Dan Mikael Larsen

Language Training and the Employment of Immigrants

An empirical analysis of the effect of participation in Norwegian language training and social studies on the employment of immigrants.

Master's thesis in Economics
Supervisor: Costanza Biavaschi
Trondheim, September 2018

Norwegian University of Science and Technology
Faculty of Economics and Management
Department of Economics

NTNU
Norwegian University of
Science and Technology

Preface

This thesis concludes my economics studies at NTNU. It has truly been an eventful and instructive period. First and foremost, I want to express my gratitude to my supervisor Costanza Biavaschi for her enthusiasm and excellent guidance during the work. Her feedback has been indispensable for the completion of the thesis.

I would further want to give credit where credit's due. Thank you Jarl Mattias Larsen (Larsen Consultance), Jonas Danielsen (Marathon Man), Marie Skara (The Office), Torstein Fjeldstad (UN Statistics), Martin Steffensen (GPT) and Alexander Aamo (Fit Economy Inc.) for much appreciated feedback and assistance. Finally, a big thank you to my father for constantly encouraging me during my studies. Without your assistance I could never have done this.

Trondheim, September 3, 2018.

Dan Mikael Larsen

Abstract

This thesis examines how participation in Norwegian language training and social studies affects immigrant employment by exploiting panel data at the municipal level between 2014 and 2016. Research on the effect of language training is complicated by several endogeneity concerns. First, unobserved heterogeneity may induce immigrants to select participation in the training. In addition, unobserved municipal characteristics may cause immigrants to select into certain municipalities. The main analysis addresses these issues by first employing fixed effects estimation and second an array of control variables. Across a battery of models and specifications, the study finds a positive, albeit economically small, impact of participation in the training on employment. There is no evidence of a gap in the returns to participation across municipalities with a relatively high share of African and Asian immigrants, male immigrants, or young immigrants of working age. Neither is there support of differences across urban and rural municipalities, nor municipalities with varying political parties in local governance. Results are found to be robust against several extensions, including an IV strategy.

Sammendrag

Denne oppgaven evaluerer hvordan deltakelse i norskopplæring påvirker antall sysselsatte innvandrere på kommunenivå. Til dette utnyttes paneldata for perioden 2014-2016. Studier på effekten av språkopplæring kompliseres på grunn av endogenitetsproblemer. For det første kan uobservert heterogenitet føre til at innvandrere selv-selekterer til deltakelse. Dernest kan uobserverte kommunale karakteristikka forårsake at innvandrere selv-selekterer til spesifikke kommuner. Hovedanalysen håndterer disse utfordringene ved bruk av kommunespesifikke faste effekter og et utvalg av sosioøkonomiske kontrollvariable. På tvers av flere modeller og spesifikasjoner finner oppgaven en positiv effekt av deltakelse i norskopplæring på sysselsetting. Effekten er dog økonomisk liten. Det finnes ikke bevis for forskjeller i effekten på tvers av kommuner med en relativt høy andel av innvandrere fra Afrika og Asia, mannlige innvandrere eller unge innvandrere i arbeidsfør alder. Det er videre ingen støtte for ulik effekt i urbane og rurale områder eller på tvers av kommuner med ulike politiske partier. Resultatene i hovedanalysen er bevist robuste gjennom flere robusthetsanalyser.

Contents

1. Introduction

The Norwegian government considers participation in the labour market as the best way to integrate immigrants (Ministry of Justice and Public Security, 2015-2016, p.13). Participation increases social skills and counteracts the development of parallel communities, in addition to improving language skills. The latter is of substantial importance as language skills are considered to be among the major obstacles to the integration of immigrants ((Chiswick, 1991, p.167); (Borjas, 2014, Chapter 2, para.1)). Language deficiency complicates the process of attaining a native social network and reduces the worker's productivity. Furthermore, employers might require a high level of language proficiency when hiring workers (NOU 2017:2, 2017, p.72). The study of the relationship between language skills and labour market performance is thus important.

