RUHR
ECONOMIC PAPERS

Matthias Giesecke
Eric Schuss

Heterogeneity in Marginal Returns to Language Training of Immigrants

rwi
#812

Imprint

Ruhr Economic Papers

Published by

RWI – Leibniz-Institut für Wirtschaftsforschung
Hohenzollernstr. 1-3, 45128 Essen, Germany

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Editors

Prof. Dr. Thomas K. Bauer
RUB, Department of Economics, Empirical Economics
Phone: +49 (0) 234/3 22 83 41, e-mail: thomas.bauer@rub.de

Prof. Dr. Wolfgang Leininger
Technische Universität Dortmund, Department of Economic and Social Sciences
Economics – Microeconomics
Phone: +49 (0) 231/7 55-3297, e-mail: W.Leininger@tu-dortmund.de

Prof. Dr. Volker Clausen
University of Duisburg-Essen, Department of Economics
International Economics
Phone: +49 (0) 201/1 83-3655, e-mail: vclausen@vwl.uni-due.de

Prof. Dr. Roland Döhrn, Prof. Dr. Manuel Frondel, Prof. Dr. Jochen Kluve
RWI, Phone: +49 (0) 201/81 49-213, e-mail: presse@rwi-essen.de

Editorial Office

Sabine Weiler
RWI, Phone: +49 (0) 201/81 49-213, e-mail: sabine.weiler@rwi-essen.de

Ruhr Economic Papers #812

Responsible Editor: Jochen Kluve

ISSN 1864-4872 (online) – ISBN 978-3-86788-941-4

Ruhr Economic Papers #812

Matthias Giesecke and Eric Schuss

Heterogeneity in Marginal Returns to Language Training of Immigrants

RWI

Bibliografische Informationen
der Deutschen Nationalbibliothek

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at **http://dnb.dnb.de**

RWI is funded by the Federal Government and the federal state of North Rhine-Westphalia.

http://dx.doi.org/10.4419/86788941
ISSN 1864-4872 (online)
ISBN 978-3-86788-941-4

Matthias Giesecke and Eric Schuss[1]

Heterogeneity in Marginal Returns to Language Training of Immigrants

Abstract

We estimate the effect of language training on subsequent employment and wages of immigrants under essential heterogeneity. The identifying variation is based on regional differences in language training availability that we use to instrument endogenous participation. Estimating marginal treatment effects along the distribution of observables and unobservables that drive individual participation decisions, we find that immigrants with higher gains are more likely to select into language training than immigrants with lower gains. We document up to 15% higher employment rates and 13% wage gains for immigrants with a high desire to participate but the positive returns vanish with increasing resistance to treatment. This pattern of selection on gains correlates with unobserved ability and motivation, promoting investments in education and job-specific skills that yield higher returns when complemented by language capital in the host country.

JEL Classification: F22, J24, J61, J68, O15

Keywords: Language training; heterogeneous returns; marginal treatment effects; continuous instrument

August 2019

1 Matthias Giesecke, RWI; Eric Schuss, IAB. *– We are grateful to John Friedman, Chris Walters and Matthias Westphal for helpful suggestions and thank the participants of several seminars and conferences for insightful comments and discussions. Matthias Giesecke thanks the faculty associates of the Center for Labor Economics at UC Berkeley for their hospitality while writing substantial parts of this paper.* *– All correspondence to: Matthias Giesecke, RWI, Hohenzollernstr. 1-3, 45128 Essen, Germany, e-mail: matthias.giesecke@rwi-essen.de*

1 Introduction

Language capital is key for the productivity of immigrants in their host country. Mastering the language of the host country complements many skill components that are important for successful labor market integration. The average effect of language training on employment and wages is informative about the overall impact of such programs but estimating these parameters still conceals substantial heterogeneity in the returns to language training. Immigrants with higher unobservable ability in language acquisition may benefit more than others and their returns are likely to be higher if they have complementing job-specific skills from pre-migration education.

In this paper, we estimate the effects of language training on subsequent employment and wages of immigrants under essential heterogeneity. To relate heterogeneous treatment effects to unobserved heterogeneity in the propensity for language course participation, we estimate marginal treatment effects (MTE) by means of parametric and semi-parametric techniques. The expansion of mandatory language training in Germany creates variation regarding the actual availability of open slots in language courses across counties. Due to constrained capacities in some regions, the supply falls behind the demand for language training and thus induces substantial excess demand. Exploiting this regional variation, we construct a continuous *language training availability* instrument (henceforth LTA instrument) to identify the local average treatment effect (LATE) of language training within the typical instrumental variable (IV) framework. Next, we estimate MTEs along the distribution of unobservables that drive individual treatment decisions. The analysis is based on a unique data set that links employment records from administrative data to survey information on the individual level.[1]

We find moderate average returns from language training of immigrants when considering the average treatment effect on the treated (ATT) and negligibly small returns

[1]In particular, we use the IAB-SOEP Migration Sample linked with Integrated Employment Biographies (IAB-SOEP-MIG-ADIAB). It is a 1:1 match of survey data from the German Socio-Economic Panel (reporting language skills, language course participation, pre-migration employment history and many socioeconomic variables) to the individuals' labor market biography, reporting the outcomes employment and wages.

regarding the LATE.[2] However, these parameters mask substantial heterogeneity which we uncover by estimating the effects for the marginal immigrant at the point of indifference for participation in language training. We document considerable positive returns from language training for those immigrants with the highest desire for language training. Among immigrants who are most ready to take the treatment, we estimate a significant 6 percentage point increase in employment[3] and gains of up to 13% in daily wages. These estimates are long-run outcomes, measured eight years after arrival. The estimated MTE curve slopes downward so that the effects vanish for those with higher latent costs of participation. This pattern reflects selection on gains, implying that immigrants with a strong desire for language training are those who benefit the most.

Our estimates are robust against several sensitivity checks. In particular, we show how the results differ when varying the follow-up period after arrival. Consistent with recent evidence on the effects of active labor market programs, the returns to language training are more pronounced in the long-run.[4] We further demonstrate that our results are robust against regional heterogeneity in financial endowment and selective movements of immigrants across regions.

We make four major contributions to the literature. First, we extend a large body of literature on the effects of language skills (Dustmann and van Soest, 2001, 2002; Bleakley and Chin, 2004, 2010; Miranda and Zhu, 2013; Yao and van Ours, 2015) and language training (Sarvimäki and Hämäläinen, 2016; Aslund and Engdahl, 2018; Lang, 2018; Lochmann et al., 2019) on a range of labor market outcomes and economic integration. These studies do not agree in all details but by and large they do find that language skills and language-related trainings have a considerable positive impact on the labor market success of immigrants. We add to this literature by explicitly allowing for effect heterogeneity to learn more about how the labor market effects of language training differ across individuals at the margin of indifference for participation. Answering the question

[2]The average effect among participants (ATT) reports a significant increase of 0.4 percentage points in employment and a 2.5% increase in wages. The LATE does not significantly differ from zero regarding both outcomes.

[3]This translates into a 15% increase when evaluating at the sample mean employment rate of 41%, considering regular employment that is subject to social security contributions.

[4]See Card et al. (2018) for a meta-analysis of the literature on active labor market programs.

who benefits the most, or not at all, is important for the design of language schemes for immigrants in the future.

Second, by estimating MTEs we adopt a research design that is more flexible than typical IV estimates and thereby allows to uncover treatment effect heterogeneity. Established by Björklund and Moffitt (1987) and generalized by Heckman and Vytlacil (1999, 2005, 2007), the MTE approach ranges somewhere in between reduced form and structural methods (see Cornelissen et al., 2016, for a summary). It has been adopted in a recent surge of studies uncovering heterogeneity in the returns to higher education[5] and regarding the effects of child care programs[6]. Our study is the first one to adopt the MTE framework for estimating the labor market returns to language training among immigrants. Previous studies on the impact of language on labor market outcomes either instrument endogenous language skills[7] or use training intensity to instrument training participation[8] or identify training effects at known policy discontinuities[9]. The parameters identified in these studies all rely on different types of LATEs. Estimating marginal returns to language training extends this literature because it permits to recover the full range of policy-relevant treatment parameters such as the population average treatment effect (ATE), the average treatment effect on the treated (ATT), and the average treatment effect on the untreated (ATU).

Third, examining language training effects along the distribution of unobservables that

[5]Examples are (Carneiro et al., 2011) on wage returns from college education in the U.S., (Nybom, 2017) on lifetime earnings returns to college in Sweden, and (Kamhöfer et al., 2019) on non-monetary and wage returns to higher education in Germany.

[6]Recent studies examine the cost-effectiveness of an early childhood education program (Head Start) in the U.S. (Kline and Walters, 2016), the effects of a universal child care (preschool) program on school readiness in Germany (Cornelissen et al., 2018) or the effects of early child care on child development in Germany (Felfe and Lalive, 2018).

[7]Age-at-arrival instruments exploit the finding from the psychobiological literature that childrens' language acquisition is easier compared to adults or adolescents. Studies that use this relationship to estimate the effects of language skills on labor market outcomes are e.g. Bleakley and Chin (2004) for the U.S. and Yao and van Ours (2015) for the Netherlands.

[8]Lang (2018) uses variation in training intensity across job agencies to instrument endogenous participation decisions, based on a language training program for unemployed immigrants in Germany.

[9]Sarvimäki and Hämäläinen (2016) examine a discontinuity from modifying active labor market policies in Finland towards improved integration of immigrants, including language training, to estimate earnings effects. Lochmann et al. (2019) study a threshold in language test scores that introduces a discontinuity in the probability of assignment to language training for immigrants upon arrival in France, estimating the effects on labor force participation.

determine selection into treatment has an attractive economic interpretation. Essentially, the MTE is a willingness to pay parameter for individuals at the margin of indifference between participating or not (see Brinch et al., 2017, who elaborate on this argument more generally). In contrast to only estimating the LATE, the MTE approach permits us to study effect heterogeneity across the entire population of immigrants with at least partial knowledge of their expected gains. Not only indicate our results that immigrants with higher gains are more likely to select into language training, but they are also more likely to have invested in medium or high education. Based on these findings, we can draw a particularly informative policy conclusion: pushing highly reluctant immigrants into participation through expanding the program will only yield positive returns if the policy is complemented by job-specific training. Since language capital complements other forms of human capital[10], complementary training would raise individual returns and thus the willingness to pay for language training. Otherwise, the marginal immigrant may not be able to adequately transmit her language skills to the labor market.

Fourth, our results can be used to evaluate the cost-effectiveness of the program. While language courses are perceived to be highly beneficial for immigrants, the optimal training intensity can only be determined by contrasting the marginal benefits to the marginal costs of the program. Among participants, we estimate average wage gains of 492 EUR per year that would outbalance the marginal costs of training an additional immigrant of 1,230 EUR after less than three years in employment.[11] The programs' benefit-cost ratio is thus pushed well above one after a fairly short period. From this we conclude that the expansion of language training would be worthwhile in terms of long-run cumulative wage gains, raising lifetime earnings and thus governmental tax revenues.

The remainder of this paper is structured as follows. Section 2 provides a brief overview on institutional details of language training in Germany. Section 3 outlines the research

[10]Chiswick and Miller (2002, 2003) provide evidence on the complementarity between language skills and other types of human capital. For further discussion of the role of language capital in the context of migration and education, see Dustmann and Glitz (2011) (pp. 56).

[11]Relating the average wage gains among participants of 2.5% (ATT) to the sample mean of annual wages of 19,682 EUR in 2014 yields annual wage gains from language training of 492 EUR. The training costs per additional immigrant of 1,230 EUR are documented in Federal Government (2007) (p. 66), for details see section 6.

design. Section 4 describes the data and provides details about all important variables including the instrument. Section 5 presents the main results and sensitivity checks. Section 6 presents treatment parameters and the cost-benefit analysis. Section 7 concludes.

2 Institutional Background and Exogenous Variation

2.1 Language Training of Immigrants in Germany

Provision and assignment to language courses is legislated in a specific migration policy that makes language training mandatory for some immigrants upon arrival. The German Law of Immigration (GLI)[12] came into effect on January 1, 2005, aiming to better integrate immigrants. The core part of this policy are integration courses that are subsidized by governmental funds and primarily intend to improve the command of the German language of immigrants through language training. Other components of the integration courses also convey values and knowledge on everyday life in Germany.[13] Both the conception and coordination of language courses are determined at the national level, ensuring nationwide quality standards that allow considering the contents to be homogeneous across regions.

