival and registration dates of the participating refugees. Therefore, we collected a sample of 500 course registration records from participants in April 2018. This sample was randomly drawn from scanned course registration records available at the BA (In 61 of the sampled records, the paperwork was incomplete or unreadable, and the participant's registered arrival date could not be determined). To estimate the participation rate, we first use raking to post-stratify the random sample of participants by managing office and the week of course start. Next, we combine the estimated monthly participant counts with population data on the number of refugee arrivals from Syria, Iraq, Iran, and Eritrea in the EASY-System (*Erstverteilung der Asylbegehrenden-System*). Figure S.2 shows the estimated participation rates in the ad hoc program by refugees' date of registration in Germany. We find that the estimated program participation rate drops sharply by about 21 percentage points at the cutoff. Figure S.3 shows a binned scatter plot with the average waiting period (number of days between the beginning of the course and the registration date) by calendar week when the course starts. To estimate the average waiting period for each week, we use the post-stratified data. The fitted linear regression line suggests that the average waiting period does not systematically vary with the week when the course starts.
- Main Effects for the RDD by Employment Type: Figure S.4 shows the estimates of the ad hoc program broken down by employment type: all jobs, marginal employment, and jobs excluding marginal employment. We see that the estimated effects are statistically insignificant across all employment types and months of residency. This holds regardless of

whether we use the conventional or bias-corrected regression discontinuity estimator (with a common MSE-optimal bandwidth) [5, 4]. This indicates that the null-finding is not an artifact of choosing a specific estimator.

- Main Effects for the RDD Adding Covariates: Figure S.5 shows that there are also few differences between estimates adjusted for observed demographic covariates and unadjusted estimates. The covariates include: Age (18-20, 21-25, 26-30, 31-35), nationality (Syrian, Iraqi, Iranian, Eritrean), schooling (no, yes, missing), sex (female, male). This lack of sensitivity to the inclusion of covariates supports the identification assumption of the RDD.
- Main Effects for the RDD Imputing for Attrition: In Figure S.6, we show the same estimates using the observed outcomes as well as imputed outcomes. The latter outcomes are constructed by imputing a zero for all years post-arrival without an observed employment outcome. The motivation for this imputation strategy is that any form of formal employment (other than self-employment) should be observed in the data. Unobserved outcome values could be, for example, due to emigration out of Germany or death. This indicates that the null-finding is not an artifact of panel attrition.
- Main Effects for the RDD Varying the Bandwidth: Figure S.7 shows the main estimates for the employment effect of the ad hoc program on refugee employment by varying the bandwidth for the RDD and including polynomials of the running variable. We find that the null effect is similar across all specifications. If anything, we observe a negative effect of ad hoc program eligibility on employment for some of the specifications. This indicates that the null-finding is not an artifact of choosing a specific specification or bandwidth.
- Main Effects for the RDD with Donut Exclusion: Figure S.8 shows the estimates of the ad hoc program when we apply a (one-sided) donut RDD [1, 7] and exclude observations to the left of the threshold (i.e., before December 31). For example, a donut size of 30 days in the figure means that we estimate the effect using a regression discontinuity design that compares refugees who arrived on December 1, 2015 with refugees who arrived on January 1, 2016, excluding all observations between December 2 and December 31. We see that the estimated effect is close to zero and far from statistical significance for all donut sizes. This indicates that the null-finding is not an artifact of choosing a specific donut period.
- Main Effects for the RDD Excluding Refugees who Also Participated in the Preexisting Program: Figure S.9 displays the effects of the ad hoc program when excluding individuals who, at some point during our study period, enrolled in the preexisting program. Once again, we find no discernible differences. This suggests that our null finding is not influenced by refugees who were ineligible for the ad hoc program but later participated in the pre-existing program. It is important to note that this test conditions on a post-treatment variable and should therefore be interpreted with caution. As an additional verification, we replicated the models to estimate the effect of eligibility for the ad hoc program on the probability of enrolling in the preexisting program. As shown in Figure S.10, we observe a small decline in the participation probability for the preexisting program at the cutoff for the ad hoc program but the estimates are statistically different from zero in only 6 out of 17 months.
- Main Effects for the RDD by Gender: Figure S.11 shows the effects of the ad hoc program separately for men and women. We observe that the estimates for women are concentrated

around zero, while there is more variability for men. As for the pooled estimates, we tend, if anything, to observe negative employment effects for men for selected months after arrival.

- Main Effects for the RDD by Refugees' State of Residence at Arrival: Figure S.12 shows the effects of the ad hoc program broken down by refugees' state of residence at arrival. For this estimation, we pool the sample across residency months and adjust for residency month fixed effects. We find that there is some heterogeneity, but the distribution of effects is centered around zero. This indicates that the null effects are not driven by particular geographic areas or the results of large treatment effect heterogeneity that averages out to zero.
- Main Effects for the RDD by Refugees' Education: Figure S.13 shows the effects of the ad hoc program broken down by refugees' education levels and for the pooled sample. Again, we find that there is some heterogeneity, but the distribution of effects is centered around zero. However, it is worth noting that information on education is only implicitly recorded as part of the CV, particularly for job seekers and participants in active labor market programs from the administrative system career path. The data quality of the education field can therefore be classified as poor. Therefore, this test should be interpreted cautiously.

| Covariate | Coef. | S.E. | p(> z) | Obs. |
|-------------|-------|------|---------|--------|
| Age | | | | |
| 18-20 | 0.02 | 0.01 | 0.14 | 209284 |
| 21-25 | -0.02 | 0.01 | 0.10 | 209284 |
| 26-30 | 0.01 | 0.01 | 0.33 | 209284 |
| 31-35 | -0.01 | 0.01 | 0.54 | 209284 |
| Gender | | | | |
| Female | -0.01 | 0.02 | 0.56 | 209284 |
| Nationality | | | | |
| Eritrea | 0.00 | 0.00 | 0.54 | 209284 |
| Iran | 0.01 | 0.01 | 0.27 | 209284 |
| Iraq | -0.01 | 0.02 | 0.48 | 209284 |
| Syrian | -0.01 | 0.02 | 0.51 | 209284 |
| Schooling | | | | |
| No | -0.04 | 0.01 | 0.01 | 209284 |
| Yes | 0.02 | 0.02 | 0.39 | 209284 |
| (Missing) | 0.02 | 0.01 | 0.13 | 209284 |

Table S.2: Placebo estimates and 95% confidence intervals from local linear regressions with (MSE)-optimal bandwidth estimating the intention-to-treat effect of the ad-hoc program eligibility at the cut-off date.

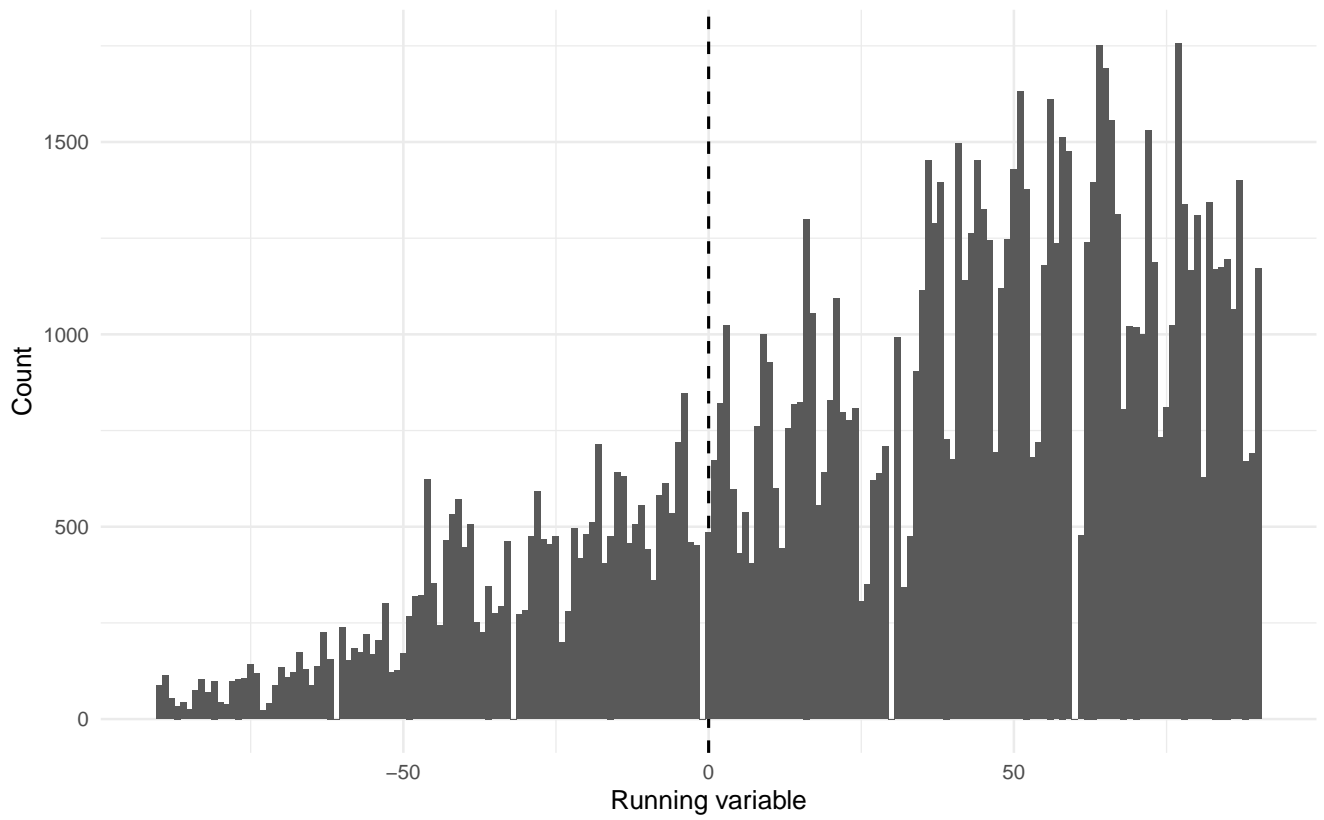


Figure S.1: Frequency histogram of the running variable 90 days around the cut-off date. Each bin corresponds to 1 day.