The recent decades have seen an upsurge of immigration to Norway. Between 2006 and 2011, net immigration accounted for 67% of the population growth in Norway (Hagelund et al., 2011, p.2). However, immigrants are often a vulnerable part of the labour force. In 2015, unemployment among immigrants amounted to 11.2%, compared to 4.5% for the total population. An important tool to meet the needs of considerable immigration and the relatively high unemployment among immigrants is the programme *Norwegian language training and social studies* (Ministry of Justice and Public Security, 2016, p. 55). The training aspires to augment proficiency in the Norwegian language and labour market participation for newly arrived immigrants.

This thesis examines how participation in *Norwegian language training and social studies* affect immigrant employment. In particular, it studies the effect on the number of employed immigrants in a municipality when the number of participants in the training increases in that municipality. The research question is answered by analysing a strongly balanced panel data set at the municipal level in the period 2014-2016.

This thesis addresses identification challenges that arise in studies on language training and labour market performance at the municipal level. First, unobserved heterogeneity may cause immigrants to self-select into participation. Second, unobserved local conditions may induce immigrants to self-select into specific municipalities. As an illustration, assume that an immigrant chooses to participate in the training. Suppose then that this immigrant enrolled in the training has a much higher probability of employment than an immigrant who chooses not to participate. However, these immigrants might differ along other dimensions such as motivation or socioeconomic background. Because such factors may affect the decision of participation and possibly employment later on, the estimated effect of participation in the training on employment may lead to biased estimates that make causal statements difficult to assert. Through a fixed effects estimation and use of socioeconomic controls, the analysis attempts to correct for self-selection.

Across various models and specifications, the study identifies a positive relationship between the number of participants in training and the number of employed immigrants in a municipality. The causal impact is nonetheless economically small. There is no evidence of a gap in returns to participation across municipalities with a relatively high share of African and Asian immigrants, male immigrants or young immigrants of working age. Neither is there support of differences across urban and rural municipalities or municipalities with varying political parties in local governance. The finding is robust throughout several extensions. A static model underlines the importance of lagging the key independent variable in the main analysis. Age cohorts and bottom coding of missing values further validate the results. Finally, the study displays how results are robust not only to fixed effects, but also to the inclusion of an instrumental variable that exploits immigrants that are exogenously allocated.

The study is interesting for several reasons. First, it is interesting from a policy point of view to assess the relationship between the training and employment. Forasmuch as the language training is among the major integration tools in Norway, knowledge about its efficacy is valuable. Second, previous studies on language skills and assimilation focus primarily on the effect on earnings, while this thesis extends the narrow scope of literature that examines the impact on employment. Moreover, there is relatively little literature on the evaluation of public policy programmes targeting immigrants. Thus, the limited research for the link between language training and employment underlines the importance of this study. Finally, opposed to the majority of literature on language and labour market performance, this thesis examines the context of a small country where English might serve as a substitute for the local language. As communication with natives in English might be an alternative for immigrants in Norway, it is highly interesting to consider the importance of language training in terms of immigrants' labour market performance. To the best of my knowledge, this thesis is the first study in Norway attempting a causal identification of the training on employment at the municipal level.

The remainder of the thesis proceeds as follows. Chapter 2 summarises relevant literature, while Chapter 3 provides background information on the Norwegian language training and social studies. Chapter 4 describes the data set and provides descriptive statistics for the population sample. Chapter 5 proceeds to discuss identification of the key independent variable and build the empirical model of the analysis. Chapter 6 presents and comments parameter estimates with OLS regression in addition to investigating model validity. Chapter 7 and 8 present the set-up and results of the main analysis respectively. The latter also provides a discussion of findings in light of previous literature. Chapter 9 examines estimates across municipal heterogeneity. Chapter 10 confirms robustness of the main findings through an extensive robustness analysis. Finally, Chapter 11 concludes.

2. Language and Labour Market Performance

This study draws on and supplements previous literature on language training and labour market performance of immigrants. A collection of former studies, though not exhaustive, is listed in Table 2.1 at the end of this chapter. Ever since the suggestion of a possible link between English language proficiency and economic assimilation of immigrants by Chiswick (1978), an extensive literature regarding the returns to language skills has emerged. Consequently, this review only highlights selected works.