According to the GLI, attendance in language training is generally mandatory for non-EU citizens, if they are classified to have a special need for integration (*besonders integrationsbedürftig*). Probably the most important reason for being assigned to language training is having insufficient German language skills that are determined by an initial placement test. Moreover, immigrants who receive government transfers (unemployment assistance, welfare benefits or benefits from the Asylum Seekers Benefits Act (*Asylbewerberleistungsgesetz, AsylbLG*)) are also obliged to participate in language courses. Refusals to attend the language training can be sanctioned by cutting social benefits or withdrawal

[12] *Gesetz zur Steuerung und Begrenzung der Zuwanderung und zur Regelung des Aufenthalts und der Integration von Unionsbürgern und Ausländern (Zuwanderungsgesetz, ZuWG)*. The exact legal text of the corresponding law is documented in §44 ZuWG (legal claim) and §44a ZuWG (mandatory assignment). For further details on language courses, see appendix A.

[13] Details on the contents and proceedings of integration courses are defined in the corresponding enactment (*Integrationskursverordnung, IntV*, December 31, 2004).

of the migrant's residency permit (sanctions were legitimated in a novel of the GLI from 2007). In addition to mandatory assignment, the law also warrants a legal claim for voluntary participation in integration courses for immigrants who arrived in Germany after December 2004. Voluntary participation is possible for EU citizens and for those non-EU citizens who are not mandated to participate because they are not classified to have special needs for integration. The only requirement for voluntary participation is that the immigrant permanently resides in Germany and attains a legal residency status for the first time.[14]

The institutional details on language training of immigrants in Germany feature some properties that are central to the empirical analysis. The fact that immigrants are primarily obliged to attend in language training if they do not have German language skills or if they obtain government transfers implies that this group is negatively selected in terms of labor market outcomes. When comparing these participants to a more positively selected pool of non-participants, this can lead to biased estimates of a causal impact of language training on employment. To circumvent these pitfalls we exploit that, until today, a considerable share of immigrants effectively does not participate in language training. The effective participation rate depends on regional characteristics and varies substantially across counties (*Landkreise*) such that a universal treatment of obliged immigrants never took place. In what follows, we explain how estimating the returns to language training builds on this regional variation in language course availability.

2.2 Exogenous Variation in Language Training Across Counties

A key feature of the GLI is that it induces regional heterogeneity in language course coverage that is arguably exogenous. In many counties the supply falls behind the demand for language training due to constrained capacities, thus creating excess demand.[15] In

[14]The claim for participation expires two years after the attainment of the first residency permit. Note also that some immigrants are excluded, especially if there is no obvious need for integration (e.g. highly qualified immigrants, see §4 (2) IntV).

[15]Aggregate statistics on integration courses suggest that excess demand for language training is substantial. In 2014, our most recent observation year, the number of admissions and obligations of 211,000 immigrants contrasted to only 142,000 new participants (Federal Office for Migration and Refugees, 2017), indicating excess demand of almost 50%.

these cases, not every immigrant is treated by language training even if the assignment is mandatory. Wherever the demand is high relative to the supply of open slots, the likelihood of treatment is lower and vice versa. This marks the identifying variation that we use to estimate the returns to language training.

From this variation we construct a continuous language training availability (LTA) instrument that exploits regional differences in available slots. Figure 1depicts the absolute number of supplied slots per square kilometer that concentrate in metropolitan areas with high population density. Dark areas indicate counties where the supply of open slots is higher and we show that this makes participation in language training considerably more likely to the individual immigrant.

The identifying variation inherent to this measure is a strong positive correlation between the county-level share of supplied slots (figure 1) and actual language course participation as reported in our data. One additional open slot per square kilometer within a county (43% increase, based on 2.3 slots on average) is associated to an 8% higher likelihood for the individual immigrant to participate in language training, conditional on spatial background variables that include important determinants of language course demand such as foreign share and population density.[16] Hence the probability to participate in language training is an upward-sloping function of language course supply. We use this link to isolate the exogenous part from otherwise endogenous participation decisions.

An important explanation for differences in the regional coverage in language training is the absence of a regulating authority at the municipality level. As a consequence, the supply of language training does not directly follow economic determinants that may reliably signal scarcity within a region (Federal Office for Migration and Refugees, 2011). Although the regional coverage by language courses primarily depends on observable characteristics such as the share of the foreign population and unemployment rates of

[16]The relationship is based on the correlation coefficient of 0.031 that is obtained from the first stage regression of the 2SLS and MTE framework (table 3) and significantly differs from zero. It indicates that one additional slot per square kilometer within a county is associated to an increase in the likelihood of language training participation of 3.1 percentage points. Relating this quantity to the sample language training participation rate of 38.2% implies an 8% increase. Dropping the full set of covariates in a regression only including spatial variables at the county level (population density, disposable income, foreign share, unemployment rate of foreigners) yields a correlation coefficient of 0.026 that also significantly differs from zero and implies a 6.8% increase.

foreigners at the municipality level, the probability to participate is largely exogenous to the individual immigrant and her labor market outcomes.

Considerable heterogeneity in the supply of language training across regions is not only due to the lack of a regional governing entity. It is also influenced by the circumstance that the supply of language training depends on a large set of factors such as overall infrastructure, the availability of qualified teachers, and the willingness of providers to organize integration courses. This complexity makes the spatial distribution of language training difficult to foresee and thus exogenous to the individual immigrant. The probability of being treated by language training is thus exogenous to labor market outcomes and makes the corresponding variation of excess demand an appropriate instrument for endogenous language skills.

One potential challenge to this approach is the financial situation at the county level that may correlate to language course provision and thus participation rates. Higher financial endowment of a county may come along with more open slots for language training and this may attract more motivated immigrants and thus induce selection at the level of variation of the instrument. To demonstrate that this does not change our results, we use information on the degree of indebtedness for each county as a direct proxy for the financial power to supply language training (see section 5.4).

Another challenge for the exogeneity of the treatment by language training may be selective moving behavior subsequent to immigration. Our primary data include information on individual mobility within Germany such that we can account for selective movements towards counties that provide higher capacities and more open slots for language training. A considerable share of 89% of the immigrants in our sample have not moved after one year and 80% have not moved after three years. Even after a considerable amount of time, the majority of immigrants still lives in the same county of first appearance upon arrival. Despite rather small shares of movers, we obviate concerns about the exogeneity of the instrument to the individual immigrants' labor market outcomes by adding information on movements across regions within the estimation framework (see section 5.4).

3 Empirical Strategy

The empirical framework of estimating marginal treatment effects (MTE) mainly builds on the discussions of Heckman and Vytlacil (2005) and Carneiro et al. (2011) and corresponding derivations therein. Our starting point is to define two potential outcomes Y^1 and Y^0 with and without treatment respectively (presupposing the index i for the individual). The observed outcome Y is equal to Y^1 if the immigrant receives language training and equal to Y^0 if she does not receive the language training. As implied by the institutional setting, the participation is not fully exogenous to the individual immigrant, or even voluntary for some of them, and thus a treatment dummy D would be endogenous in a simple linear regression.

To make the voluntary treatment explicit, we use a latent index model for the two potential outcomes

$$Y^1 = X'\beta_1 + U_1 \tag{1}$$

$$Y^0 = X'\beta_0 + U_0 \tag{2}$$

$$D^* = Z'\delta - V, \text{ where } D = \mathbb{1}[D^* \geq 0] = \mathbb{1}[Z'\delta \geq V] \tag{3}$$

where X denotes a vector of observables and U_1, U_0 are unobservables affecting each potential outcome respectively. The variable D^* is the latent desire to participate in language training that is explained by observable variables Z and unobservables V. Z includes the instrument that satisfies an exclusion restriction[17] and all observable variables that are part of X. Only if D^* exceeds a specific threshold (assumed to be zero, for simplicity), the immigrant will participate in language training. The fact that U_1, U_0, and V are probably correlated but also unobservable is the fundamental challenge to the analysis here. Although we can observe the outcome Y we cannot observe both Y^1 and

[17]That the instrument is not part of the underlying causal relationship is arguably a plausible assumption. In particular, regional variation in excess demand for language training, as reflected by the LTA instrument, should not be part of an equation that aims at explaining labor market outcomes. It only affects these outcomes indirectly by isolating the exogenous part from otherwise endogenous language skills that do affect labor market performance.

Y^0 for one individual at the same time ($Y = DY^1 + (1 - D)Y^0$).

The estimation framework exploits that individuals who react to a shift in the instrument, being pushed into treatment at the margin of indifference, also reveal their rank in the distribution of unobservables. Even though the unobservables are unknown by their nature, they are fixed by the propensity score that is based on the observables. This enables us to examine the outcome for those who are pushed into treatment by the instrument at any quantile of the distribution of U_D. That instrument-induced changes in language training participation are identifiable across the distribution of U_D illustrates how the MTE can be interpreted as a willingness-to-pay parameter: an immigrant will evaluate the costs (e.g. foregone leisure or income from a job) and the benefits from acquiring language skills at each point of indifference across the distribution of unobservables. At this point, we have that $Z'\delta \geq V$.

The MTE is then defined as the treatment effect for an individual with observable characteristics $X = x$ who is just indifferent to receiving the treatment when having a propensity score $P(X, Z)$ that is equal to the unobserved resistance to treatment U_D. Formally, it is defined as the derivative of the conditional expectation of the outcome with respect to the propensity score

$$\text{MTE}(X = x, U_D = p) = \frac{\partial E(Y|X = x, P(Z) = p)}{\partial p} \tag{4}$$

and thus indicates a change in the outcome relative to a marginal change of the propensity score.

The MTE recovers treatment effect heterogeneity along the distribution of U_D. Essential heterogeneity arises when idiosyncratic responses to the treatment differ across individuals with at least partial knowledge about their returns (Heckman et al., 2006). This structural property of the MTE implies that the marginal immigrant needs increasingly more compensation to participate in language training when the observed propensity of participation $P(Z)$ decreases. If immigrants already react at values of the instrument that imply a low observed treatment probability, i.e. when excess demand is high and open slots in language courses are rare, then they must have low unobserved latent costs

V. This would be the case, for example, if the immigrant has high unobserved motivation or ability based on partial knowledge of her own expected idiosyncratic gains. In the notation from above, participation only requires that the values of the unobserved latent costs V are marginally lower than the observed part $Z'\delta$. Thus, choosing to participate in language training becomes increasingly unlikely as $P(Z)$ decreases because it requires increasingly more unobserved motivation to make V sufficiently low to be willing to participate.

Under essential heterogeneity, the average treatment effect (ATE), the average treatment effect on the treated (ATT), the average treatment effect on the untreated (ATU) and the local average treatment effect (LATE) differ from each other. These common treatment parameters are simply weighted averages of the MTE which is a more structural parameter that reveals all local switching effects by intrinsic willingness to participate in language training. Aggregating the MTE into the different parameters is not only interesting in its own right but also allows for summaries that are consistent with large parts of the treatment effect literature.

To obtain an estimable expression, we follow the derivations of Heckman and Vytlacil (2007) and plug in the counterfactual outcomes from (1) and (2) into the conditional expectation $E(Y|X = x, P(Z) = p)$. Rearranging and imposing an exclusion restriction of p on Y yields the expression

$$\begin{aligned} E(Y|X = x, P(Z) = p) &= X\beta_0 + X(\beta_1 - \beta_0) \times p + E(U_1 - U_0|D = 1, X) \times p \quad (5) \\ &= X\beta_0 + X(\beta_1 - \beta_0) \times p + K(p) \end{aligned}$$

where $K(p)$ is some flexible function of the propensity score. Two properties of this expression are important. First, the interaction of X and p identifies $(\beta_1 - \beta_0)$ which is the intercept of the MTE-curve, showing that the MTE is only shifted by the observables while the shape of the MTE does not depend on X. This is implied by the full independence assumption $(X, Z) \perp (U_0, U_1, V)$, which means that X is exogenous and that the way in

which U_1 and U_0 depend on V (i.e. the shape of the MTE-curve) is independent on X.[18] Second, the function $K(p)$ does not depend on X, reflecting the assumption that the slope of the MTE-curve does not depend on the observables. This means that, by conditioning on X in a parametric linear way, we only need unconditional full common support of the propensity score across all values of $X = x$ (see Cornelissen et al., 2016, for a discussion).