| Month | Coef. | S.E. | $p(> z)$ | 95% CI | | Obs. |
|-------|-------|------|-----------|--------|-------|--------|
| | | | | Low | High | |
| 6 | -0.02 | 0.16 | 0.92 | -0.33 | 0.29 | 209284 |
| 7 | 0.03 | 0.19 | 0.88 | -0.34 | 0.40 | 208623 |
| 8 | -0.03 | 0.26 | 0.92 | -0.53 | 0.48 | 207829 |
| 9 | 0.13 | 0.31 | 0.68 | -0.49 | 0.74 | 206971 |
| 10 | -0.15 | 0.35 | 0.67 | -0.84 | 0.54 | 206137 |
| 11 | -0.08 | 0.39 | 0.85 | -0.84 | 0.69 | 205331 |
| 12 | -0.00 | 0.38 | 1.00 | -0.75 | 0.75 | 204549 |
| 13 | -0.37 | 0.41 | 0.38 | -1.17 | 0.44 | 203787 |
| 14 | -0.58 | 0.45 | 0.19 | -1.46 | 0.30 | 202973 |
| 15 | -0.74 | 0.49 | 0.13 | -1.70 | 0.22 | 202231 |
| 16 | -0.37 | 0.59 | 0.54 | -1.53 | 0.80 | 201505 |
| 17 | -0.92 | 0.64 | 0.15 | -2.17 | 0.33 | 200806 |
| 18 | -0.87 | 0.71 | 0.23 | -2.27 | 0.53 | 200195 |
| 19 | -1.86 | 0.71 | 0.01 | -3.26 | -0.47 | 199580 |
| 20 | -2.04 | 0.92 | 0.03 | -3.84 | -0.24 | 197342 |
| 21 | -0.96 | 1.08 | 0.37 | -3.08 | 1.16 | 194824 |
| 22 | -2.01 | 1.09 | 0.06 | -4.14 | 0.12 | 191708 |

Table S.3: Main estimates and 95% confidence intervals from local linear regressions with (MSE)-optimal bandwidth estimating the intention-to-treat effect of program eligibility at the cut-off date 6 to 22 month after arrival.

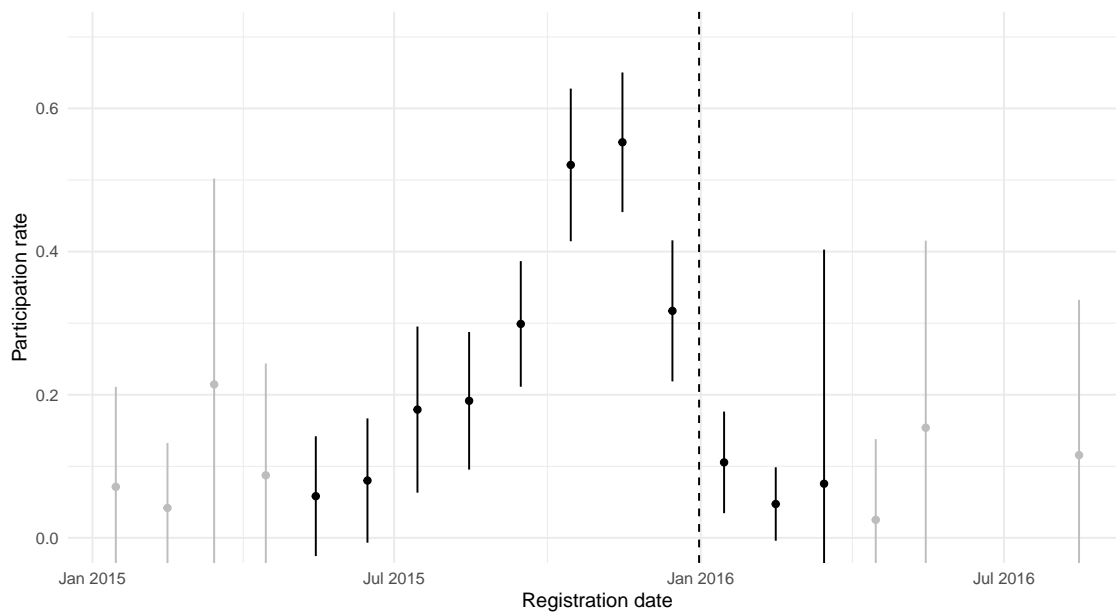


Figure S.2: Estimated participation rate in the ad-hoc program by registration date. Some confidence intervals are clipped to increase readability.

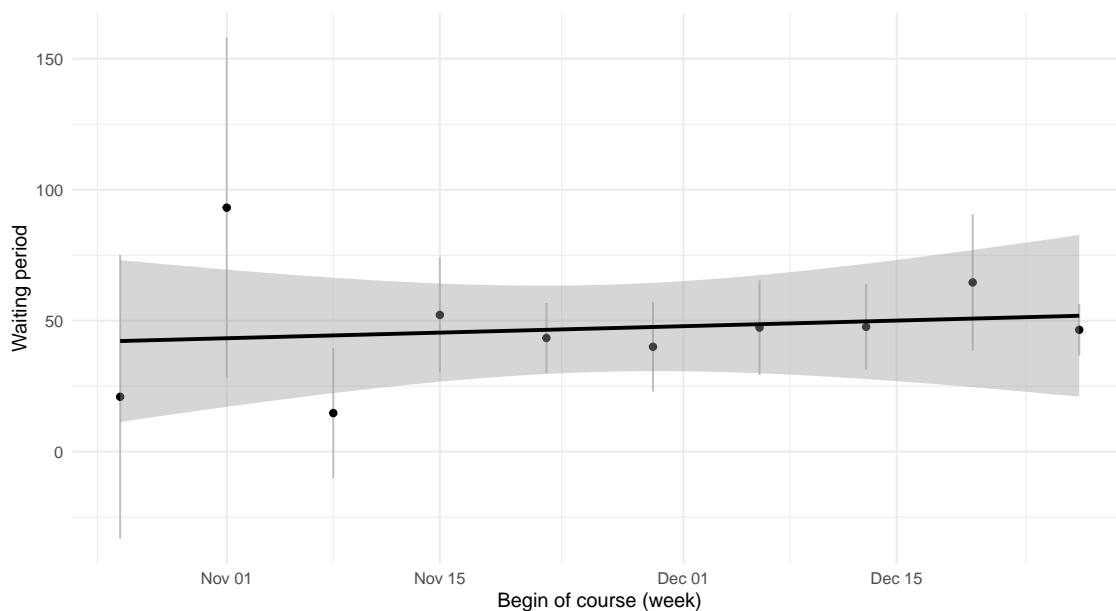


Figure S.3: Binned scatter plot showing the estimated difference between beginning of ad hoc course and registration date with fitted linear regression line.

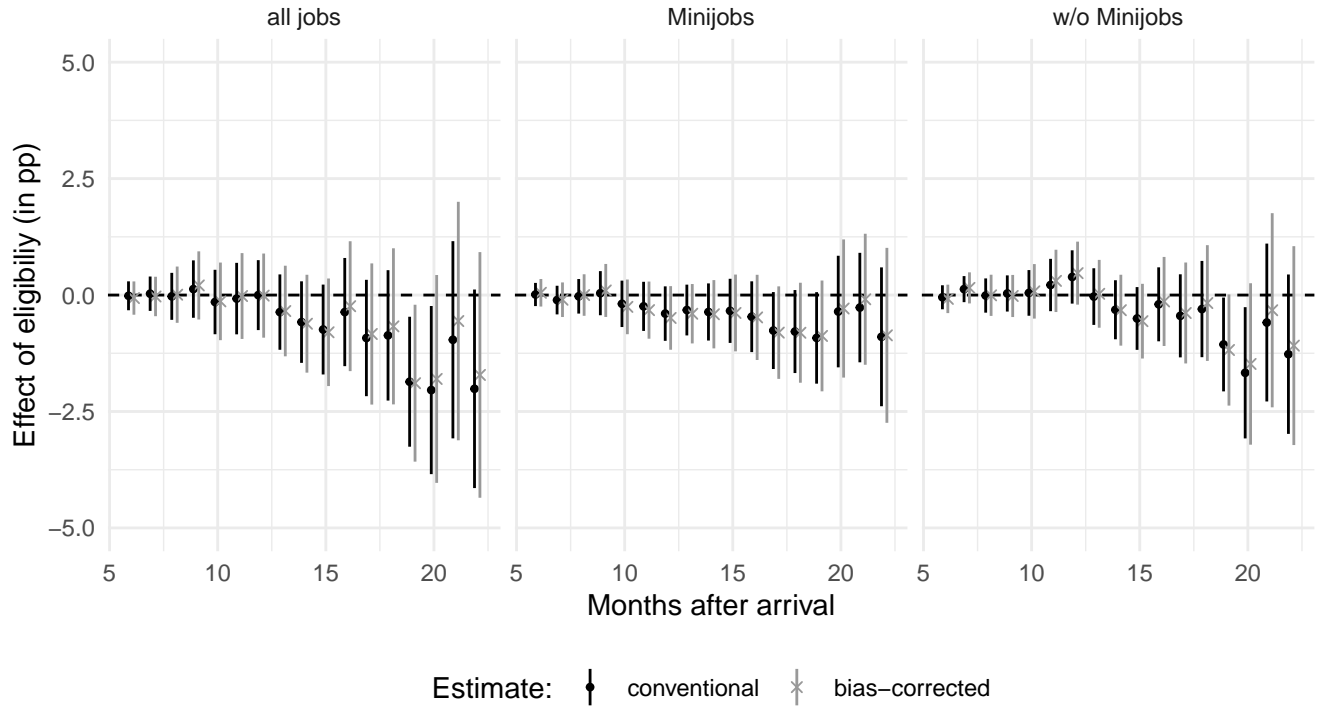


Figure S.4: Estimates for three employment outcomes and corresponding 95% confidence intervals from conventional and bias-corrected local linear regressions with (MSE)-optimal bandwidth estimating the intention-to-treat effect of the ad-hoc program eligibility at the cut-off date.