The Impact of Language Skills

Much of the literature examining returns to language proficiency is based on self-reported measures from surveys, in which immigrants may over- or underestimate their true level of fluency. In a study from 2003, Dustmann & Fabbri (2003) use cross-sectional data of immigrants from two UK surveys[1] on ethnic minorities to assess the effect of English proficiency on earnings and employment probability. The paper examines individuals already in the labour force, but distinguishes between employed and unemployed immigrants actively searching for employment. In order to address the problem with self-reported proficiency, the authors utilise an instrumental variable to eliminate bias resulting from measurement error. Furthermore, they apply a matching estimator to deal with endogeneity in language acquisition. Their estimates show that English proficiency increments employment probabilities by 22 percentage points. Moreover, results indicate a downward bias of estimates when measurement error is not corrected for.

Motivated by the possible different effects of a smaller language, Yao & van Ours (2015) analyse the link between Dutch deficiency and labour market performance. Relying on the Longitudinal Internet Studies for the Social sciences (LISS) panel data survey, the authors identify three threats to estimating the effect of language skills on labour market performance. First, unobserved characteristics may cause upward bias in estimates. Second, employment experience implies reverse causality. Finally, since the study relies on self-reported information, measurement errors might lead to underestimation. Addressing these issues, Yao

[1] Fourth National Survey on Ethnic Minorities (FNSEM) collected between 1993 and 1994, and Family and Working Lives Survey (FWLS) collected between 1994 and 1995.

& van Ours (2015) use an instrumental variable as the interaction between childhood language and age at arrival in the Netherlands. In contrast to the paper by Dustmann & Fabbri (2003), they find no significant effect on employment. The authors argue that these findings indicate no disadvantage for immigrants deficient in small European languages. A speculative reason might be substitutability between the local language and English in countries excelling high proficiency in the latter.

The Impact of Language Training

Albeit the literature on language skills and labour market performance is vast, Chiswick & Miller (2014, p. 7-8) point out that studies on the returns of language training is noticeably restricted due to limited data. In particular, data sets rarely report immigrant participation in language courses.

In an early study by Beenstock (1996) in Israel, Hebrew proficiency is estimated to increase with participation in language training. Furthermore, the finding indicates that immigrants completing the training improved proficiency relative to dropouts. The study was carried out with an ordered probit analysis based on the panel data survey Immigrant Absorption Survey (IAS). However, the author does not address probable self-selection into language training. In the same way, Beiser & Hou (2000) do not correct for self-selection into English as a Second Language (ESL) courses when estimating the effect on language acquisition of Southeast Asian refugees arriving in Canada 1981. The study draws upon self-reported panel data over ten years conducted by the Refugee Resettlement Project (RRP) staff, examining proficiency at arrival, after two years and then a decade after arrival. After ten years, ordered logistic regressions estimate that ESL classes are likely to increase English proficiency relatively more for females. Polytomous logistic regressions predict increased employment with increased proficiency for both genders, but females are more likely to enter the labour market than males as a result of English proficiency.

Unlike the two aforementioned papers, Gonzalez (2000) and Hayfron (2001) correct for self-selection into language training. The former study investigates determinants of proficiency in understanding, speaking, reading and writing English, and the wage returns to proficiency. Based on cross-sectional data from the National Adult Literacy Survey (NALS), the author addresses self-selection into language training by employing an instrumental variable of speaking another language at work or when shopping. Thus, he corrects for the lowered incentive of participating in language training for immigrants settled down in ethnic enclaves. Estimates based on a probit link function indicate that attendance in ESL classes increases English language skills. The analysis also hints towards higher earning returns to oral proficiency than literacy.

Drawing on the finding of Gonzalez (2000), Hayfron (2001) points out that not correcting for language instruction may lead to omitted variable bias. By adding participation and completion as two measures of language training, Hayfron evades this issue in his study linking Norwegian language training with proficiency and earnings. The author obtains data through a questionnaire for Third World immigrant men.[2] Justifying this selection of immigrants, he presents evidence of poor labour market performance by this particular immigrant group. Hayfron further refers to lack of Norwegian language proficiency as the main obstacle for employment among this specific immigrant group. In order to assess the impact of Norwegian language training on proficiency, he estimates a probit model with an instrumental variable constructed by unemployment and other social benefits dummies. When estimating the impact of proficiency on earnings, however, the author uses ethnicity of wife and mother tongue dummy variables as instruments. In contrast to the broad literature, the study in Norway finds no significant effect on earnings. On the other hand, immigrants attending the government-sponsored language training programme are anticipated to attain a higher level of Norwegian proficiency than those who do not attend the training. Nevertheless, two shortcomings of the study are the limited subsample of the total immigrant population and that it relies on self-reported proficiency scores.