Based on equation (4), an estimable expression of the MTE is given by

$$\text{MTE}(X = x, U_D = p) = \frac{\partial E(Y|X = x, P(Z) = p)}{\partial p} = X(\beta_1 - \beta_0) + \frac{\partial K(p)}{\partial p} \tag{6}$$

To take this approach to the data, we follow the recent applications of Brinch et al. (2017); Cornelissen et al. (2018); Kamhöfer et al. (2019) that do not aim at causally interpreting the separate sources of effect heterogeneity. In this case, an instrument that satisfies an exclusion restriction (i.e. is not part of the underlying causal relationship of the outcome equation) is sufficient to identify the level and the curvature of the MTE curve. The first step is to estimate the participation decision D in (3) as a Probit model. From this first stage selection equation, we obtain estimates of the propensity score $\hat{p}$ that permits us to estimate the parameters of the outcome equation

$$Y = X\beta_0 + X(\beta_1 - \beta_0)\hat{p} + K(\hat{p}) + I\alpha + O\gamma + S\delta + F\lambda + \varepsilon \tag{7}$$

where $K(\hat{p})$ is a polynomial in $\hat{p}$ of degree k. Throughout, we use a linear specification of the propensity score to model the relationship between the outcome and the propensity to participate in language training.[19] Further components of the outcome equation are fixed effects of the immigration year I (α), country of origin O (γ), occupational Sector S (δ) and region R (federal state, λ). In our baseline specification, we estimate a fully parametric version of (7) and then contrast these estimates to more flexible semi-parametric

[18] The full independence assumption is stronger than only assuming conditional independence $Z \perp (U_0, U_1, V)|X$ that would be required for a causal interpretation of IV and estimating the MTE non-parametrically.

[19] Previous applications have also used higher order polynomials (see for example Cornelissen et al., 2018). In our application, the relationship between the outcome and the treatment probability is essentially linear so that modeling higher order polynomials does not improve estimation.

estimates based on techniques developed by Robinson (1988).[20]

4 Data

4.1 Data and Sample Restrictions

The empirical framework is based on administrative employment records (Integrated Employment Biographies, IEB) that are linked to survey data from the German Socio-Economic Panel (SOEP) at the individual level. The combined data source, the IAB-SOEP-MIG-ADIAB migration sample[21], is restricted to immigrant respondents that are representative for Germany.

The key advantage of this data set is that it combines individual employment histories from social security records to individual surveys. Due to the emphasis on the immigrant population, the data set is smaller in size compared to only using employment records from social security data. However, it allows for detailed descriptions of the immigrant population and, necessary for our analysis, includes information on whether immigrants have received language training or not. It also includes details on the pre-migration employment history and a rich set of variables on initial conditions upon arrival after immigration. Based on this information we can control for pre-migration employment and language skills at arrival to explicitly capture heterogeneity that drives both language skills and subsequent labor market performance.[22]

Our final sample includes 1,570 immigrants for whom we observe both outcomes employment status and wages. We focus on first generation migrants to make the sample homogeneous in terms of language skill acquisition after arrival. We further restrict the sample to immigrants of age 16 and above, because children grasp language capabilities

[20]Both parametric and semi-parametric estimates are obtained using the Stata command *margte* (see Brave and Walstrum, 2014) and the more flexible extension *mtefe* (see Andresen, 2018).

[21]For a detailed overview on the data source including a description of the content, the sampling design and methodology, see Brücker et al. (2014); Trübswetter and Fendel (2016). Recent papers on the economics of migration using this data source are Dustmann et al. (2016), Battisti et al. (2018), Brücker et al. (2018) and Riphahn and Saif (2018).

[22]If these characteristics remain unobserved, estimates of labor market outcomes can suffer from down- or upward bias, depending on the type of selection (Willis and Rosen, 1979; Borjas, 1994; Chiswick and Miller, 1995).

much more quickly and they acquire these skills through channels other than formalized language training (e.g. in kindergarten or school). Finally, the sample of immigrants under study is characterized by a relatively high share of non-EU immigrants, reflecting more recent immigration inflows during the 1990s and 2000s[23], making the paper arguably more relevant for current migration policy.

4.2 Control Variables and Language Skills

Table 1 provides an overview on the sample of immigrants that we use throughout. Descriptive statistics are reported separately by treatment status, showing that language course participants are slightly older and include a higher share of social benefit recipients. The share of EU citizens is much lower among participants while the share of ethnic Germans, predominantly from Russia, is larger. Despite non-significant differences in education, it is worthwhile to note that participants tend to have medium education more often (52%) than non-participants (47%). Graphical evidence also shows that immigrants with higher education (medium and high) have a higher propensity to participate (figure 4, panel e). The estimation results later-on further support this view, leading us to conclude that the pool of participants consists of two divergent groups: a negatively selected one with poor language skills and bad integration with mandatory assignment and a positively selected one with higher education.

Table 2 (Panel B) reports details on the distribution of language skills, indicating that a good command of German is lower among participants both at arrival and currently. Good command of German is an indicator that is equal to one if the command of German is good or very good in at least one out of the three skill categories speaking, writing and reading. These categories range between 1 (very good) and 5 (very poor). The fact that a good command of German is significantly less prevalent in the group of language course participants is consistent with the assignment rules to language training.

[23]Hence the immigrant sample we use is less dominated by guest workers from Turkey or Yugoslavia, who immigrated predominantly throughout the 1960s and 1970s.

4.3 Dependent Variables

The outcomes employment and wages are depicted in table 2 (Panel A), showing their distribution for different follow-up periods after the arrival of immigrants. Throughout the analysis, employment is defined as regular employment that is subject to social security contributions (*sozialversicherungspflichtige Beschäftigung*), which has a share of 41% in the baseline follow-up period (8 years after arrival). Evidently, regular employment is lower among those who receive language training (37.3%) compared to the non-participants (44.1%). Wages are defined as daily wages, averaging to 33.4 EUR in the baseline follow-up period. Consistent with employment rates, average wages are lower among participants in language courses (31.6 EUR) compared to those who do not receive the treatment (35.5 EUR). This is, once again, in line with the rules of the GLI that primarily assigns immigrants to language training that have a stronger need for integration.

4.4 Instrument

The expansion of language training creates exogenous variation across German counties (see section 2). From this variation in the supply of open slots across counties, we construct a language training availability instrument Z_{jt} that refers to the individual immigrant living in county j at time t. The instrument is defined as

$$Z_{jt} = \frac{\#\text{supplied slots}_{jt}}{km^2} \tag{8}$$

based on county level information on the number of open slots in integration courses.[24] The number of supplied slots is conditional on a set of spatial background variables that include important determinants of language course demand at the local level, each referring to county j at time t. These variables are the foreign share $\left(\frac{\#foreigners_{jt}}{\#inhabitants_{jt}}\right)$, the unem-

[24]The annual reports from the Federal Office for Migration and Refugees (BAMF) are supplemented by spreadsheets that include the actual number of participants for each of the 401 German counties (see Federal Office for Migration and Refugees, 2013, 2014).

ployment rate of foreigners, the overall population density and an indicator for wealth (disposable income per inhabitant). Including spatial background variables accounts for the fact that they indirectly determine the likelihood of the individual immigrant to obtain an open slot in a language course. All background variables are used to estimate the first stage selection equation to determine the propensity of participation in language training. They all satisfy an exclusion restriction and are thus not part of the outcome equation.

Larger values of Z_{jt} indicate more supplied slots and thus higher availability of language training, depicted as dark areas in figure 1. Further descriptive statistics on Z_{jt} and spatial background variables that are implicitly included in the instrument are presented in table 2 (Panel C). The average language course supply among treated immigrants (with language training) amounts to 2.7 slots per square kilometer and to only 1.4 participants among non-treated immigrants (without language training).[25]

4.5 Limitations

Estimating the returns to language training involves a few caveats from data limitations. First of all, since we use administrative employment records, the sample is restricted to immigrants with an individual employment history after arrival. Immigrants who never registered at the employment agency and thus never became part of the labor force (either through employment or unemployment in combination to active job search) do not appear in the employment records. Thus, immigrants without any labor force attachment are not part of the analysis. However, since these immigrants are likely to be the ones with the highest resistance to language training, such as illiterate family members who follow their predecessors, we argue that our MTE estimates would extrapolate to this population towards the extreme margin of non-participation.

Second, self-reported language skills involve measurement error to the extent that each respondent has her own scale of evaluation. This may introduce downward bias in the estimated effect that is likely to overcompensate the upward bias from unobserved ability

[25]The Germany-wide mean supply is 2.3 slots per square kilometer, averaged over all counties. One additional slot thus corresponds to an increase of 43%.

(see Dustmann and Glitz, 2011, for a discussion). Since we use panel data with repeated information on each individual we are able to account for this type of measurement error by including individual-specific fixed effects, assuming that individual scaling of language proficiency remains constant over time.

Third, regional variation in language training supply is only available for the years 2013 and 2014. Due to this limitation, our instrument relies on the assumption that variation across counties is stable over time. It is unlikely, however, that differences in language training supply changed considerably across counties in the period under study. First, the budgets were largely constant over time and second the training capacities only change slowly since language teachers are not flexibly available in the short term.

Finally, selection induced from the unobserved intention to stay in the host country permanently is likely to correlate with language training choices. However, our approach of estimating marginal treatment effects is a new way of dealing exactly with this problem. It reveals marginal gains from language training participation for individuals who are marginally shifted into treatment by a marginal change of the propensity score as a function of the instrument. This permits us to examine how individuals who do not participate in language training would benefit from treatment.

5 Results

5.1 OLS and 2SLS

We start presenting OLS and 2SLS results as a benchmark and then estimate the returns to language training in the MTE framework. The first stage results of the 2SLS estimation (table 3) are based on identical samples and are thus similar for both outcomes employment and wages. The main indication from this is that language course availability coincides to a higher treatment probability. This correlation expresses the identifying variation of the instrument: one additional slot per square kilometer (43% increase) is associated to a 3.1 percentage point increase in the participation probability. Evaluating this quantity at the average participation rate of 38.2% in the sample, the individual like-

lihood of participating in a language course increases by 8% whenever the supply increases by one slot at the local level. In addition to the considerable correlation between the LTA instrument and endogenous participation, the instrument is also individually significant (t-statistic: 2.8, F-statistic: 22.5).

Naive OLS estimates of endogenous language training on both outcomes employment and wages reflect negative selection into language training (table 3, column (1) and (2)). The 2SLS estimates in column (3) and (4) report positive coefficients for employment and wages by accounting for negative selection. However, none of these estimates (LATEs) significantly differ from zero. Although this suggests zero average returns from language training, we will come to different conclusions when looking at the marginal immigrant who is just indifferent regarding participation in language training.

The bottom part of table 3 also reports a strong and significant correlation between language skills (to be distinguished from training) and subsequent employment and wages. It indicates that language skills upon arrival coincide to much higher employment rates and wages: strong German language skills (Good command of German at arrival = 1) are associated to employment rates that are 15 percentage points higher and wages that are 63% – 70% higher later-on, compared to immigrants with poor German language skills (Good command of German at arrival = 0, see section 4.2 and table 2 for the definition of the language indicator). Departing from this considerable correlation, we now estimate marginal returns to language training based on the LTA instrument, exploiting regional variation in language course availability.

5.2 Marginal Treatment Effects

The density distribution of the propensity score is shown separately by treatment status in panel (a) of figure 2. The two distributions are obtained from a Probit regression of the first stage selection equation and the common support ranges from 0.15 to 0.91, thus covering a large part of the distribution. Panel (b) illustrates the distribution of the propensity score that is solely predicted by the instrument Z. This indicates the support for the identifying variation ranges from 0.26 to 0.75. Although the propensity

score solely predicted by Z (identifying variation) does not cover the entire range of the propensity score predicted from the full model (based on X and Z), we can still identify marginal treatment effects. The only limitation regarding this is that we have to rely on extrapolations to regions of the propensity score where we cannot draw on identifying variation. For this reason, our baseline estimates are based on fully parametric estimation.