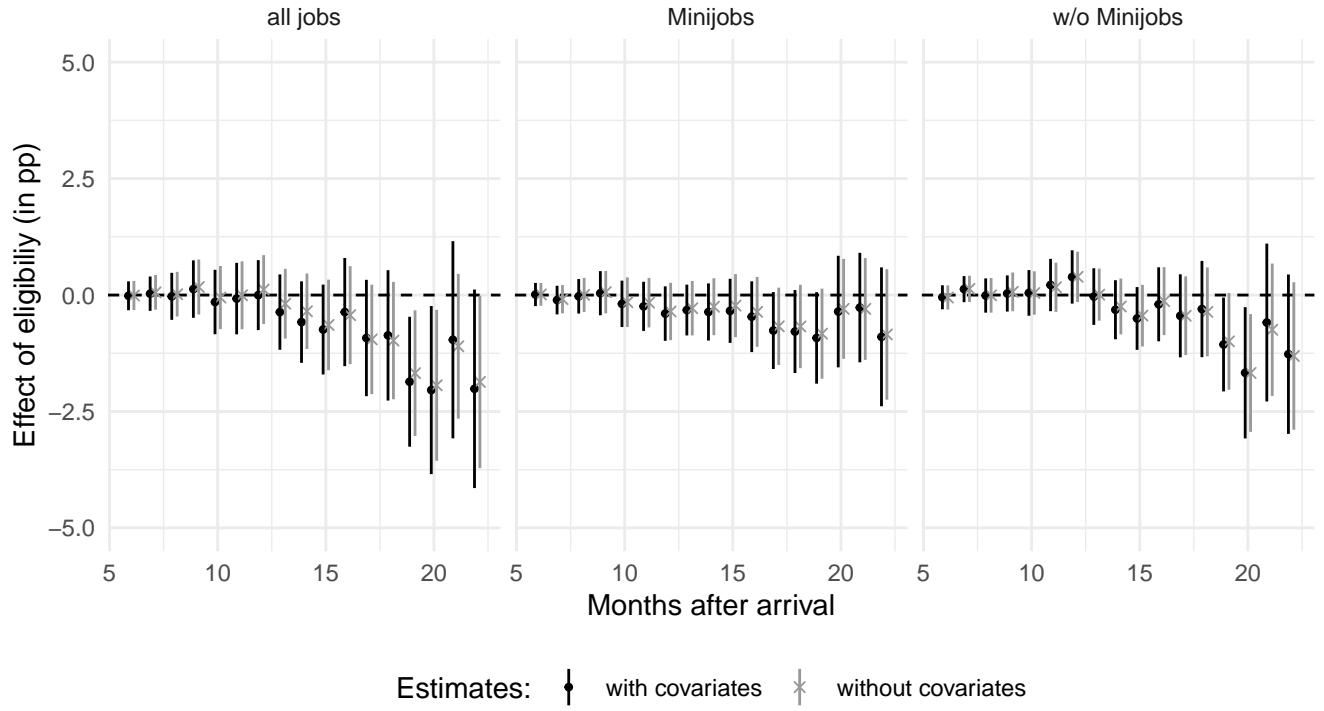


Figure S.5: Estimates for three employment outcomes and corresponding 95% confidence intervals from covariate-adjusted and unadjusted local linear regressions with (MSE)-optimal bandwidth estimating the intention-to-treat effect of the ad-hoc program eligibility at the cut-off date.

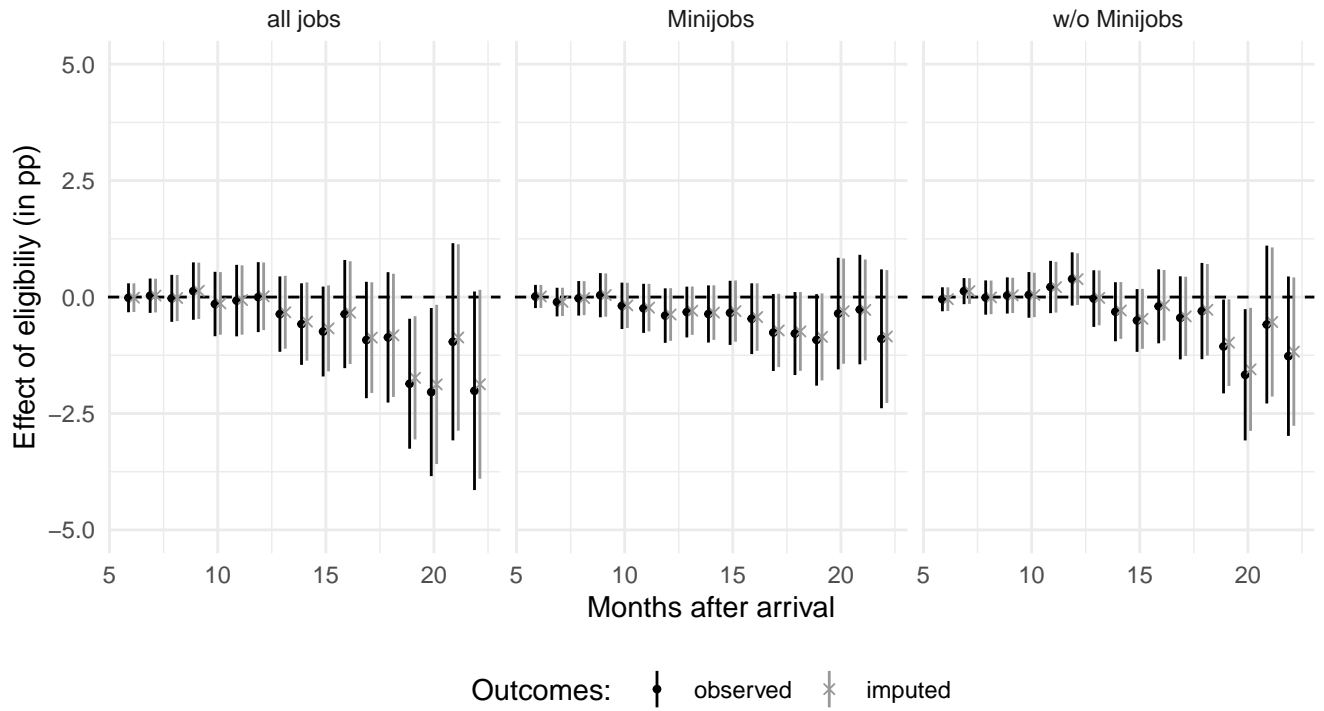


Figure S.6: Estimates for three employment outcomes (observed and imputed) and corresponding 95% confidence intervals from local linear regressions with (MSE)-optimal bandwidth estimating the intention-to-treat effect of the ad-hoc program eligibility at the cut-off date.

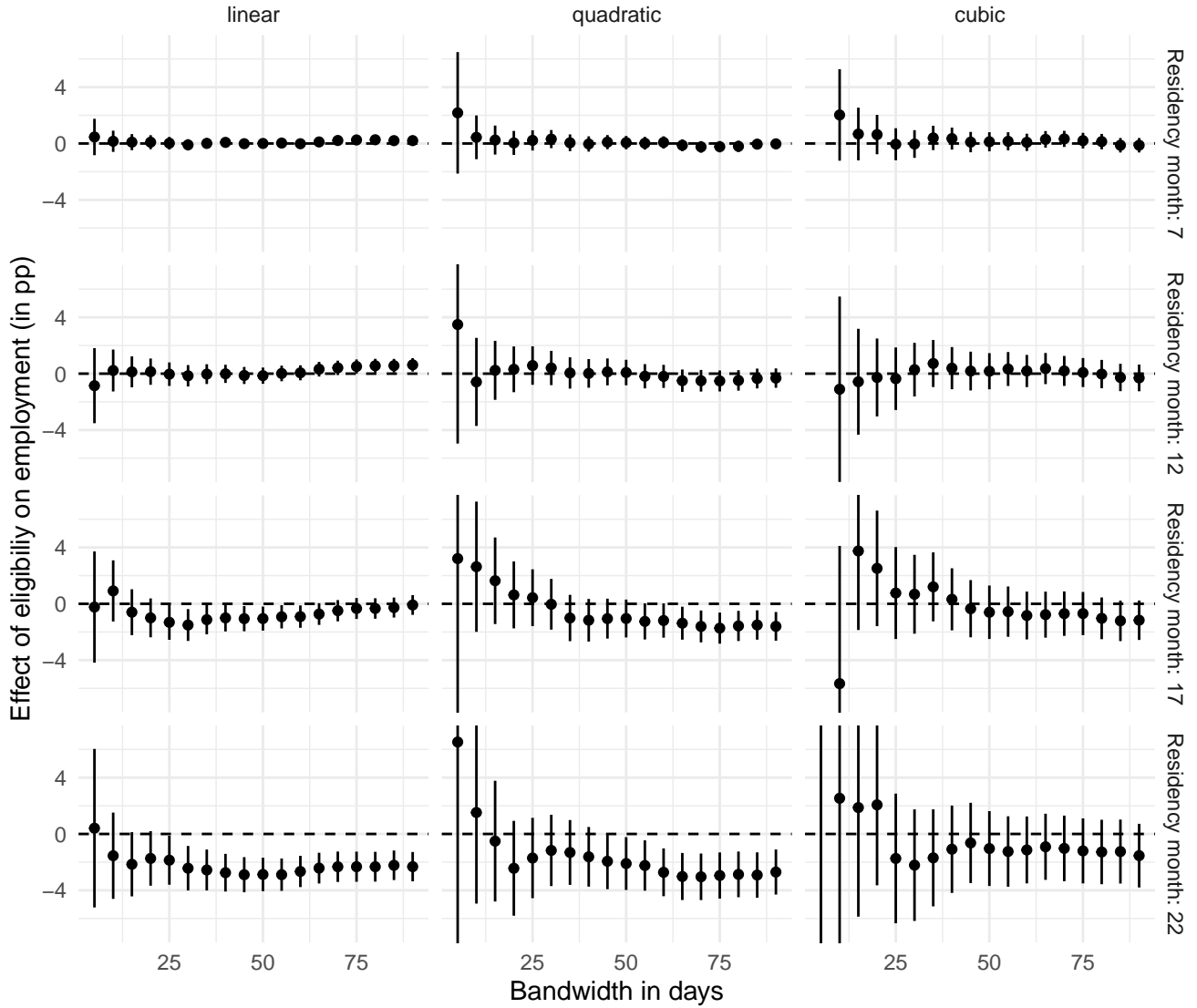


Figure S.7: Sensitivity of the main estimates to different bandwidth choices and polynomials of the local linear regression 7, 12, 17, and 22 month after arrival. The figure shows the estimates and 95% confidence intervals from local linear regressions estimating the intention-to-treat effect of the ad-hoc program eligibility on employment for symmetric bandwidths choices of 5 to 90 days around the cutoff and for polynomials of degree 1, 2 and 3.