In two recent studies, Sarvimäki & Hämäläinen (2016) and Lochmann et al. (2018) employ fuzzy regression discontinuity designs (RDD) to analyse programme participation on labour market outcomes. Sarvimäki & Hämäläinen (2016) examines the effect of active labour market programmes (ALMP) on earnings and employment in Finland, in which immigrants' eligibility were based on arrival date. The paper follows participants over a period of ten years and considers the implementation of individual tutoring with caseworkers. The total amount of training remained the same, but the programmes increased the time spent in language courses. In contrast to the other Nordic study by Hayfron (2001), Sarvimaki/Hamalainen suggest that active labour market programs are highly efficient in terms of earnings, as participants' income

[2] Males aged 17-65 from Morocco, Pakistan and Chile, residing in Oslo, Akershus or Bergen.

increased considerable at the same time as their reception social benefits decreased. More important for the purpose of this thesis, however, is the lack of impact on employment. Different from that paper, per contra, Lochmann et al. (2018) employ test scores of an initial language test as an assignment variable for participation in language training. In order to examine the effect on the economic integration of immigrants in France, they use the database Enquête Longitudinale Sur l'Intégration des Primo-Arrivants (ELIPA), that contains detailed socio-demographic information combined with data from test results offered by L'Office Français de l'Immigration et de l'Intégration (OFII). The local randomised experiment allows the authors to omit unobserved skill dimensions and find evidence of a positive connection between language training and labour force participation. The authors point to an information effect as a major mechanism in which interaction with classmates and teachers enhances the immigrant's job search abilities. On the other hand, estimates hint towards a negative effect on employment although results show little robustness throughout different specifications in terms of significance.

Closer to the identification method of this study is the work by Akresh (2007). The paper uses panel data from the New Immigrant Survey-Pilot to investigate the short-run effect of English classes on immigrant earnings the first year after receiving permanent residence in the US. Correcting for unobserved ability and motivation, the author runs a fixed effects regression with an unbalanced panel. Casewise deletion was performed concerning individuals with missing data on income. However, estimates suggest no significant link between ESL attendance and earnings. The author proposes a small sample size and short-term as possible explanations for the insignificant effect.

Contributions of This Thesis

The contribution of this thesis to the aforementioned studies is fivefold. First, it provides more depth to the literature on language training and immigrant employment. Second, it extends the literature by employing panel data at the municipal level, whereas previous studies rely on individual data. Hence, the relationship between language training and employment is examined in a bigger picture. From a policy standpoint, knowledge of municipal alteration is useful for budget planning. Third, the municipal focal point ensures, to my knowledge, a unique model within this branch of literature. The estimated model includes socioeconomic variables and is estimated with a fixed effects approach. This corrects for self-selection into participation unlike the studies of Gonzalez (2000) and Hayfron (2001). Indeed, this thesis adopts the fixed effects estimation as applied by Akresh (2007), but the present sample size is sufficiently large to avoid the issue with imprecise estimates. Fourth, motivated by Yao & van Ours (2015), the study complements the literature by evaluating language training in a small host-country. Hence, the examination of language training itself, in lieu of integration plans, sets this thesis apart from the work of Sarvimäki & Hämäläinen (2016) in a comparable small country. Finally, the design of the training allows this thesis to analyse newly arrived immigrants as opposed to Gonzalez (2000) and Hayfron (2001). Unless a pre-training measure of proficiency is implemented, estimates on immigrants with several years in the host-country before attending language classes might be dubious.

Table 2.1: Summary of previous studies of language skills and language training on labour market performance.