Figure 3 shows the marginal treatment effects of language training on the probability of being employed eight years after arrival (panel a). The MTE curve slopes downward[26], indicating that immigrants with the highest desire to participate in language training are those who benefit the most. At the maximum, we estimate a significant 6 percentage point increase in employment eight years after arrival, referring to a 15% increase when relating to the mean employment rate of 41% in the sample. These gains vanish with increasing unobserved resistance to participation.[27] Regarding wages (figure 3, panel b), we estimate significant wage gains of up to 13% for those immigrants who are most ready to take the treatment. Similar to the results on employment, the effects vanish for those with the highest resistance to participation. In conclusion, the results reflect selection on gains similarly for employment and wages: immigrants with higher expected gains are more likely to select into language training.[28]

5.3 Selection on Observables and Language Training

The first stage results yield a meaningful description of participation in language training that can either occur by the assignment through the law of integration or by self-selection. The selection equation (table 4 and 5) shows that EU-immigrants are much less likely to participate in language training than non-EU immigrants (13 percentage point difference) and indicates that immigrants with German language skills at arrival are less likely to

[26]The test of heterogeneity (table 4) shows that the decline of the MTE curve is significant.

[27]The effects even reverse to negative values, implying that the employment probability reduces through participation. More flexible semi-parametric estimates, however, indicate that the employment effects only become insignificant with declining treatment probability (see section 5.4). Negative employment effects could only be rationalized for immigrants with poor employment who do not find a job despite facing the opportunity costs of the intense 600 hour training-load.

[28]Although we take an ex-post perspective in the empirical framework where the returns are realized ones, we assume that individuals form expectations about their returns based on which they make decisions about participation.

receive language training compared to immigrants without any knowledge of the German language (12 percentage point difference).[29] Both the origin (EU vs. non-EU) and language capabilities at arrival are important determinants on whether an immigrant is assigned to language training or not and these first stage estimates are consistent to graphical evidence on the group-specific distribution of the propensity score (figure 4, panel b and d).

We also document a positive relationship between age at arrival and the propensity to participate. Immigrating one age-year later is associated with a 0.5 percentage point increase of the probability of receiving language training. Similarly, when splitting the sample into young immigrants (below 25 at arrival) and older immigrants (25 or above at arrival), the distribution of the propensity score makes this finding graphically explicit (figure 4, panel f). Higher participation rates among older immigrants are consistent with the idea of age-at-arrival instruments for identifying the effects of language skills on labor market success (see e.g. Bleakley and Chin, 2004; Yao and van Ours, 2015). Young immigrants grasp languages more quickly and less burdensome, thus making language training obsolete for the them in comparison to older immigrants who are much more likely to participate.

Social benefit receipt at arrival does not significantly correlate to participation on the first stage. This is noteworthy because receiving social benefits is a criterion for compulsory assignment due to incomplete enforcement of actual participation by the responsible authorities. As expected, social benefit receipt is negatively associated to employment (table 4) but those immigrants who participate in language training benefit substantially in terms of a higher employment probability.[30] This finding strongly suggests that local authorities should ensure more thoroughly that immigrants who receive social benefits do participate in language training given that their benefits are so high.

[29]Lower participation propensities are also reported for low-skilled immigrants (less communicative tasks, compared to medium-skilled), family members (low labor force attachment, compared to other immigrants e.g. with job offer), asylum seekers (uncertain residency status makes investments into host country specific human capital less attractive, compared to other immigrants e.g. with job offer) or job searchers (negative selection (search is required by employment agency), compared to other immigrants e.g. with job offer).

[30]This is indicated by the positive and significant coefficient on the interaction between the propensity score and social benefit receipt (table 4, Panel C, outcome equation).

Finally, immigrants with low education are associated with a lower likelihood to participate in language training than those with medium education (reference group). This is consistent with the distribution of the propensity score, indicating that immigrants with higher education (medium and high) have more probability mass at higher values of the propensity score (figure 4, panel e). This pattern has strong implications for the central result that individuals with higher gains are more likely to select into language training. Strikingly, it reveals that the participants are composed of two different groups of immigrants. The first group includes persons who are assigned to language training by law and who tend to be a negatively selected pool with particular need for integration. The second group includes immigrants with high levels of unobserved motivation and talent that have at least partial knowledge about their high expected gains from language training. These individuals also have made higher investments in observed education which strongly supports the view that language capital complements other types of human capital such as education. Consequently, the gains from language training are much higher for those with higher education and more job-specific skills. A strong policy conclusion from this is that pushing more reluctant immigrants into treatment should be complemented by training on job-specific skills. Otherwise, immigrants may not be able to transmit their language skills to the labor market.

5.4 Sensitivity Analysis

The precision of the estimates increases along the lines of the follow-up period. To show how the returns from language training materialize more explicitly as time since migration elapses, figures 5 (employment) and 6 (wages) depict the MTE curves for different follow-up periods. Moving from less precisely measured MTE only five years after arrival, the estimates are increasingly more pronounced after eight years (the baseline, see above) and 10 years. The pattern is very similar for employment and wages and consistently indicates how immigrants integrate over time, suggesting that the labor market effects of language training do have a pronounced long-run component. This is consistent with the findings in Card et al. (2018), who show that the average effects of (more broadly defined) active

labor market programs are small in the short-run and become larger after 2 - 3 years after completion.

To demonstrate that our results are robust against potential selection bias from county-specific financial power, we present further MTE estimates from regressions that include public debt per capita at the county level. The MTE-curve in figure 7 shows that the baseline estimates change only little when indebtedness is included as a proxy for financial means in language training supply. The MTEs are slightly smaller in magnitude for employment outcomes (panel a) at the extreme ends but yet precisely estimated. Regarding wage returns (panel b), the estimates are slightly less precisely estimated for high resistance individuals but are otherwise very similar. Although the financial situation at the county level may correlate to language course provision and thus participation rates this seems not to affect our results. If one were to argue that more motivated immigrants select into counties that provide better opportunities for language training (more open slots), then we can rule out that this affects our results along the lines of financial means.

We also test to what extent selective movements of immigrants across regions may confound the analysis. The issue is particularly important because selective movements of more motivated immigrants could make the treatment endogenous. If an immigrant moves to a more prosperous county, this may not only involve a higher language course capacity and thus a higher treatment probability, but it would also imply better employment and wage opportunities. Figure 8 depicts MTE-curves when estimating the baseline specification but accounting for selective movements. In particular, the specifications in panel (a) and (b) include an indicator that is equal to one if the immigrant has moved across counties within the first three years after arrival and zero otherwise. Although the baseline results on employment are robust and change only little when accounting for the potential of selective movements, the wage outcome is estimated with less precision. The relevant source of selection becomes more evident in the panels (c) and (d) of figure 8, showing MTE-curves that are estimated on a sample that is restricted to non-movers over the entire follow-up period of eight years (N:679). When ruling out potentially selective movements, the patterns regarding both employment and wages become clearer and are much more precisely estimated compared to the baseline specification. In summary, selec-

tive movements within Germany seem to play a role. Based on information about moving behavior, however, we are able to show that the principal results are robust against this source of selection and that the conclusions remain unchanged.

Finally, we show semi-parametric estimates of the MTE that are depicted in figure 9. To make the identification of the semi-parametric MTE transparent, we only plot the region of the MTE curve where common support of the propensity score is given and thus do not extrapolate. The negatively sloping MTE-curves recovers a pattern that is largely similar to previous parametric estimates, identifying strong positive employment effects and positive but less precisely estimated wage effects among immigrants with a high desire for participation in language training. Interestingly, the semi-parametric estimates indicate that the positive returns vanish for immigrants with a high unobserved resistance against participation. For high values of U_D, the confidence bands are large and suggest that the MTE does not differ from zero. Negative returns for highly reluctant individuals (high values of U_D), as suggested by fully parametric estimates, are generated from extrapolations to the extreme margins of the MTE curve where our data do not provide common support of the propensity score. Since the semi-parametric estimates are more flexible than the fully parametric estimates, we conclude that those immigrants with high resistance to participation have zero returns from language training.

6 Treatment Parameters and Cost-Effectiveness

Table 6 reports common treatment parameters that are estimated using the MTE and corresponding weights, following the derivations and explanations of Heckman et al. (2006). We calculate the ATE, the ATT, the ATU, and the LATE, according to the definition of each parameter, thus summarizing the MTE from different perspectives based on the weights depicted in figure 10.

The parameters in table 6 are informative because they aggregate the MTE for different sub-populations and thereby provide a meaningful summary of effect heterogeneity. It is noteworthy, first of all, that the population ATE does not significantly differ from zero regarding both employment and wages. While this result indicates that there are

no language training effects on average when looking at the entire population, it does not rule out that there may be effects on specific subgroups. And indeed, the ATT reveals that there are considerable positive effects for those immigrants who participate in language training. Receiving language training increases the employment probability by 0.4 percentage points and wages by 2.5% on average. Among immigrants without language training, the estimates (ATU) are small, negative and insignificant and hence we conclude that non-participants do not benefit from language training. This result is also consistent with semi-parametric estimates of the MTE from above.

We now relate the marginal returns from language training, as reported above, to the marginal costs for each additional participant. To obtain a measure of marginal costs, we use the hourly rate of 2.05 EUR for each participant of integration courses as used by the German government for budget calculations (Federal Government, 2007, p. 66). For a 600 hour course, the hourly rate totals to 1,230 EUR per immigrant.

The cost-benefit analysis is informative on how benefits and costs of language training relate to each other for different sub-populations. An extreme case would be looking at low-resistance individuals with a high desire to participate. Those individuals benefit the most from language training, with baseline estimates suggesting wage gains of up to 13%. These wage gains amount to 9.8 EUR per day when evaluated at the average daily wage rate of 75.7 EUR in the sample of employed immigrants in 2014.[31] For immigrants with a particularly high desire to participate, the annual wage gains would thus accumulate to 2,558 EUR[32] on average, thus doubling the monetary costs of language training within only one year.

These particularly large gains are obtained when looking at the maximum gains for those immigrants with the highest motivation in terms of the observed propensity score and the unobserved resistance (U_D). In the sub-population of immigrants with language training, the wage gains are a significant 2.5% (ATT) and thus amount to 1.9 EUR per day (again, evaluated at the mean daily wage rate of 75.7 EUR). These wage gains

[31] In particular, the mean daily wage is defined for those persons who are in employment that is subject to social security contributions. We cannot distinguish by hours of work and thus the mean includes both full- and part-time work.

[32] In this example, we have 9.84 EUR x 5 days x 52 weeks = 2,558 EUR per year.

accumulate to 492 EUR per year[33], such that it still takes less than three years in the job to outbalance the training costs per additional immigrant of 1,230 EUR. This result is highly policy relevant because it indicates that the benefit-cost ratio of the program is pushed well above one within a fairly short period, even disregarding other non-monetary effects that language has on successful integration.

7 Conclusions

In this paper, we estimate the returns to language training of immigrants under essential heterogeneity. For this purpose we adopt the framework of marginal treatment effects, an approach that has been introduced by Björklund and Moffitt (1987) and advanced by Heckman and Vytlacil (1999, 2005, 2007). The empirical analysis is based on a unique data source linking integrated employment biographies of immigrants to survey data from the German Socio-Economic Panel. The outcomes examined are employment and daily wages. Both of these measures show a strong and positive correlation to language skills, indicating that a good command of the host countries' language (German) is associated to higher employment rates and higher wages. Based on this correlation, we estimate the returns to language training based on an instrument that exploits regional variation in language training availability.

The main results of this paper are summarized as follows: although the average returns from language training of immigrants are moderate (the ATT implies a 2.5% increase of wages in the long-run) or even negligible (the LATE does not significantly differ from zero), these estimates mask substantial heterogeneity. To uncover this heterogeneity, we estimate the effects for the marginal individual at each point of indifference for participation in language training. We document considerable positive long-run returns from language training for those immigrants with the highest desire for language training. Eight years after arrival, immigrants who are most ready to take the treatment show a significant 15%

[33]Wage gains accumulate to 5,000 EUR in a 10-year follow-up period and for those with a high desire to participate these gains would even sum to 25,000 EUR. Sarvimäki and Hämäläinen (2016) extrapolate their LATE of restructuring active labor market policies towards more immigrant-specific training (including language training) to cumulative gains in gross earnings of 21,000 EUR within 10 years, thus ranging somewhere between our estimates.

increase in employment and wage gains of up to 13%. The effects vanish for individuals with a high resistance to participation. This reflects selection on gains where immigrants with a high desire for language training are those who benefit the most.