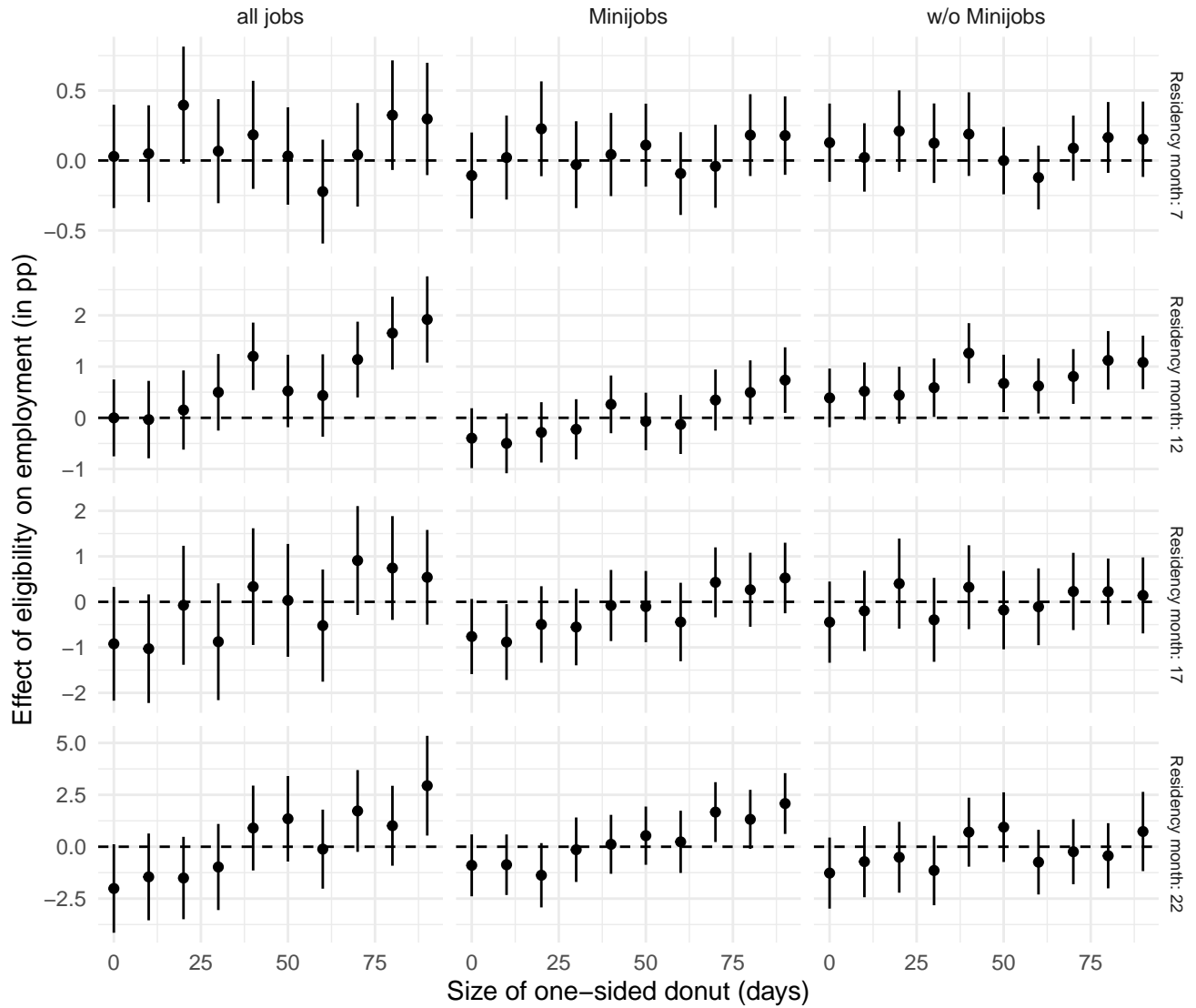


Figure S.8: Sensitivity of the main estimates to different donut sizes in the the local linear regression 7, 12, 17, and 22 month after arrival. The figure shows estimates and 95% confidence intervals from local linear regressions with (MSE)-optimal bandwidth estimating the intention-to-treat effect of program eligibility on employment at the cut-off date for donut sizes of length 0 to 90 days in 10 day increments.

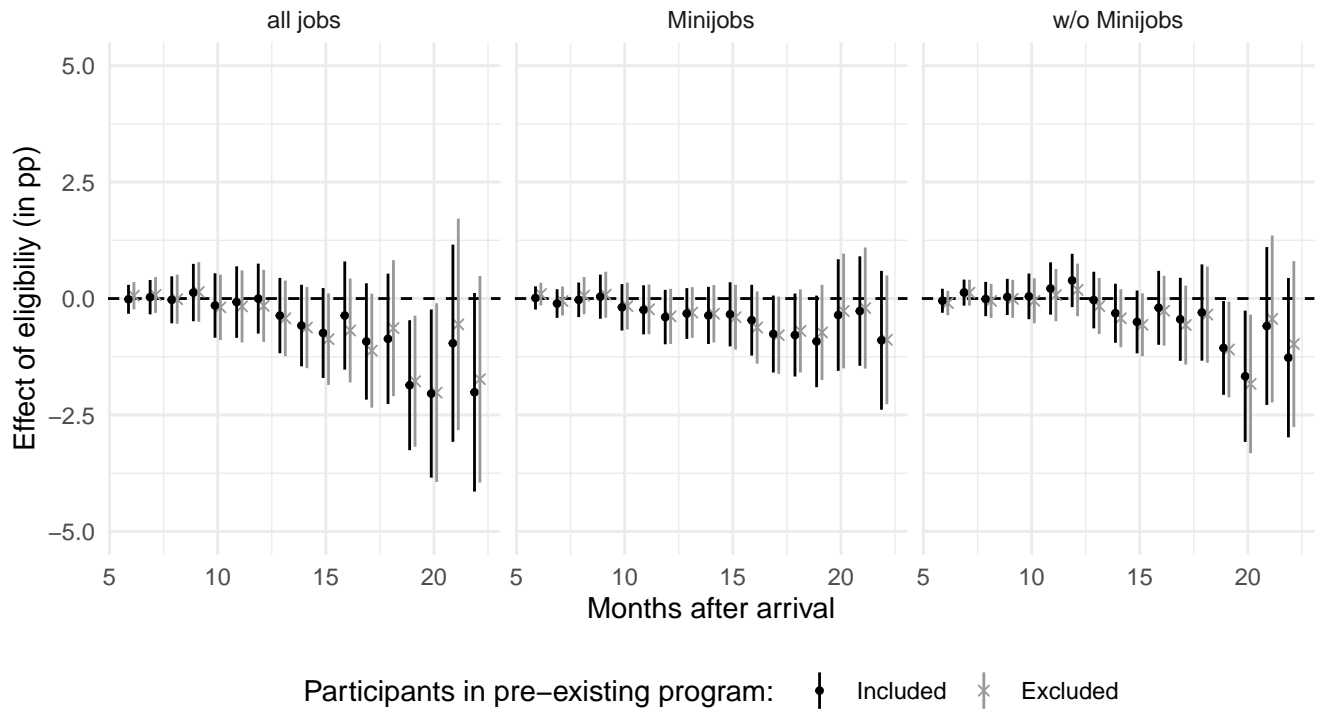


Figure S.9: Estimates for three employment outcomes and corresponding 95% confidence intervals from local linear regressions with (MSE)-optimal bandwidth estimating the intention-to-treat effect of the ad-hoc program eligibility at the cut-off date separately for a sample that includes all individuals and a sample that excludes everyone enrolling in the pre-existing program during the study period.

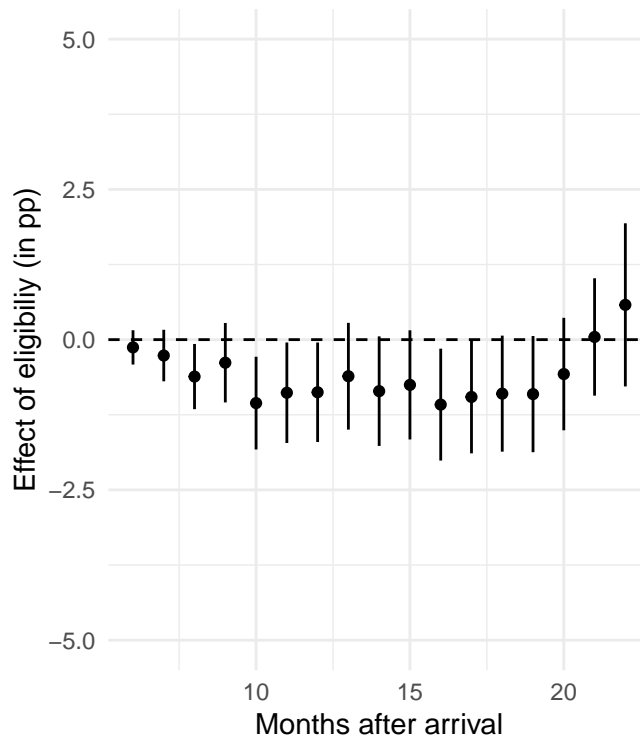


Figure S.10: RDD plots showing the intention-to-treat effect of the ad-hoc program eligibility at the cut-off date on the probability of enrolling in the preexisting program.