Study	Research question	Type of data	Empirical strategy	Results
Language skills				
Dustmann and Fabbri (2003)	Effect of English proficiency on earnings and employment probability in UK	Cross-sectional data from FNSEM and FWLS surveys	Instrumental variables (number of children and minority concentration) and propensity score estimator	Positive effect of English proficiency on employment and earnings
Yao and Van Ours (2015)	Effect of Dutch deficiency on employment, hours of work and hourly earnings in the Netherlands	Panel data from LISS survey	2SLS with instrumental variable equal (childhood language*age at arrival)	Negative effect of Dutch deficiency on hourly earnings for females; no effect on employment or hours of work
Language training				
Beenstock (1996)	Effect of language training on Hebrew proficiency in Israel	Panel data from IAS survey	Ordered probit analysis	Positive effect of language training on Hebrew proficiency
Beiser and Hou (2000)	Effect of ESL courses on language acquisition in Canada for Southeast Asian refugees	Panel data from survey conducted by RRP staff	Ordered and polytomous logistic regressions	Ordered: positive effect of ELS classes on female English proficiency; Polytomous: positive effect of English proficiency on employment
Gonzalez (2000)	Effect of ESL attendance on English proficiency and wage returns to proficiency in the US	Cross-sectional data from NALS survey	Probit model with instrumental variable (speaking another language at work or when shopping)	Positive effect of ESL attendance on English proficiency; relatively higher earnings due to oral proficiency
Hayfron (2001)	Effect of language training on proficiency and wage returns to proficiency in Norway	Cross-sectional survey data	Probit model with instrumental variables (unemployment and other social benefits dummies; ethnicity of wife and mother tongue)	Positive effect of language training on Norwegian proficiency; no effect of language skills on earnings
Akresh (2007)	Effect of ESL classes on earnings in the US	Panel data from NIS-P survey	Fixed effects	No effect of ESL attendance on earnings
Sarvimaki and Hamalainen (2016)	Effect of ALMP on earnings and employment in Finland	Pooled cross-sectional data from Statistics Finland	Fuzzy RDD (assignment variable is arrival date)	Positive effect of ALMP on earnings; no effect of ALMP on employment
Lochman et al. (2017)	Effect of language traning on labour force participation, employment and earnings	Panel data from ELIPA survey and OFFI database	Fuzzy RDD (assignment variable is test scores) with fixed effects (country of origin)	Positive effect of language training on labour force participation; negative effect on employment; ambiguous effect on earnings

3. Norwegian Language Training and Social Studies

As shown in the previous chapter, certain contributions of this thesis may be attributed to the arrangement of the Norwegian language training and social studies. This section provides a rough outline of the training for the years in question, 2014-2016, which is the background for the analysis. For a detailed discussion see appendix A.

The introduction act of 2003 comprises two principle integration tools: the introduction programme and Norwegian language training and social studies. The act aspires to enhance the possibility of labour market participation for newly arrived immigrants.[3] Due to data availability, the analysis only examines Norwegian language training and social studies.[4] The Norwegian language training and social studies is often referred to as the training later in the thesis.

The purpose of the Norwegian language training and social studies is to qualify immigrants for participation in the labour market and in doing so encouraging a faster transition into work or education (Ministry of Justice and Public Security, 2016, p.16).

Eligibility

Roughly, immigrants may be clustered in three groups on the basis of age, country of origin and type of residence permit. The groups differ in respect to eligibility of training, more specifically the right to free training and the expectancy of participation. Immigrants without the right to free training may be required to pay tuition costs.[5] Table 3.1 displays eligibility of free training and expectancy of participation for the three immigrant groups. Immigrants between 16 and 55 years of age granted permit of residence as refugees, and families migrating to reunite with such immigrants, have the right to free training and are expected to enroll in the training. Immigrants between 55 and 67 years old granted permit of residence as refugees, and their families, are not expected to undertake the training. Still they have the right to training free of

[3] The act also aims to increase the propensity to undertake education as this may further prepare participants for the labour market. As a result, the labour market and education are both objectives in the act. The focus of this analysis is nonetheless labour market participation.

[4] A brief discussion on component choice is provided in appendix A. Notwithstanding, the reader should note that the *introduction programme* is of limited relevance in the thesis.

[5] Tuition costs vary between municipalities. For instance, the cost per hour is 46 NOK in Lier, while it is 75 NOK in Bergen.

charge. Finally, immigrants between 15 and 66 years old from outside the European Economic Area (EEA) that do no migrate as refugees are expected to enroll without free training.