Studying heterogeneity in the returns to language training allows for particularly informative policy conclusions. The expansion of the program would be worthwhile due to the large long-run cumulative wage gains that raise lifetime earnings and thus governmental tax revenues. Our estimates imply that the costs per additional immigrant in language training are compensated after less than three years in employment, shifting the benefit-cost ratio above one within a fairly short period. Pushing immigrants with a weak desire for language training into participation should, however, be complemented by job-specific training to raise their returns. Immigrants who prefer non-participation may not be able to adequately transmit their language skills to the labor market because they also tend to have made little pre-migration investments in education and job-specific skills. We conclude that improving the immigrants' language capabilities is important for economic integration. This is highly topical because communicative non-routine tasks and corresponding skill requirements become increasingly more valuable through technological change.

Some limitations should be kept in mind when interpreting the results of this paper. First, language acquisition may involve further positive non-monetary effects regarding successful integration that we are not able to capture. Second, we only measure the lower bound of the true effect in the presence of close substitutes. Although our data report language training information for a whole range of different programs, it is likely that immigrants receive language training outside these courses, for example through social interaction with natives. In this case, our estimates must be interpreted as the average effect for compliers relative to a mix of relevant alternatives.[34]

[34]Kline and Walters (2016) point out that, in the presence of close substitutes, typical IV estimates only yield the average effect of the program for compliers relative to their own counterfactual training choice.

References

Andresen, M. E. (2018). Exploring Marginal Treatment Effects: Flexible Estimation Using Stata. *The Stata Journal 18*(1), 118–158.

Aslund, O. and M. Engdahl (2018). The Value of Earning for Learning: Performance Bonuses in Immigrant Language Training. *Economics of Education Review 62*, 192–204.

Battisti, M., G. Peri, and A. Romiti (2018). Dynamic Effects of Co-ethnic Networks on Immigrants' Economic Success. *NBER Working Paper 22389*.

Björklund, A. and R. Moffitt (1987). The Estimation of Wage Gains and Welfare Gains in Self-Selection Models. *The Review of Economics and Statistics 69*(1), 42–49.

Bleakley, H. and A. Chin (2004). Language Skills and Earnings: Evidence From Childhood Immigrants. *Review of Economics and Statistics 86*(2), 481–496.

Bleakley, H. and A. Chin (2010). Age at Arrival, English Proficiency, and Social Assimilation Among US Immigrants. *American Economic Journal: Applied Economics 2*(1), 165–192.

Borjas, G. (1994). The Economics of Immigration. *Journal of Economic Literature 32*(4), 1667–1717.

Brave, S. and T. Walstrum (2014). Estimating Marginal Treatment Effects Using Parametric and Semiparametric Methods. *The Stata Journal 14*(1), 191–217.

Brücker, H., A. Glitz, A. Lerche, and A. Romiti (2018). Occupational Recognition and Immigrant Labor Market Outcomes. *IZA Discussion Paper No. 12030*.

Brücker, H., M. Kroh, S. Bartsch, J. Goebel, S. Kühne, E. Liebau, P. Trübswetter, I. Tucci, and J. Schupp (2014). The New IAB-SOEP Migration Sample: An Introduction into the Methodology and the Contents. *SOEP Survey Papers Series C*(216).

Brinch, C. N., M. Mogstad, and M. Wiswall (2017). Beyond LATE with a Discrete Instrument. *Journal of Political Economy 125*(4), 985–1039.

Card, D., J. Kluve, and A. Weber (2018). What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. *Journal of the European Economic Association 16*(3), 894–931.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011). Estimating Marginal Returns to Education. *American Economic Review 101*, 2754–2781.

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 (2002). Immigrant Earnings: Language Skills, Linguistic Concentrations and the Business Cycle. *Journal of Population Economics 15*(1), 31–57.

Chiswick, B. R. and P. W. Miller (2003). The Complementarity of Language and Other Human Capital: Immigrant Earnings in Canada. *Economics of Education Review 22*(5), 469 – 480.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2018). Who Benefits from Universal Child Care? Estimating Marginal Returns to Early Child Care Attendance. *Journal of Political Economy 126*(6), 2356–2409.

Cornelissen, T., C. Dustmann, A. Raute, and U. Schönberg (2016). From LATE to MTE: Alternative Methods for the Evaluation of Policy Interventions. *Labour Economics 41*, 47–60.

Dustmann, C. and A. Glitz (2011). Migration and Education. In: E. A. Hanushek, S. Machin and L. Woessmann (Eds.). *Handbook of the Economics of Education Vol. 4, Chapter 4, 327-439, North-Holland*.

Dustmann, C., A. Glitz, U. Schönberg, and H. Brücker (2016). Referral-based Job Search Networks. *The Review of Economic Studies 83*(2), 514–546.

Dustmann, C. and A. van Soest (2001). Language Fluency and Earnings: Estimation with Misclassified Language Indicators. *The Review of Economics and Statistics 83*(4), 663–674.

Dustmann, C. and A. van Soest (2002). Language and the Earnings of Immigrants. *Industrial and Labor Relations Review 55*(3), 473–492.

Federal Government (2007). Erfahrungsbericht der Bundesregierung an den Deutschen Bundestag zu Durchführung und Finanzierung der Integrationskurse nach § 43 Abs. 5 Aufenthaltsgesetz. *Field Report of the Federal Government to the Federal Parliament.*

Federal Office for Migration and Refugees (2008). Das Integrationspanel. Ergebnisse zur Integration von Teilnehmern zu Beginn ihres Integrationskurses. *Bundesamt für Migration und Flüchtlinge (BAMF) Working Paper 19.*

Federal Office for Migration and Refugees (2011). Das Integrationspanel. Ergebnisse einer Längsschnittstudie zur Wirksamkeit und Nachhaltigkeit von Integrationskursen. *Bundesamt für Migration und Flüchtlinge (BAMF) Forschungsbericht 11.*

Federal Office for Migration and Refugees (2013). Integrationskursgeschäftsstatistik (Integration Course Statistics) 2013. *Bundesamt für Migration und Flüchtlinge (BAMF).*

Federal Office for Migration and Refugees (2014). Integrationskursgeschäftsstatistik (Integration Course Statistics) 2014. *Bundesamt für Migration und Flüchtlinge (BAMF).*

Federal Office for Migration and Refugees (2017). Bericht zur Integrationskursgeschäftsstatistik (Report on Integration Course Statistics). *Bundesamt für Migration und Flüchtlinge (BAMF).*

Felfe, C. and R. Lalive (2018). Does Early Child Care Affect Children's Development? *Journal of Public Economics 159*, 33 – 53.

Heckman, J. J., S. Urzua, and E. Vytlacil (2006). Understanding instrumental variables in models with essential heterogeneity. *The Review of Economics and Statistics 88*(3), 389–432.

Heckman, J. J. and E. Vytlacil (1999). Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects. *Proceedings of the National Academy of Sciences USA 96*(8), 4730–4734.

Heckman, J. J. and E. Vytlacil (2005). Structural Equations, Treatment Effects, and Econometric Policy Evaluation. *Econometrica 73*(3), 669–738.

Heckman, J. J. and E. J. Vytlacil (2007). Econometric Evaluation of Social Programs, Part II. Volume 6 of *Handbook of Econometrics*, pp. 4875 – 5143. Elsevier.

Kamhöfer, D. A., H. Schmitz, and M. Westphal (2019). Heterogeneity in Marginal Non-Monetary Returns to Higher Education. *Journal of the European Economic Association 17*(1), 205–244.

Kline, P. and C. R. Walters (2016). Evaluating Public Programs with Close Substitutes: The Case of Head Start. *The Quarterly Journal of Economics 131*(4), 1795–1848.

Lang, J. (2018). Employment Effects of Language Training for Unemployed Immigrants. *IAB Discussion Paper 21/2018 Institute for Employment Research, Nürnberg*.

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.

Miranda, A. and Y. Zhu (2013). English Deficiency and the Native Immigrant Wage Gap. *Economics Letters 118*(1), 38–41.

Nybom, M. (2017). The Distribution of Lifetime Earnings Returns to College. *Journal of Labor Economics 35*(4), 903–952.

Riphahn, R. T. and S. Saif (2018). Naturalization and Labor Market Performance of Immigrants in Germany. *SOEP Papers 980*.

Robinson, P. M. (1988). Root-N-Consistent Semiparametric Regression. *Econometrica 56*(4), 931–954.

Sarvimäki, M. and K. Hämäläinen (2016). Integrating Immigrants: The Impact of Restructuring Active Labor Market Programs. *Journal of Labor Economics 34*(2), 479–508.

Trübswetter, P. and T. Fendel (2016). IAB-SOEP Migrationsstichprobe verknüpft mit administrativen Daten des IAB, Version 1 (IAB-SOEP-MIG-ADIAB 7514). *FDZ-Datenreport 11/2016, Institute for Employment Research (IAB), Nürnberg.*

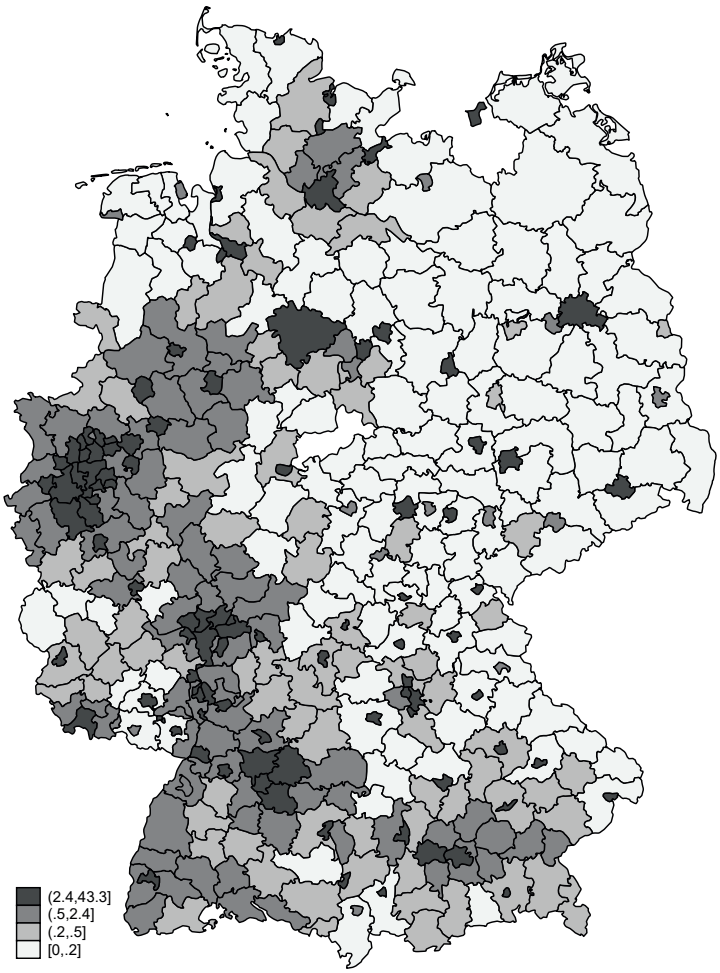
Willis, R. and S. Rosen (1979). Education and Self-Selection. *Journal of Political Economy 87*(5), S7–36.

Yao, Y. and J. C. van Ours (2015). Language Skills and Labor Market Performance of Immigrants in the Netherlands. *Labour Economics 34*, 76–85.