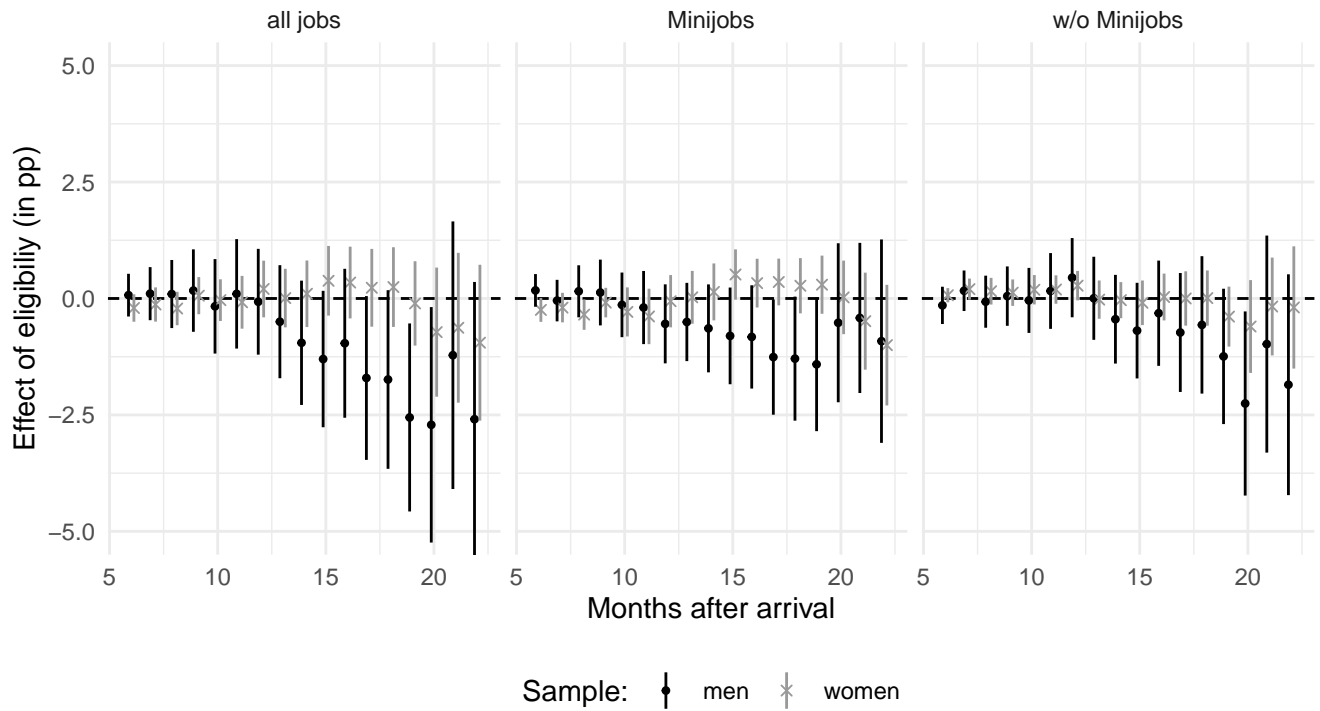


Figure S.11: Estimates for three employment outcomes and corresponding 95% confidence intervals from local linear regressions with (MSE)-optimal bandwidth estimating the intention-to-treat effect of the ad-hoc program eligibility at the cut-off date separately for men and women.

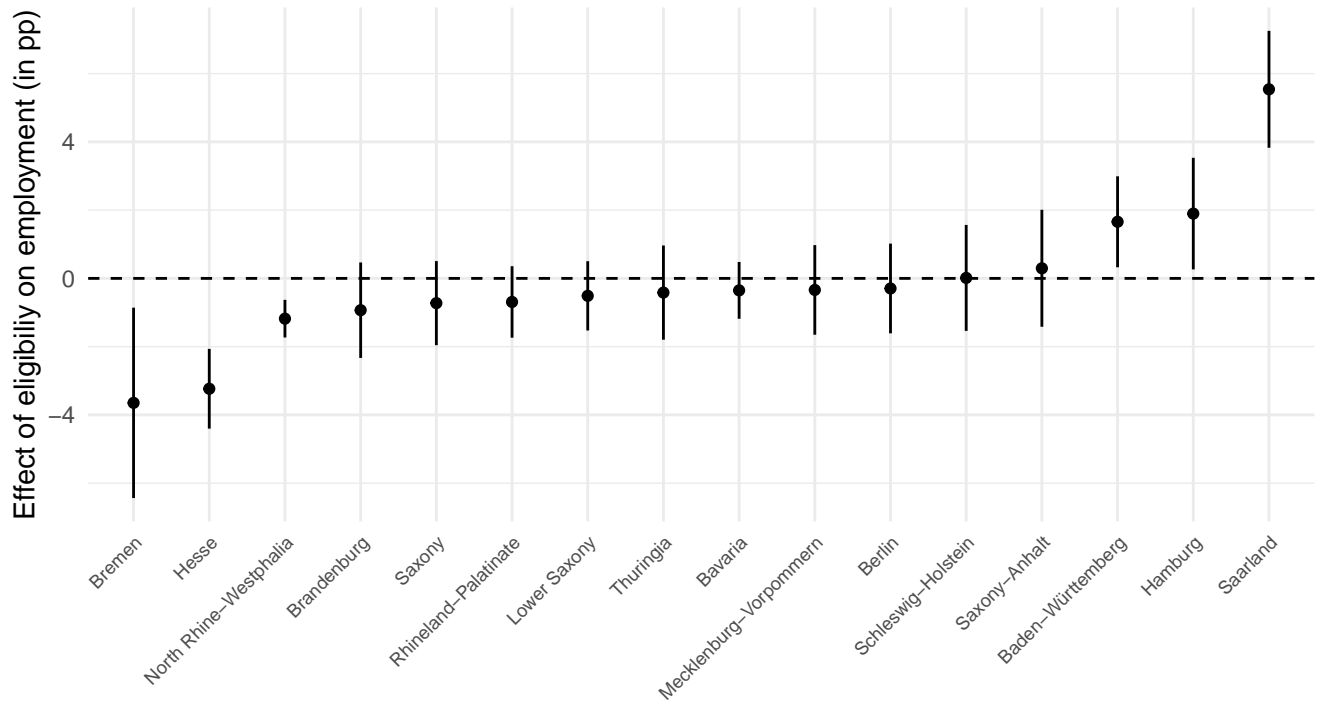


Figure S.12: Heterogeneity of estimates by state of residence and corresponding 95% confidence intervals from local linear regressions with (MSE)-optimal bandwidth estimating the intention-to-treat effect of program eligibility at the cut-off date and pooling all residency periods.

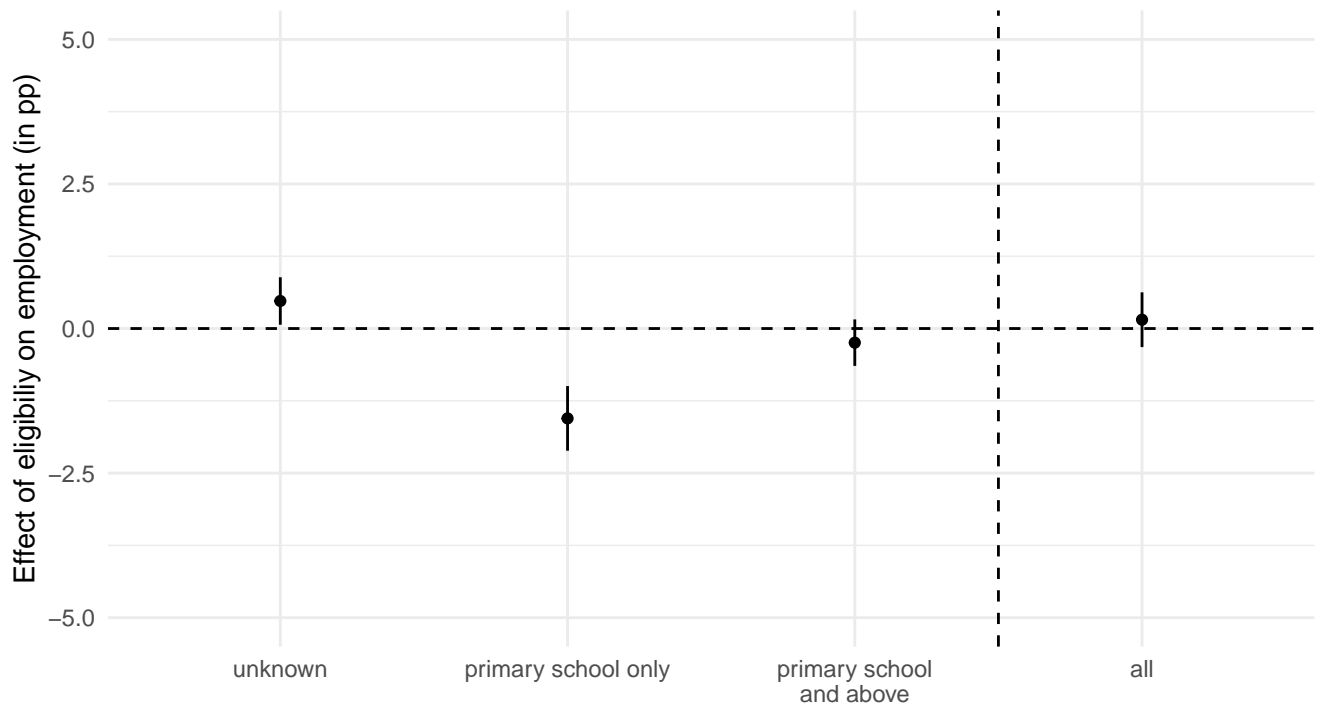


Figure S.13: Heterogeneity of estimates by measured schooling and corresponding 95% confidence intervals from local linear regressions with (MSE)-optimal bandwidth estimating the intention-to-treat effect of program eligibility at the cut-off date and pooling all residency periods.