Table 3.1: Summary of right and obligation to participate

Type of immigrants	Free training	Expectancy of participation	Extent of training
Persons escaping severe social or human conditions and their families, 16-55 years old	x	x	600 hours (550 hours Norwegian language training and 50 hours social studies), additional 2400 hours if necessary
Persons escaping severe social or human conditions and their families, 55-67 years old	x		600 hours (550 hours Norwegian language training and 50 hours social studies), additional 2400 hours if necessary
Non-refugees outside EEA regulations, 16-55 years old		x	300 hours (250 hours Norwegian language training and 50 hours social studies)

In principle, participation in the training is voluntary. However, in order to attain a permanent residence or Norwegian citizenship, the immigrant groups described above must complete the training. On the other hand, immigrants and their families migrating in accordance with EEA regulations are not covered by the *introduction act*. Hence, they are not required to complete the training to obtain permanent residence, still they remain free to participate if willing.[6]

Contents and Duration of the Training

The minimum duration of the training is 600 hours, the equivalent of four months of full-time work. The exception is training for labour immigrants from outside the EEA, which amounts to a minimum of 300 hours. Chapter 1 stated that the Norwegian government emphasises proficiency in Norwegian as a key factor for the integration of immigrants. Consequently, schooling in the Norwegian language is the major component. Language training amount to 550 hours, whereas 50 hours are used on training in social studies. Social studies is to be done early in the course of training in a language the participant understands. It is important to note that the participant can apply for further training if needed, amounting up to a maximum of 3000 training hours.[7]

[6] Due to cultural similarities, it would be reasonable to assume that western countries such as USA, Canada, Australia and New Zealand applies to the same category as the EEA. The Directorate of Integration and Diversity (IMDi), however, rejects such an assumption (stated by IMDi per email 2018-03-26).

[7] Hence, individuals may receive different number of training hours which is unobserved in the data. Although this does not pose an identification threat in the analysis as it is carried out at the municipal level, one would not capture marginal effects of the training.

4. Data

The analysis exploits a strongly balanced panel data set at the municipal level without completely missing values for the dependent and key independent variable in any given municipality. Unavailable data before 2014 restricts the analysis to the period 2014-2016.[8] The number of municipalities is 370 and total observations amount to 1110.

The study is based on population registers from the Directorate of Integration and Diversity (IMDi), the Norwegian Directorate of Immigration (UDI), the Norwegian Centre for Research Data (NSD), the Norwegian government, and Statistics Norway (SSB). When presenting the variables, their specific source is stated. All sources provide publicly available data online with the exception of NSD that requires an application. SSB reports annual data for the dependent and key independent variable. Consequently, variables given on a monthly or quarterly basis is aggregated to yearly data. In order to use socioeconomic control variables, municipal data from SSB on immigrant employment and participation must be merged with socioeconomic data for the same municipalities. As a result, the data set has been constructed through a time-demanding process of connecting each variable with its corresponding municipality.