Appendix

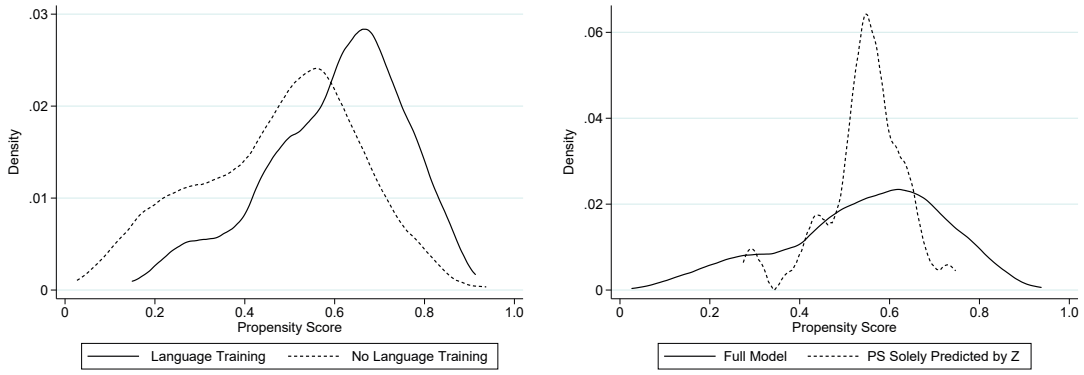
Figures

Figure 1: Regional Variation in Language Training Availability



Source: Own illustration based on county level information from integration course statistics (Federal Office for Migration and Refugees, 2013, 2014). The original spreadsheets including county level information are available from the authors. For Germany-wide statistics on language course supply, see also Federal Office for Migration and Refugees (2017), p.3.
Note: The figure depicts regional variation in language training availability. The spatial distribution is measured as the number of supplied slots that are bounded by the total number of participants, measured per square kilometer at the level of the 401 German counties. The number of open slots, conditional on a set of spatial background variables such as foreign share and population density, define language training availability (LTA) and thus the instrument.

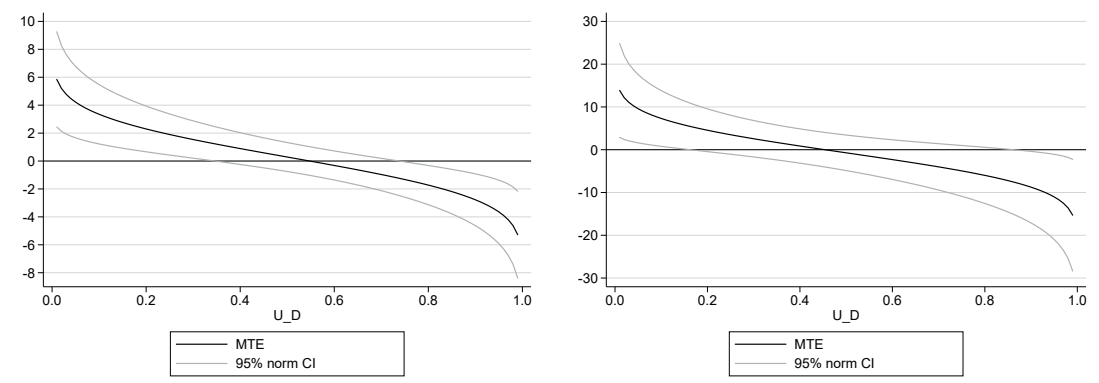
Figure 2: Common Support of the Propensity Score



(a) Support by Treatment Status (b) Overall Variation and Identifying Variation

Source: Own calculation based on IAB-SOEP-MIG-ADIAB 7514, Version 1, 2013-2014.
Note: Panel (a) plots the density distribution of the propensity score by treatment status, predicted from the first-stage selection equation of the baseline specification (Probit). The solid line in panel (b) plots the joint density distribution of the propensity score (overall variation) and for only relying on variation from Z (identifying variation) when integrating out variation from covariates X (dashed line).

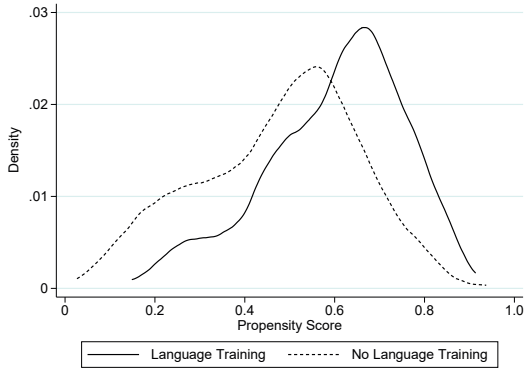
Figure 3: Marginal Treatment Effects of Language Training (Parametric)



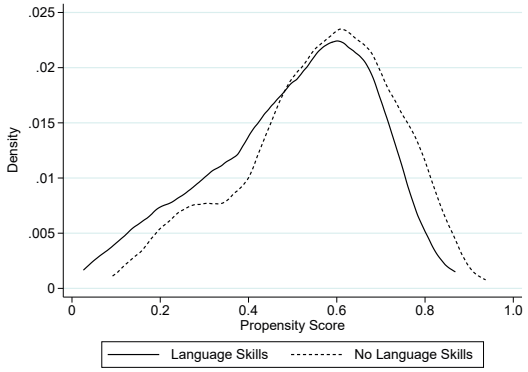
(a) Employment: 8 Years after Arrival (b) Wage: 8 Years after Arrival

Source: Own calculation based on IAB-SOEP-MIG-ADIAB 7514, Version 1, 2013-2014.
Note: The figures plot marginal treatment effects of language training on the employment probability (panel a) and on log daily wages (panel b). The baseline follow-up period is eight years after arrival.

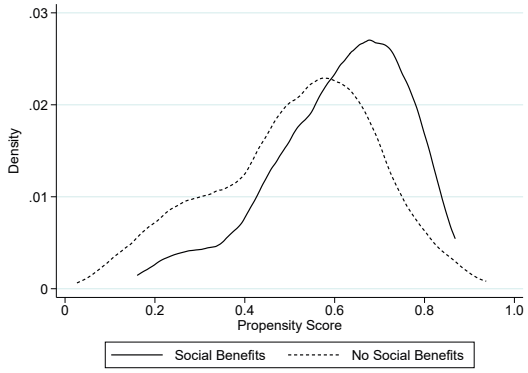
Figure 4: Common Support of the Propensity Score by Observable Characteristics



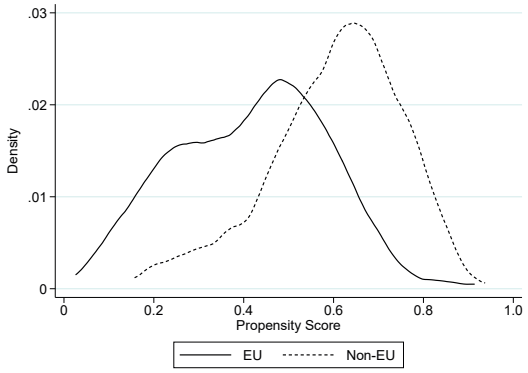
(a) Language Training (Treatment)



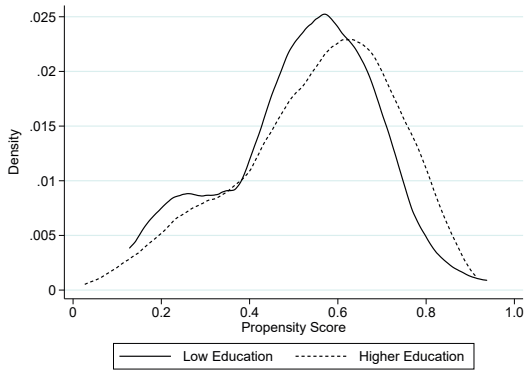
(b) Language Skills at Arrival



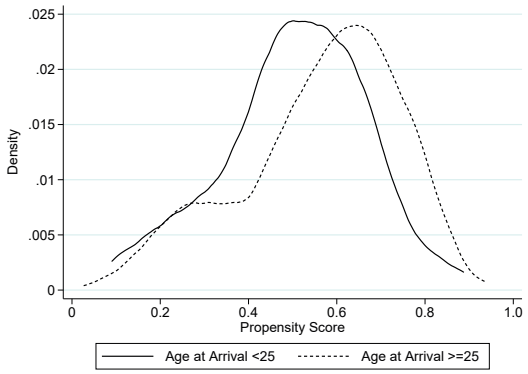
(c) Social Benefit Receipt



(d) Region of Origin



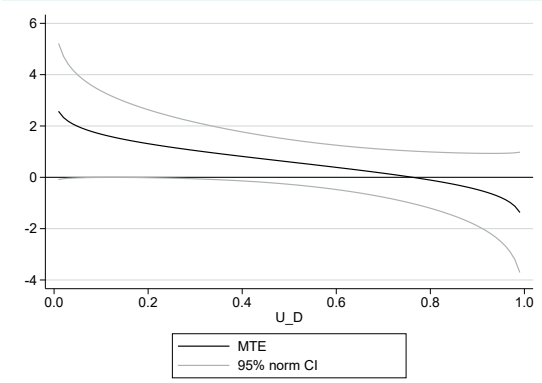
(e) Education



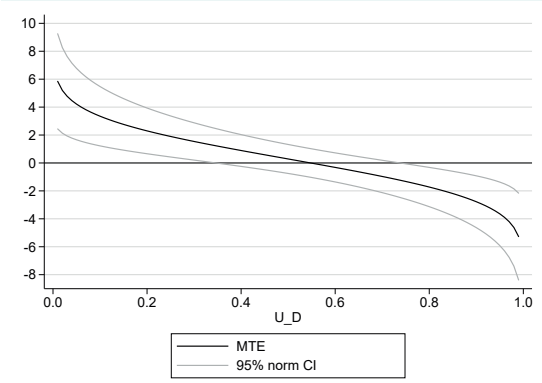
(f) Age at Arrival

Source: Own calculation based on IAB-SOEP-MIG-ADIAB 7514, Version 1, 2013-2014.
Note: The figures plot the density distribution of the propensity score by treatment status (baseline, panel a) and socio-economic characteristics as indicated. Densities of the propensity score are predicted from the first-stage selection equation of the baseline specification (Probit).

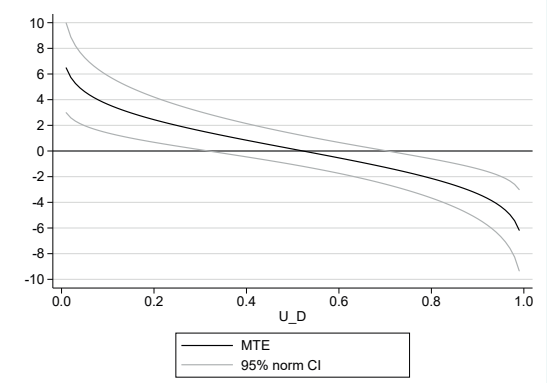
Figure 5: Marginal Treatment Effects on Employment by Follow-up Period



(a) Employment: 5 Years



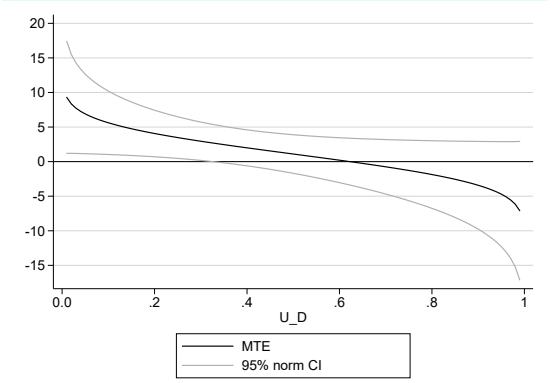
(b) Employment: 8 Years (Baseline)



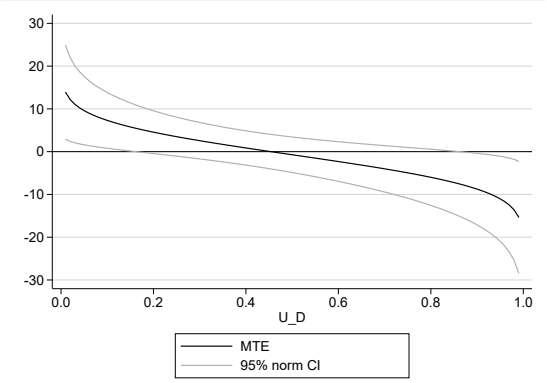
(c) Employment: 10 Years

Source: Own calculation based on IAB-SOEP-MIG-ADIAB 7514, Version 1, 2013-2014.
Note: The figures plot marginal treatment effects of language training on the employment probability for different follow-up periods, measured in years after arrival.