Detailed Results: Preexisting Language Training Program

- Main Effect Estimates: Tables S.4 and S.5 report the main estimates as displayed in Figure 4 in the main text.
- Tests for Pre-treatment Differences: Table S.6 reports formal F-tests for three selected pre-treatment periods (3 to 1 month before enrollment, 9 to 1 month before enrollment, and 29 to 1 month before enrollment) and all three employment outcomes based on the two-dimensional IFEct estimator. We observe one statistically significant difference in the pre-treatment estimates across the three employment outcomes but the substantive magnitude of this difference is small, as shown in Figure 4 of the main text. Overall, this indicates support for the identification assumption.
- Estimates from the CS DR estimator when varying the control group: Figure S.14 presents the estimates from the CS DR estimator using the not-yet-treated and the never-treated units as controls. Although the numerical estimates differ slightly, the pattern remains identical across all three employment outcomes. This consistency indicates that the results are robust across both control groups.
- Main Effect Estimates Imputing for Attrition: Figure S.15 shows the estimates when imputing a zero for all years post-arrival without an observed employment outcome. The results are very similar, indicating that findings are not an artifact of panel attrition.
- Main Effect Estimates by Gender: Figure S.16 shows the estimates for samples of men and women separately. Although the estimates for women are somewhat elevated in magnitude (and come with larger confidence intervals due to the reduced estimation sample size), the pattern is similar to that observed for men across all three employment outcomes.
- Correlates of Enrollment: Table S.7 describes the correlates of preexisting program enrollment. Although we find no pronounced differences across age groups, Syrians tend to be more likely to participate in the preexisting program compared to Iraqis, Iranians, and Eritreans. Those reported to have at least some schooling and those without any schooling information are somewhat less likely to participate than those with no reported schooling. Women are less likely to participate compared to men. There are also differences across states: in the city states of Berlin, Hamburg, and Bremen, the propensity for enrollment is lower compared to Saarland and Hesse.

| Month | all jobs | | | w/o Minijob | | | Minijob | | |
|-------|----------|------|---------|-------------|------|---------|---------|------|---------|
| | ATT | S.E. | p(> z) | ATT | S.E. | p(> z) | ATT | S.E. | p(> z) |
| -20 | 0.21 | 0.24 | 0.37 | 0.04 | 0.20 | 0.86 | 0.05 | 0.16 | 0.77 |
| -19 | 0.25 | 0.20 | 0.22 | 0.19 | 0.15 | 0.23 | -0.04 | 0.14 | 0.80 |
| -18 | 0.15 | 0.18 | 0.41 | 0.12 | 0.13 | 0.33 | -0.04 | 0.14 | 0.80 |
| -17 | 0.36 | 0.17 | 0.04 | 0.16 | 0.11 | 0.17 | 0.15 | 0.14 | 0.28 |
| -16 | 0.33 | 0.18 | 0.06 | 0.23 | 0.11 | 0.04 | 0.06 | 0.14 | 0.66 |
| -15 | 0.26 | 0.15 | 0.08 | 0.10 | 0.09 | 0.28 | 0.21 | 0.13 | 0.11 |
| -14 | 0.22 | 0.16 | 0.15 | 0.01 | 0.09 | 0.92 | 0.27 | 0.14 | 0.05 |
| -13 | 0.18 | 0.17 | 0.27 | -0.06 | 0.10 | 0.54 | 0.26 | 0.14 | 0.06 |
| -12 | 0.09 | 0.16 | 0.59 | -0.08 | 0.11 | 0.46 | 0.20 | 0.14 | 0.15 |
| -11 | 0.02 | 0.15 | 0.92 | -0.05 | 0.11 | 0.64 | 0.12 | 0.11 | 0.30 |
| -10 | -0.02 | 0.14 | 0.89 | 0.05 | 0.10 | 0.66 | -0.01 | 0.10 | 0.93 |
| -9 | -0.04 | 0.12 | 0.75 | 0.04 | 0.10 | 0.70 | -0.02 | 0.09 | 0.84 |
| -8 | -0.19 | 0.14 | 0.18 | -0.05 | 0.10 | 0.61 | -0.08 | 0.11 | 0.44 |
| -7 | -0.17 | 0.14 | 0.24 | -0.01 | 0.09 | 0.93 | -0.11 | 0.11 | 0.31 |
| -6 | -0.26 | 0.14 | 0.06 | -0.07 | 0.09 | 0.47 | -0.15 | 0.11 | 0.15 |
| -5 | -0.16 | 0.15 | 0.27 | 0.03 | 0.09 | 0.73 | -0.18 | 0.11 | 0.12 |
| -4 | 0.02 | 0.15 | 0.91 | 0.13 | 0.10 | 0.21 | -0.09 | 0.12 | 0.45 |
| -3 | 0.06 | 0.16 | 0.68 | 0.09 | 0.09 | 0.35 | -0.07 | 0.13 | 0.60 |
| -2 | 0.03 | 0.16 | 0.86 | 0.03 | 0.09 | 0.77 | -0.05 | 0.14 | 0.73 |
| -1 | 0.04 | 0.14 | 0.75 | -0.16 | 0.08 | 0.05 | 0.15 | 0.11 | 0.18 |
| 0 | -0.21 | 0.16 | 0.20 | -0.23 | 0.09 | 0.01 | -0.02 | 0.14 | 0.87 |
| 1 | -0.19 | 0.30 | 0.52 | -0.33 | 0.21 | 0.10 | 0.12 | 0.26 | 0.64 |
| 2 | -0.22 | 0.42 | 0.61 | -0.27 | 0.29 | 0.36 | -0.07 | 0.34 | 0.84 |
| 3 | 0.14 | 0.53 | 0.79 | -0.17 | 0.39 | 0.67 | 0.10 | 0.43 | 0.81 |
| 4 | 0.52 | 0.65 | 0.42 | 0.33 | 0.49 | 0.50 | -0.11 | 0.53 | 0.84 |
| 5 | 0.95 | 0.77 | 0.21 | 0.54 | 0.58 | 0.35 | -0.01 | 0.62 | 0.99 |
| 6 | 1.32 | 0.87 | 0.13 | 0.91 | 0.64 | 0.16 | -0.13 | 0.70 | 0.86 |
| 7 | 0.86 | 0.89 | 0.34 | 0.63 | 0.65 | 0.33 | -0.27 | 0.76 | 0.73 |
| 8 | 0.62 | 0.89 | 0.49 | -0.09 | 0.57 | 0.88 | 0.19 | 0.77 | 0.80 |
| 9 | 1.42 | 0.85 | 0.10 | 0.32 | 0.55 | 0.56 | 0.66 | 0.73 | 0.37 |
| 10 | 2.15 | 0.84 | 0.01 | 0.78 | 0.59 | 0.18 | 0.98 | 0.70 | 0.16 |
| 11 | 2.88 | 0.83 | 0.00 | 0.73 | 0.61 | 0.24 | 1.96 | 0.67 | 0.00 |
| 12 | 4.78 | 0.91 | 0.00 | 1.85 | 0.69 | 0.01 | 2.76 | 0.69 | 0.00 |
| 13 | 5.73 | 1.02 | 0.00 | 2.80 | 0.77 | 0.00 | 3.19 | 0.78 | 0.00 |
| 14 | 6.57 | 1.08 | 0.00 | 3.36 | 0.82 | 0.00 | 3.40 | 0.86 | 0.00 |
| 15 | 7.51 | 1.21 | 0.00 | 4.10 | 0.92 | 0.00 | 3.93 | 0.99 | 0.00 |
| 16 | 8.01 | 1.30 | 0.00 | 5.39 | 1.04 | 0.00 | 3.43 | 1.09 | 0.00 |
| 17 | 8.81 | 1.42 | 0.00 | 6.99 | 1.14 | 0.00 | 2.76 | 1.14 | 0.02 |
| 18 | 9.23 | 1.53 | 0.00 | 8.11 | 1.29 | 0.00 | 1.77 | 1.13 | 0.12 |
| 19 | 8.42 | 1.62 | 0.00 | 7.77 | 1.43 | 0.00 | 1.40 | 1.11 | 0.21 |
| 20 | 9.32 | 1.91 | 0.00 | 9.55 | 1.73 | 0.00 | -0.33 | 1.31 | 0.80 |

Table S.4: Point estimates, standard errors and p-values from two-dimensional IFECT estimator for three employment outcomes and twenty pre/post-treatment periods.