4.1 Dependent Variable

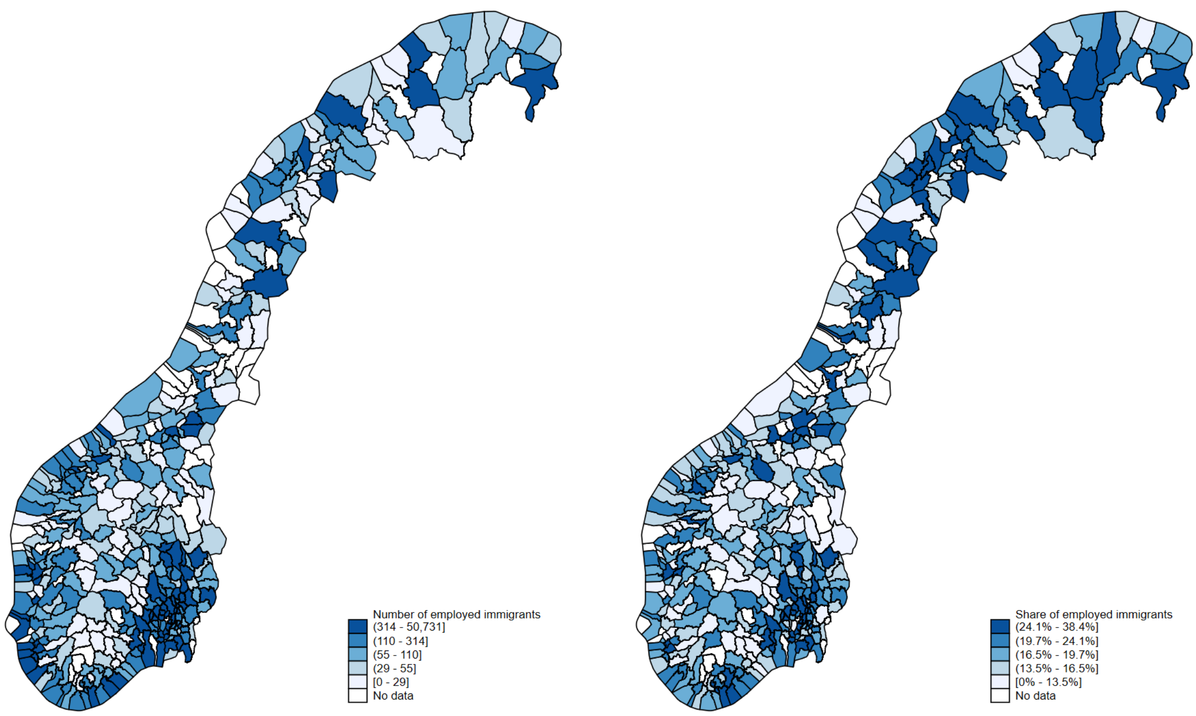
The dependent variable is the number of employed immigrants eligible for participation in Norwegian language training and social studies, in which eligibility is defined as being covered by the *introduction act*.[9] References to immigrant employment later in the thesis indicate therefore employment for immigrants eligible for the training. The empiric variable is an approximate match for the theoretical variable for two reasons. First, data is given for

[8] Previously, SSB merged data for newly arrived immigrants and asylum seekers that both participated in the Norwegian language training and social studies. As of 2014, however, SSB records the two groups separately. Considering that SSB does not define asylum seekers as immigrants, data before 2014 is inadequate for the research question of this analysis.

[9] Consequently, immigrants facing right and/or obligation of participation are considered eligible. However, note that the immigrant does not need to be enrolled in spite of being eligible and vice versa.

immigrants between 15 and 74 years old instead of 16-67.[10] Second, data excludes certain countries of origin that satisfy eligibility.[11] An immigrant is defined as a person born abroad by foreign parents. SSB further defines employment as paid work for at least one hour in the reference week, that is the week of reporting.[12]

Figure 4.1 maps the median number of employed immigrants and the median share of employed immigrants of the total immigrant population in each municipality between 2014-2016. The number of employed immigrants tends to be greater in coastal areas and the Greater Oslo Region in particular, which often are more urban than inland regions. A similar pattern for the share of employed immigrants suggests that urban areas employ immigrants eligible for participation to a greater extent of the total immigrant population, relative to rural areas.



(a) Number of employed immigrants

(b) Share of employed immigrants

Figure 4.1: Maps of employed immigrants. Shapefiles extracted from Geonorge includes municipalities' maritime boundaries.

[10] Data for the variable, extracted from SSB, contains information on two different age groups. The first group encompasses people aged between 15 and 74, whereas the second group covers people from 20 to 66 years of age. The analysis considers the group of immigrants between 15 and 74 years of age. The choice is based on arguments provided in appendix B.