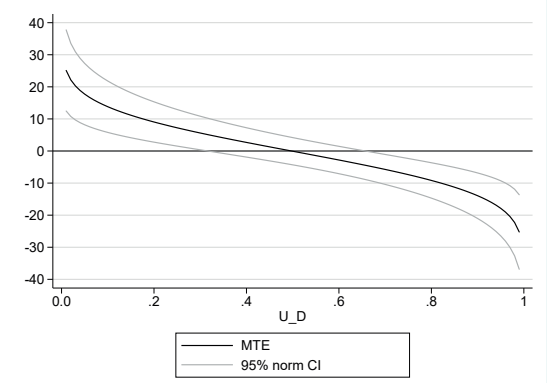
Figure 6: Marginal Treatment Effects on Wages by Follow-up Period



(a) Wages: 5 Years



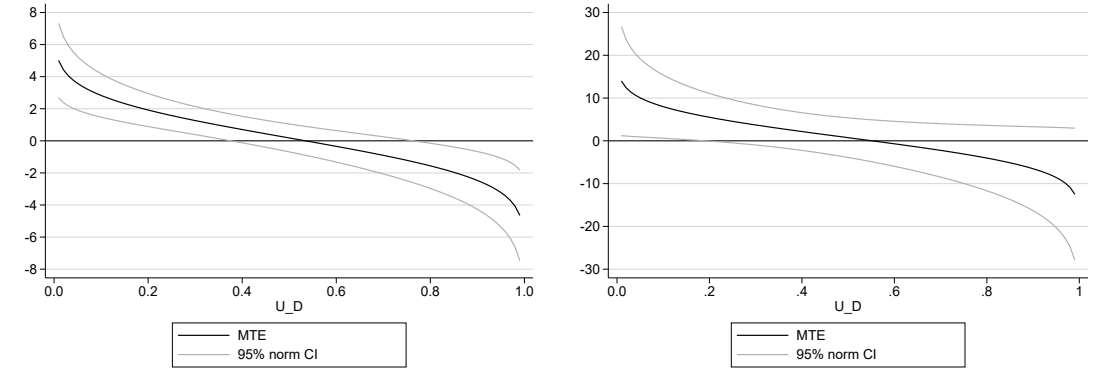
(b) Wages: 8 Years (Baseline)



(c) Wages: 10 Years

Source: Own calculation based on IAB-SOEP-MIG-ADIAB 7514, Version 1, 2013-2014.
Note: The figures plot marginal treatment effects of language training on log daily wages for different follow-up periods, measured in years after arrival.

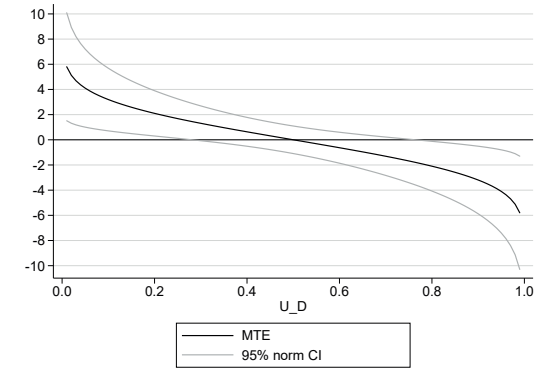
Figure 7: Marginal Treatment Effects When County Debt is Included



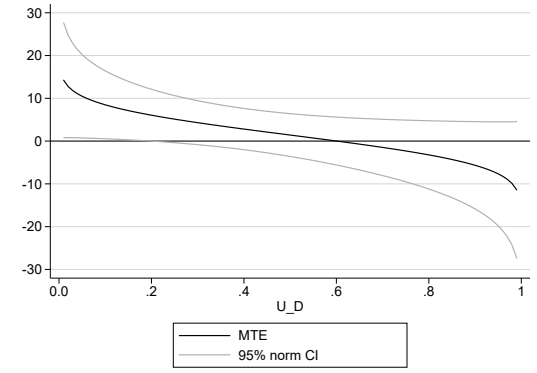
(a) Employment: 8 Years (Baseline) (b) Wages: 8 Years (Baseline)

Source: Own calculation based on IAB-SOEP-MIG-ADIAB 7514, Version 1, 2013-2014.
Note: The figures plot marginal treatment effects of language training on both outcomes (as indicated) from regressions that include public debt per capita at the county level.

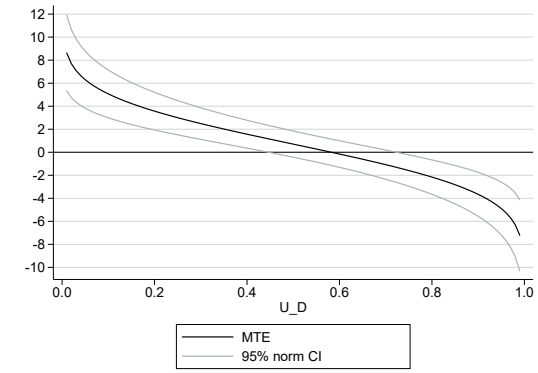
Figure 8: Marginal Treatment Effects Accounting for Selective Movements Across Regions



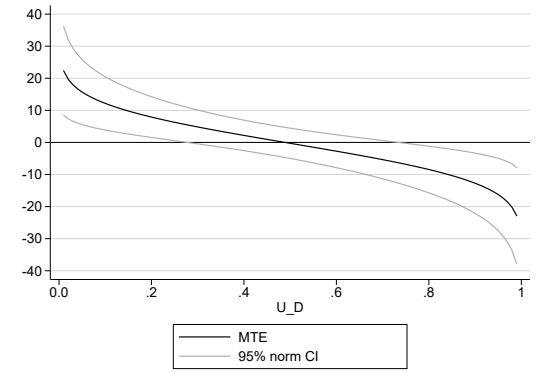
(a) Employment: Moving-Indicator



(b) Wages: Moving-Indicator



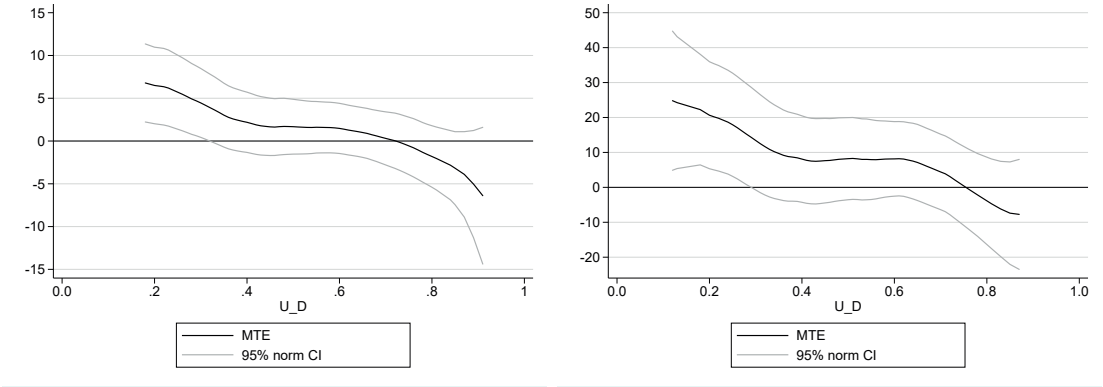
(c) Employment: Non-Mover Sample



(d) Wages: Non-Mover Sample

Source: Own calculation based on IAB-SOEP-MIG-ADIAB 7514, Version 1, 2013-2014.
Note: All estimates use the baseline follow-up period of 8 years. The figures plot marginal treatment effects of language training on both outcomes (as indicated) from regressions that include a variable that indicates whether an immigrant has moved to a different county within 3 years after arrival (= 1) or not (= 0) in panel a and b. Estimated MTE-curves in panel c and d are restricted to a sample of non-movers (N:679).

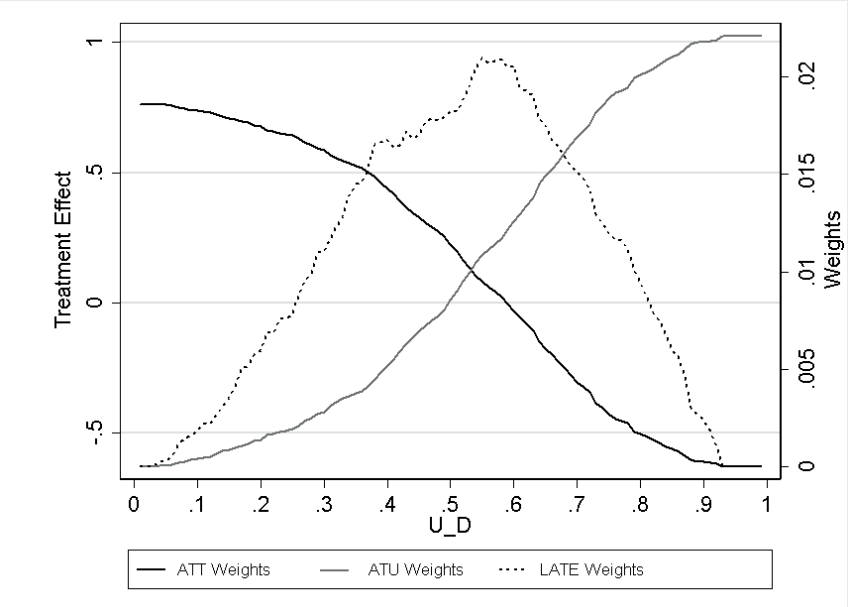
Figure 9: Marginal Treatment Effects of Language Training (Semi-Parametric)



(a) Employment: 8 Years after Arrival (b) Wages: 8 Years after Arrival

Source: Own calculation based on IAB-SOEP-MIG-ADIAB 7514, Version 1, 2013-2014.
Note: The figures plot marginal treatment effects of language training on the employment probability (panel a) and on log daily wages (panel b). The baseline follow-up period is eight years after arrival. MTEs are estimated semi-parametrically for the region with common support of the propensity score.

Figure 10: Treatment Parameter Weights



Source: Own calculation based on IAB-SOEP-MIG-ADIAB 7514, Version 1, 2013-2014.
Note: The figure depicts treatment parameter weights conditional on the propensity score. Left vertical axis: treatment effect. Right vertical axis: weights. The weights were calculated using the entire estimation sample for the baseline specification.

Tables

Table 1: Descriptive Statistics by Participation in Language Training

	Language Training	No Language Training	Difference in Means
Socio-economic and Migration Variables			
Male	43.3	45.4	2.0
Married/Partner (%)	72.2	73.7	1.5
Age	42.3	40.7	−1.6**
Age at arrival	30.5	28.8	−1.7***
High education, ISCED 5–6 (%)	26.7	28.9	2.1
Middle education, ISCED 3–4 (%)	51.9	47.2	-4.8
Low education, ISCED 1–2 (%)	21.1	23.5	2.3
Social benefits at the beginning of the residence (%)	35.3	27.1	−8.3***
Years of residence	11.8	11.9	0.1
Residency Status at Arrival (%)			
Family member	34.7	35.6	0.9
Asylum seeker	6.5	7.5	-1.0
Ethnic Germans	35.3	18.8	−16.5***
Job searcher	4.1	12.6	8.5***
With job commitment	9.5	11.9	2.4
Other status groups	9.1	13.4	4.4**
Country of Origin (%)			
EU	25.2	41.0	15.8***
EU founder nations	3.9	7.0	3.1**
Countries of EU enlargement 2004	9.3	16.0	6.7***
Turkey	4.3	3.1	-1.2
Arabic	1.9	2.6	0.6
Guest-worker countries	11.4	12.1	0.7
Russia and (former) USSR	53.4	36.9	−16.7***
Occupational Sectors at Arrival (%):			
Employed at home	73.7	66.8	−7.0**
Blue collar	26.1	22.4	−3.7
White collar	40.5	38.1	-2.4
Public sector	3.2	1.5	−1.7
Self-employed	3.9	4.6	-0.8
Observations	599	971	1,570

Source: Own calculation based on IAB-SOEP-MIG-ADIAB 7514, Version 1, 2013-2014. *Note:* Sample means are reported by treatment status for the corresponding groups with language training and without language training. Two-sided t-tests indicate whether differences in means differ significantly: $^{*}p < 10\%$, $^{**}p < 5\%$, $^{***}p < 1\%$.