| Month | all jobs | | | w/o Minijob | | | Minijob | | |
|-------|----------|------|---------|-------------|------|---------|---------|------|---------|
| | ATT | S.E. | p(> z) | ATT | S.E. | p(> z) | ATT | S.E. | p(> z) |
| -20 | 0.10 | 0.44 | 0.81 | 0.22 | 0.43 | 0.61 | -0.12 | 0.01 | 0.00 |
| -19 | -0.25 | 0.02 | 0.00 | -0.09 | 0.01 | 0.00 | -0.16 | 0.01 | 0.00 |
| -18 | 0.27 | 0.31 | 0.38 | 0.08 | 0.29 | 0.79 | 0.19 | 0.30 | 0.53 |
| -17 | -0.04 | 0.45 | 0.93 | 0.19 | 0.25 | 0.44 | -0.23 | 0.24 | 0.33 |
| -16 | -0.02 | 0.23 | 0.93 | -0.17 | 0.01 | 0.00 | 0.28 | 0.39 | 0.47 |
| -15 | -0.02 | 0.30 | 0.94 | -0.08 | 0.16 | 0.60 | 0.06 | 0.17 | 0.74 |
| -14 | -0.02 | 0.16 | 0.88 | -0.03 | 0.14 | 0.80 | -0.09 | 0.25 | 0.73 |
| -13 | -0.06 | 0.23 | 0.80 | -0.01 | 0.12 | 0.94 | -0.06 | 0.14 | 0.67 |
| -12 | -0.15 | 0.20 | 0.46 | 0.07 | 0.12 | 0.57 | -0.22 | 0.25 | 0.37 |
| -11 | -0.08 | 0.18 | 0.67 | 0.10 | 0.15 | 0.51 | -0.19 | 0.16 | 0.26 |
| -10 | -0.15 | 0.17 | 0.38 | -0.05 | 0.14 | 0.74 | -0.10 | 0.15 | 0.48 |
| -9 | -0.31 | 0.14 | 0.03 | -0.17 | 0.12 | 0.16 | -0.15 | 0.13 | 0.25 |
| -8 | -0.13 | 0.22 | 0.56 | -0.04 | 0.18 | 0.82 | -0.09 | 0.12 | 0.42 |
| -7 | -0.30 | 0.20 | 0.12 | -0.20 | 0.14 | 0.14 | -0.11 | 0.16 | 0.48 |
| -6 | -0.09 | 0.18 | 0.62 | -0.06 | 0.14 | 0.68 | -0.05 | 0.19 | 0.80 |
| -5 | -0.11 | 0.20 | 0.58 | -0.16 | 0.13 | 0.21 | 0.07 | 0.14 | 0.62 |
| -4 | -0.24 | 0.20 | 0.23 | -0.35 | 0.13 | 0.01 | 0.07 | 0.17 | 0.67 |
| -3 | -0.40 | 0.21 | 0.06 | -0.45 | 0.12 | 0.00 | 0.06 | 0.19 | 0.74 |
| -2 | -0.38 | 0.21 | 0.07 | -0.66 | 0.13 | 0.00 | 0.25 | 0.19 | 0.18 |
| -1 | -0.70 | 0.22 | 0.00 | -0.60 | 0.13 | 0.00 | -0.13 | 0.19 | 0.52 |
| 0 | -0.43 | 0.25 | 0.09 | -0.62 | 0.16 | 0.00 | 0.19 | 0.21 | 0.37 |
| 1 | -0.84 | 0.35 | 0.01 | -0.97 | 0.22 | 0.00 | 0.05 | 0.28 | 0.87 |
| 2 | -0.85 | 0.36 | 0.02 | -1.28 | 0.24 | 0.00 | 0.34 | 0.33 | 0.31 |
| 3 | -0.72 | 0.45 | 0.11 | -1.09 | 0.32 | 0.00 | 0.23 | 0.33 | 0.48 |
| 4 | -0.28 | 0.53 | 0.60 | -0.84 | 0.35 | 0.01 | 0.48 | 0.39 | 0.22 |
| 5 | 0.27 | 0.55 | 0.63 | -0.54 | 0.41 | 0.19 | 0.67 | 0.42 | 0.11 |
| 6 | -0.17 | 0.58 | 0.77 | -0.95 | 0.38 | 0.01 | 0.66 | 0.48 | 0.17 |
| 7 | 0.28 | 0.63 | 0.66 | -1.21 | 0.42 | 0.00 | 1.36 | 0.53 | 0.01 |
| 8 | 0.99 | 0.68 | 0.15 | -0.35 | 0.48 | 0.47 | 1.20 | 0.54 | 0.03 |
| 9 | 1.50 | 0.76 | 0.05 | 0.20 | 0.57 | 0.72 | 1.18 | 0.60 | 0.05 |
| 10 | 2.36 | 0.75 | 0.00 | 0.49 | 0.63 | 0.43 | 1.93 | 0.60 | 0.00 |
| 11 | 4.01 | 0.87 | 0.00 | 1.67 | 0.66 | 0.01 | 2.40 | 0.72 | 0.00 |
| 12 | 4.42 | 0.92 | 0.00 | 2.53 | 0.72 | 0.00 | 2.23 | 0.69 | 0.00 |
| 13 | 4.78 | 0.97 | 0.00 | 2.93 | 0.82 | 0.00 | 2.03 | 0.71 | 0.00 |
| 14 | 5.66 | 1.08 | 0.00 | 3.16 | 0.90 | 0.00 | 2.87 | 0.84 | 0.00 |
| 15 | 5.98 | 1.21 | 0.00 | 4.18 | 0.91 | 0.00 | 2.35 | 0.85 | 0.01 |
| 16 | 6.38 | 1.37 | 0.00 | 5.12 | 1.19 | 0.00 | 2.02 | 0.96 | 0.03 |
| 17 | 7.15 | 1.48 | 0.00 | 6.26 | 1.27 | 0.00 | 1.30 | 1.09 | 0.23 |
| 18 | 6.12 | 1.59 | 0.00 | 5.77 | 1.48 | 0.00 | 0.90 | 1.20 | 0.45 |
| 19 | 7.09 | 1.81 | 0.00 | 6.87 | 1.69 | 0.00 | 0.12 | 1.35 | 0.93 |
| 20 | 9.32 | 2.37 | 0.00 | 7.55 | 2.23 | 0.00 | 2.10 | 1.84 | 0.25 |

Table S.5: Point estimates, standard errors and p-values from CS DR estimator for three employment outcomes and twenty pre/post-treatment periods.

| Outcome | Period | F-Stat | p-value |
|--------------|---------|--------|---------|
| all jobs | [-3,-1] | 0.13 | 0.94 |
| all jobs | [-9,-1] | 0.83 | 0.59 |
| all jobs | [29,-1] | 0.86 | 0.42 |
| w/o Minijobs | [-3,-1] | 1.41 | 0.24 |
| w/o Minijobs | [-9,-1] | 0.79 | 0.62 |
| w/o Minijobs | [29,-1] | 5.27 | 0.01 |
| Minijobs | [-3,-1] | 0.61 | 0.61 |
| Minijobs | [-9,-1] | 1.13 | 0.34 |
| Minijobs | [29,-1] | 0.87 | 0.42 |

Table S.6: Testing for pre-treatment trends across three employment outcomes and three different definitions of pre-treatment period. The F-statistic and corresponding p-value are based on the two-dimensional IFECT estimator.

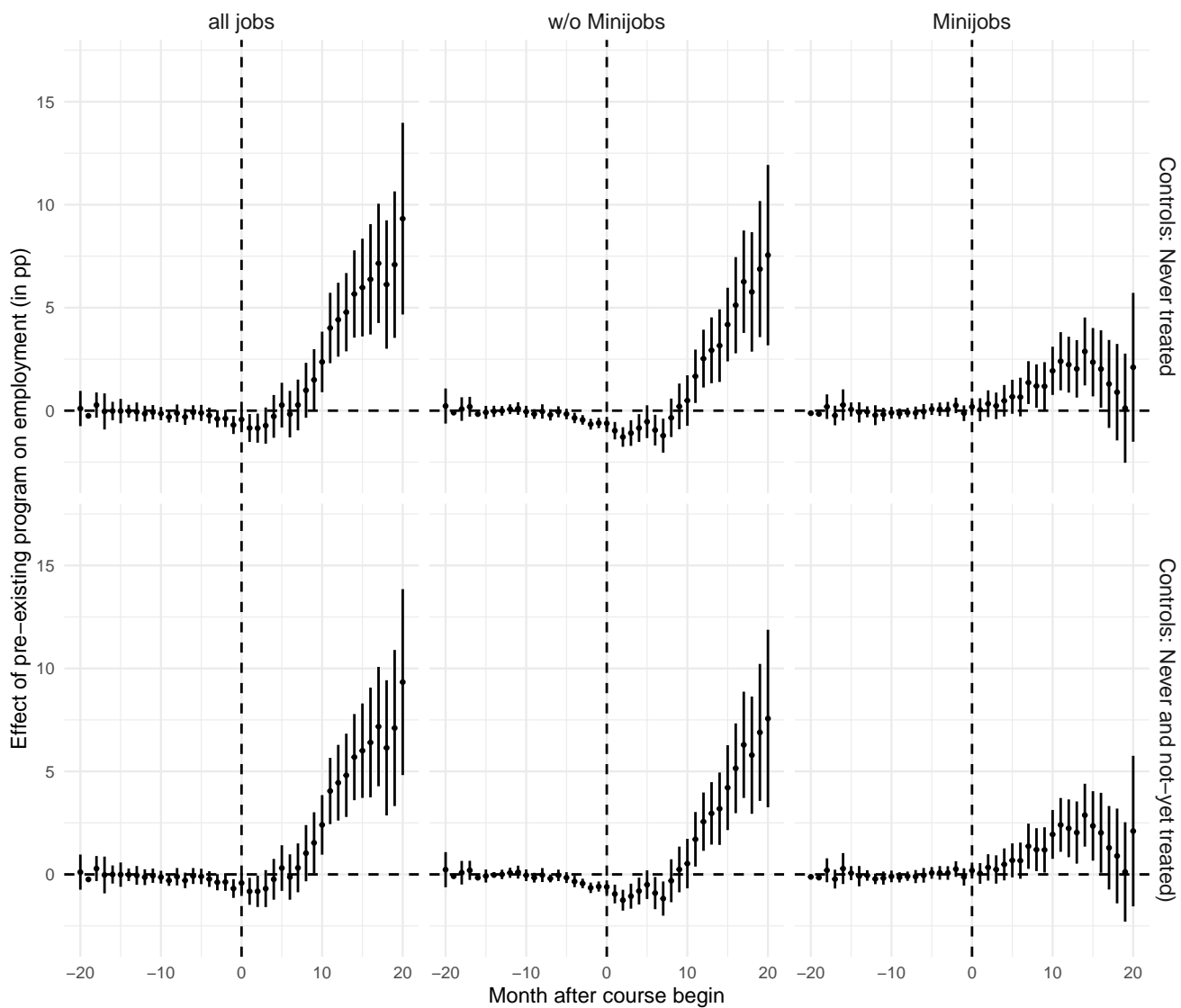


Figure S.14: Estimates and corresponding 95% confidence intervals from the CS DR estimator using either never-treated or never and not-yet treated persons as control group.