Table 2: Distribution of Outcome Variables, Language Skills, and Instruments

	(1)	(2)	(3)	(4)	(5)
	Mean				
	Language Training	No Language Training	Mean Diff. (1) - (2)	Min	Max
Panel A: Labor Market Outcomes					
Employed 5 years after arrival (%)	45.3	55.7	−10.4***	0	1
Regularly Employed (%)	35.1	44.3	−9.2***	0	1
Marginally Employed (%)	10.1	11.3	-1.2	0	1
Unemployed (%)	33.4	21.4	12.0***	0	1
Apprenticeship (%)	1.9	2.3	-0.4	0	1
Employable (%)	1.9	2.3	-0.4	0	1
Employed 8 years after arrival (%)	**53.0**	**58.2**	**-5.2**	**0**	**1**
Regularly Employed (%)	37.3	44.1	−6.8**	0	1
Marginally Employed (%)	15.7	14.2	1.6	0	1
Unemployed (%)	22.0	23.7	-1.7	0	1
Apprenticeship (%)	2.4	2.3	0.1	0	1
Employable (%)	2.4	2.3	0.1	0	1
Employed 10 years after arrival (%)	53.7	56.3	-2.6	0	1
Regularly Employed (%)	37.2	44.7	−7.5**	0	1
Marginally Employed (%)	16.5	11.6	4.9*	0	1
Unemployed (%)	26.1	29.4	-3.3	0	1
Apprenticeship (%)	1.0	0.9	0.1	0	1
Employable (%)	16.7	12.8	3.9	0	1
Daily Wage 5 years after arrival (EUR)	26.7	30.9	-4.2	0	180.8
Daily Wage 8 years after arrival (EUR)	**31.6**	**35.5**	**-3.9**	**0**	**194.5**
Daily Wage 10 years after arrival (EUR)	32.4	32.7	-0.3	0	195.6
Panel B: German Language Skills					
Good command of German – currently (%)	77.2	83.0	−5.8**	0	1
Good command of German – at arrival (%)	22.3	27.7	−5.4*	0	1
Speaking (1: very good – 5: very poor)	4.0	3.6	0.3***	1	5
Writing (1: very good – 5: very poor)	4.0	3.7	0.3***	1	5
Reading (1: very good – 5: very poor)	3.8	3.5	0.3***	1	5
Panel C: Instrument and Spatial Variables					
LTA Instrument (Z_{jt})	2.7	1.4	1.3***	0	27
Foreign share (%)	9.3	11.1	−1.7***	1.1	32.3
Unemployment rate of foreigners (%)	14.9	14.0	0.9**	3.8	31.6
Population density (inhabitants per km^2)	910	1,285	−375***	37.5	4,601.2
Disposable income per inhabitant (EUR)	21,009	21,367	−358*	15,734	41,707
Observations	599	971	1,570		

Source: Own calculation based on IAB-SOEP-MIG-ADIAB 7514, Version 1, 2013-2014. *Note:* Sample means are reported by treatment status for the corresponding groups with language training and without language training. The LTA instrument measures language training availability as the number of open slots at the county level. The baseline outcomes are regular employment and daily wage, measured eight years after arrival. Two-sided t-tests indicate whether differences in means differ significantly: $^{*}p < 10\%$, $^{**}p < 5\%$, $^{***}p < 1\%$.

Table 3: OLS and 2SLS Estimates

	OLS		2SLS	
	Employment	Wage	Employment	Wage
	(1)	(2)	(3)	(4)
First Stage:				
LTA Instrument	–	–	0.031***	0.031***
			(0.011)	(0.011)
F-Statistic	–	–	22.5	22.5
Second Stage:				
Language Training	-0.044	-0.174	0.171	0.861
	(0.032)	(0.123)	(0.138)	(0.534)
Good command of German – at arrival	0.151***	0.628***	0.157***	0.706***
	(0.038)	(0.144)	(0.040)	(0.155)

Source: Own calculation based on IAB-SOEP-MIG-ADIAB 7514, Version 1, 2013-2014. *Note:* Both outcomes (employment dummy and log daily wage) are measured eight years after arrival, corresponding to the baseline specification of the marginal treatment effects estimation below. The LTA instrument measures language training availability as the number of open slots at the county level. All specifications include socio-economic variables, residency status at arrival, spatial characteristics, country of origin fixed effects and state fixed effects. $^{*}p < 10\%$, $^{**}p < 5\%$, $^{***}p < 1\%$. Standard errors in parentheses.

Table 4: Marginal Treatment Effects on Employment

	Dependent Variable: Employment (0/1)			
	(1) Selection		(2) Outcome	
Panel A: Instruments and Spatial Variables				
LTA Instrument (Z_{jt})	0.036***	(0.013)		
Population Density	−0.0002***	(0.0001)		
Foreign Share	0.529	(0.866)		
UE Rate Foreigners	1.015	(0.660)		
Disposable Income * 1,000	-0.014	(0.01)		
Panel B: Variables on Integration, Migration and Education				
Age at Arrival	0.005**	(0.002)	-0.008	(0.007)
EU	−0.126*	(0.070)	0.185	(0.213)
German Language Skills at Arrival	−0.117***	(0.045)	-0.058	(0.113)
Social Benefits at Arrival	0.026	(0.044)	−0.329***	(0.119)
High Education	-0.011	(0.045)	0.152	(0.126)
Low Education	−0.080*	(0.048)	−0.454***	(0.115)
Family Member	−0.197*	(0.114)	0.172	(0.138)
Asylum Seeker	−0.385*	(0.212)	0.567**	(0.238)
Job Searcher	−0.903***	(0.210)	-0.220	(0.183)
Panel C: Interactions by Treatment Status ($\hat{p}$)				
Age at Arrival * $\hat{p}$			0.001	(0.011)
EU * $\hat{p}$			-0.024	(0.416)
German Language Skills at Arrival * $\hat{p}$			0.469**	(0.235)
Social Benefits at Arrival * $\hat{p}$			0.500**	(0.216)
High Education * $\hat{p}$			-0.118	(0.230)
Low Education * $\hat{p}$			0.660***	(0.223)
Family Member * $\hat{p}$			-0.365	(0.254)
Asylum Seeker * $\hat{p}$			−1.127***	(0.401)
Job Searcher * $\hat{p}$			0.555	(0.503)
$\hat{p}$			-0.287	(0.476)
Fixed Effects:				
Immigration Year (I)	Yes		Yes	
Country of Origin (O)	Yes		Yes	
Occupational Sector (S)	Yes		Yes	
Region (R)	Yes		Yes	
Chi-Squared for Test of Excluded Instruments	113.3***			
Test of Heterogeneity		−2.164***	(0.798)	

Source: Own calculation based on IAB-SOEP-MIG-ADIAB 7514, Version 1, 2013-2014. *Note:* Reported values for the first stage selection equation (1) are marginal effects from Probit regressions, defining participation in language training as treatment. The first stage estimates predict the treatment probability (propensity score $\hat{p}$). The LTA instrument measures language training availability as the number of open slots at the county level. The LTA instrument and spatial variables satisfy an exclusion restriction and are not part of the outcome equation (2). For the second stage outcome equation, OLS coefficients are reported. The dependent variable is a dummy that equals one for observations that are employed eight years after arrival and zero otherwise. The calculation of the MTE is based on first-order polynomial estimates. Estimates for gender and marital status and their interactions with $\hat{p}$ are not shown for brevity but are available from the authors. $^{*}p < 10\%$, $^{**}p < 5\%$, $^{***}p < 1\%$. Standard errors in parentheses.

Table 5: Marginal Treatment Effects on Wages

	Dependent Variable: Log Daily Wage			
	(1) Selection		(2) Outcome	
Panel A: Instruments and Spatial Variables				
LTA Instrument (Z_{jt})	0.036^{***}	(0.013)		
Population Density	-0.0002^{***}	(0.0001)		
Foreign Share	0.529	(0.866)		
UE Rate Foreigners	1.015	(0.660)		
Disposable Income * 1,000	-0.014	(0.01)		
Panel B: Variables on Integration, Migration and Education				
Age at Arrival	0.005^{**}	(0.002)	-0.050^{*}	(0.030)
EU	-0.126^{*}	(0.070)	0.798	(0.723)
German Language Skills at Arrival	-0.117^{***}	(0.045)	0.909^{**}	(0.458)
Social Benefits at Arrival	0.026	(0.044)	-1.526^{***}	(0.586)
High Education	-0.011	(0.045)	0.992^{*}	(0.551)
Low Education	-0.080^{*}	(0.048)	-1.023^{*}	(0.548)
Family Member	-0.197^{*}	(0.114)	0.536	(0.471)
Asylum Seeker	-0.385^{*}	(0.212)	1.371	(0.931)
Job Searcher	-0.903^{***}	(0.210)	-1.102	(0.702)
Panel C: Interactions by Treatment Status ($\hat{p}$)				
Age at Arrival * $\hat{p}$			0.027	(0.051)
EU * $\hat{p}$			0.416	(1.361)
German Language Skills at Arrival * $\hat{p}$			-0.095	(0.781)
Social Benefits at Arrival * $\hat{p}$			1.936^{*}	(1.027)
High Education * $\hat{p}$			-0.965	(0.895)
Low Education * $\hat{p}$			1.157	(0.885)
Family Member * $\hat{p}$			-1.198	(0.874)
Asylum Seeker * $\hat{p}$			-3.102^{*}	(1.764)
Job Searcher * $\hat{p}$			3.794^{*}	(1.999)
$\hat{p}$			0.129	(1.735)
Fixed Effects:				
Immigration Year (I)	Yes		Yes	
Country of Origin (O)	Yes		Yes	
Occupational Sector (S)	Yes		Yes	
Region (R)	Yes		Yes	
Chi-Squared for Test of Excluded Instruments	113.3^{***}			
Test of Heterogeneity		-5.358^{*}	(2.827)	

Source: Own calculation based on IAB-SOEP-MIG-ADIAB 7514, Version 1, 2013-2014. *Note:* Reported values for the first stage selection equation (1) are marginal effects from Probit regressions, defining participation in language training as treatment. The first stage estimates predict the treatment probability (propensity score $\hat{p}$). The LTA instrument measures language training availability as the number of open slots at the county level. The LTA instrument and spatial variables satisfy an exclusion restriction and are not part of the outcome equation (2). For the second stage outcome equation, OLS coefficients are reported. The dependent variable is log daily wage, assuming zero wages for those who are not employed. The calculation of the MTE is based on first-order polynomial estimates. Estimates for gender and marital status and their interactions with $\hat{p}$ are not shown for brevity but are available from the authors. $^{*}p < 10\%$, $^{**}p < 5\%$, $^{***}p < 1\%$. Standard errors in parentheses.

Table 6: Estimated Treatment Parameters

	(1)	(2)	(3)	(4)
	Treatment Parameter			
	ATE	ATT	ATU	LATE
Outcome				
Employment	0.199	0.429**	-0.074	0.161
	(0.139)	(0.197)	(0.183)	(0.133)
Wage	1.416	2.503***	-0.093	0.934*
	(2.380)	(0.741)	(0.691)	(0.503)

Source: Own calculation based on IAB-SOEP-MIG-ADIAB 7514, Version 1, 2013-2014. *Note:* Both outcomes, employment and daily wage, are measured eight years after arrival, corresponding to the baseline specification of the marginal treatment effects estimation above. $^{*}p < 10\%$, $^{**}p < 5\%$, $^{***}p < 1\%$. Standard errors in parentheses.

A Institutional Details: The German Law of Immigration

The conception and central coordination of language courses is governed by the Federal Office for Migration and Refugees (Bundesamt für Migration und Flüchtlinge, BAMF). This institution determines the content of language training, quality standards regarding courses and teachers, and teachers' salaries at the national level. Furthermore, the BAMF requires cooperation between the providers of language training, the German federal employment agency, the providers of welfare benefits, and the foreigners' registration office. This ensures that quality standards are set at the national level to make language training homogeneous across regions.

Initial placement tests are conducted by providers of language training to assign migrants to their corresponding level of language. To increase the effectiveness, there are different types of trainings available for specific needs such as alphabetization courses (for illiterate persons) or courses only for women. The curriculum consists of a basic language training followed by an advanced course, aiming to shift immigrants language skills to the European reference level B2. This level ensures that immigrants obtain upper intermediate language skills that can be used independently. After a workload of about 600 lessons (one lesson á 45 minutes), the course is completed by a final test (Federal Office for Migration and Refugees, 2008, 2011).

Language courses can be conducted by private or public providers. Allowances are granted for three years upon application at the BAMF. Admissions and the corresponding subsidies are conditional on quality standards that are controlled and supervised by the BAMF in cooperation to a central evaluation commission.

The obligation to attend in language training depends on the specific reason and is thus declared by the foreigners' registration office, by the provider of welfare benefits (benefit recipients) and by the provider of asylum seeker benefits (acquirer of asylum status).