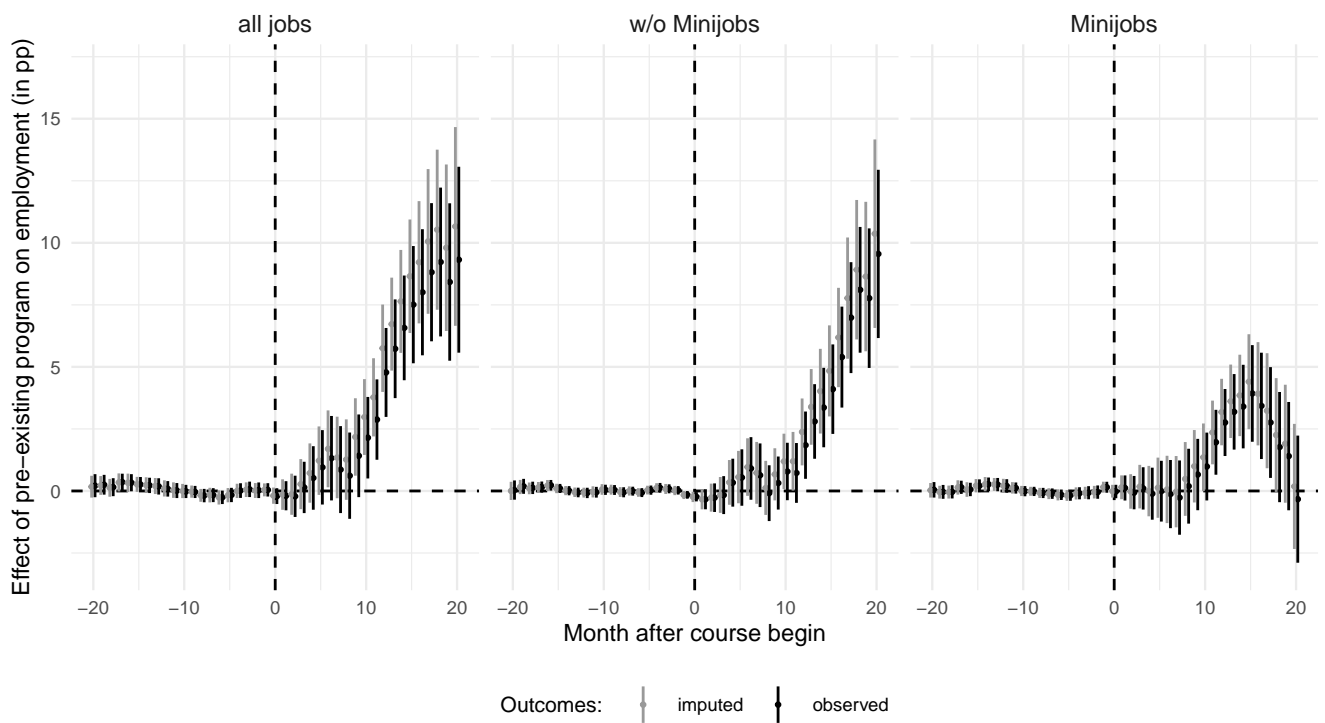


Figure S.15: Estimates and corresponding 95% confidence intervals from the two-dimensional IFECT estimator using either observed and imputed outcome.

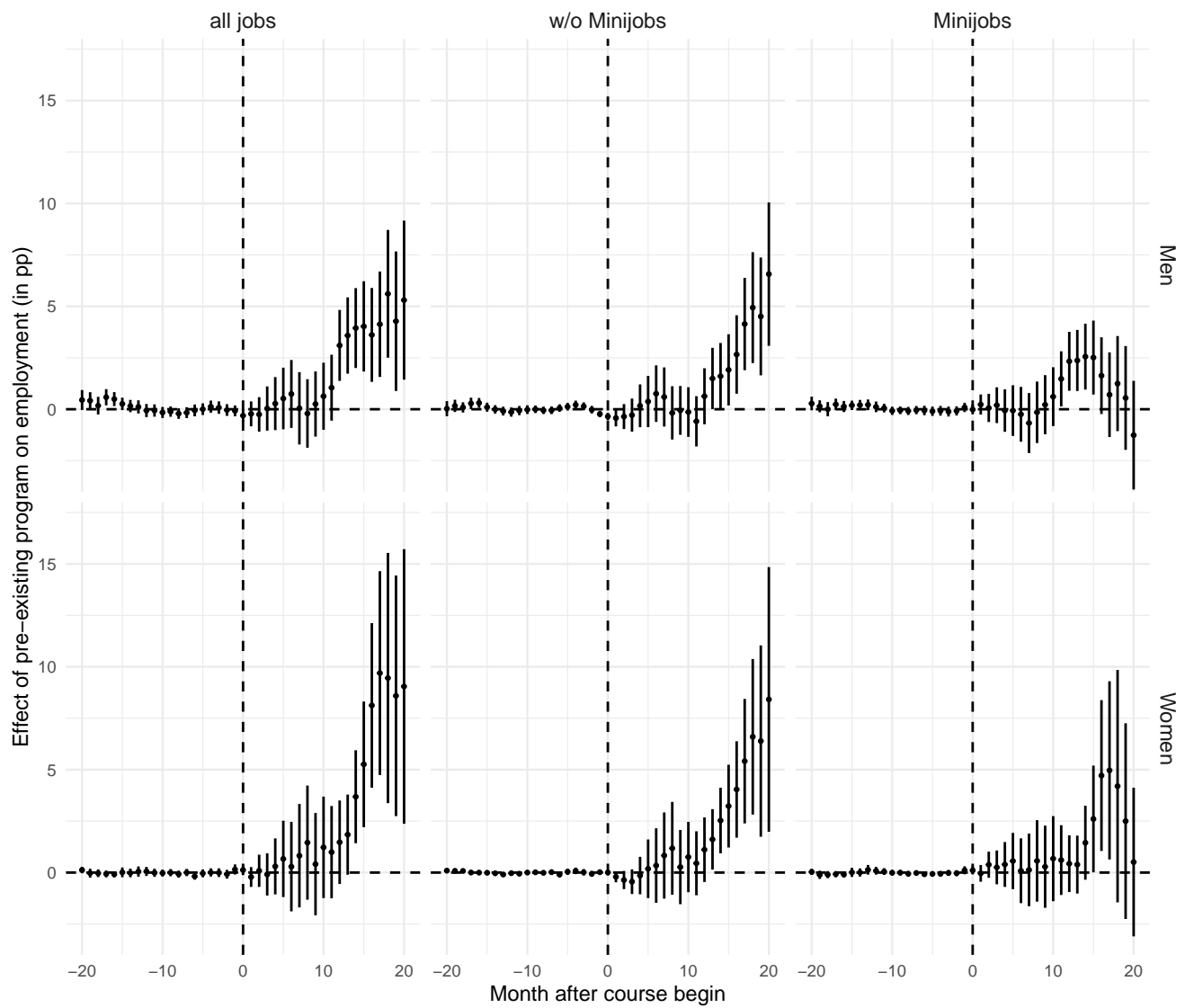


Figure S.16: Estimates and corresponding 95% confidence intervals from the two-dimensional IFEct estimator separate for men and women.

| | Model 1 | Model 2 |
|--------------------------|-----------------|-----------------|
| Age | | |
| 18-20 | | |
| 21-25 | 0.01 (0.00)* | 0.01 (0.00)* |
| 26-30 | −0.00 (0.00) | −0.00 (0.00) |
| 31-35 | −0.00 (0.00) | −0.00 (0.00) |
| Nationality | | |
| Syria | | |
| Iraq | −0.03 (0.00)*** | −0.03 (0.00)*** |
| Iran | −0.04 (0.00)*** | −0.04 (0.00)*** |
| Eritrea | −0.02 (0.01)** | −0.01 (0.01) |
| Schooling | | |
| No | | |
| Yes | −0.01 (0.00)*** | −0.01 (0.00)*** |
| (Missing) | −0.09 (0.00)*** | −0.09 (0.00)*** |
| Gender | | |
| Female | −0.03 (0.00)*** | −0.03 (0.00)*** |
| State of first residency | | |
| Schleswig-Holstein | | |
| Hamburg | | −0.02 (0.00)*** |
| Niedersachsen | | 0.07 (0.01)*** |
| Bremen | | −0.02 (0.01)*** |
| Nordrhein-Westfalen | | 0.06 (0.00)*** |
| Hessen | | 0.11 (0.01)*** |
| Rheinland-Pfalz | | 0.03 (0.01)*** |
| Baden-Württemberg | | 0.04 (0.01)*** |
| Bayern | | 0.01 (0.00)* |
| Saarland | | 0.15 (0.01)*** |
| Berlin | | −0.04 (0.00)*** |
| Brandenburg | | 0.07 (0.01)*** |
| Mecklenburg-Vorpommern | | 0.06 (0.01)*** |
| Sachsen | | 0.02 (0.01)*** |
| Sachsen-Anhalt | | 0.04 (0.01)*** |
| Thüringen | | 0.06 (0.01)*** |
| (Intercept) | 0.11 (0.00)*** | 0.07 (0.01)*** |
| N | 1335073 | 1335073 |
| R2 | 0.02 | 0.04 |

*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$

Table S.7: OLS estimates from regressing a time-constant indicator of preexisting program enrollment on time-constant demographic characteristics. Standard errors are clustered by person.

References

- [1] A. I. Barreca, M. Guldi, J. M. Lindo, and G. R. Waddell. Saving Babies? Revisiting the Effect of Very Low Birth Weight Classification. The Quarterly Journal of Economics, 126(4):2117–2123, 2011.
- [2] Bundesagentur für Arbeit. Presseinfo Nr. 48: Sprachförderung als Basis für Integration in Arbeit, 2016.
- [3] Bundesagentur für Arbeit. Weisung 201612003 vom 12.12.2016 - Überarbeitung von Bewerberdatensätzen zum Aufenthaltsstatus und Arbeitsmarktzugang im IT-Verfahren VerBIS, 2016.
- [4] S. Calonico, M. D. Cattaneo, D. M. Farrell, and R. Titiunik. rdrobust: Software for regression-discontinuity designs. The Stata Journal, 17(2):372–404, 2017.
- [5] S. Calonico, M. D. Cattaneo, and R. Titiunik. Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. Econometrica, 82(6):2295–2326, 2014.
- [6] M. D. Cattaneo, M. Jansson, and X. Ma. Simple Local Polynomial Density Estimators. Journal of the American Statistical Association, 115(531):1449–1455, 2020.
- [7] A. C. Eggers, R. Freier, V. Grembi, and T. Nannicini. Regression Discontinuity Designs Based on Population Thresholds: Pitfalls and Solutions. American Journal of Political Science, 62(1):210–229, 2018.