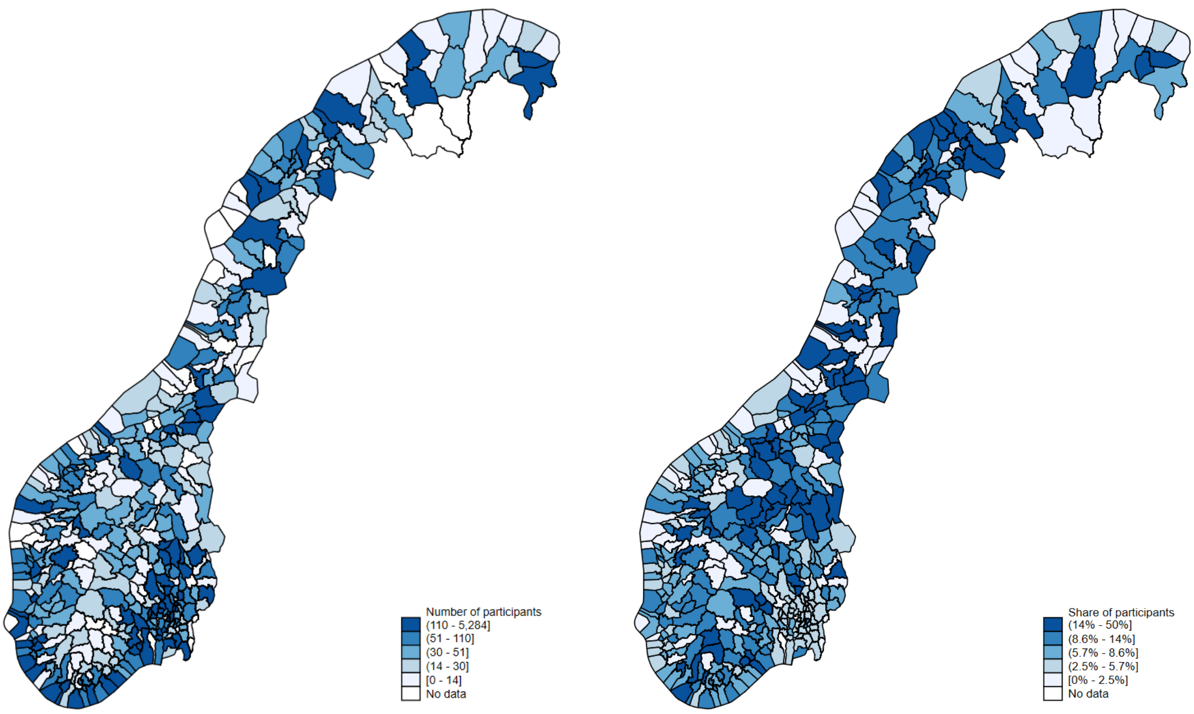
[11] SSB data contains two regions of origin. The first group comprises the EEA, North-America, Australia and New Zealand. The second group contains the rest of the world. The analysis excludes the first group due to arguments presented in appendix B.

[12] Stated by SSB per email 2018-03-19.

4.2 Key Independent Variable

The key independent variable is the number of participants in Norwegian language training and social studies for the years 2014-2016. Similar to the employment of immigrants, data is provided by SSB. However, SSB (2017-11-23) informs that figures are not shown where there are fewer than four units. Thus, it should be noted that municipalities with one, two or three participants are registered with missing values.

Figure 4.2 maps the median number of participants and the median share of participants in training of the total immigrant population in each municipality between 2014 and 2016. As expected and analogous to the employment of immigrants, the number of participants in training apparently augments with the urban status of a municipality. On the contrary, it appears that such municipalities exhibit a relatively low share of participants of the total immigrant population compared to rural municipalities.



(a) Number of participants

(b) Share of participants

Figure 4.2: Maps of participants. Shapefiles extracted from Geonorge includes municipalities' maritime boundaries.

4.3 Control Variables

The benchmark model employs several control variables to explain changes in the number of employed immigrants. The following section briefly presents control variables. The selected control variables are justified in Chapter 5.

Total Immigrant Population

The explanatory variable *totimmpop* controls for the total immigrant population in the municipality. This population is assumed to be the sum of eligible and non-eligible immigrants, hence the variable controls for both groups. The data source is SSB and the definition of an immigrant is thus equal to the one given in Section 4.1.

Production Index

The Norwegian government provides data on a production index, *prodindex*, controlling for the quality of local public services in any given municipality. The index is comprised of data on kindergarten, elementary school, child care, cultural activities, health care, primary care and social services. To distinguish the quality of services among municipalities, the index contains each municipality's annual percentage deviation from the national mean. A positive deviation indicates superior public services relative to the mean, while the opposite is the case for a negative deviation. Certain municipalities lack data for all years in question, while others lack data for specific years.[13]

Income

The variable *income* contains the annual median household income for the total population in a municipality after tax. Data extracted from SSB does not take into account the number of people in a household. Furthermore, numbers are not adjusted for inflation. In order to have a basis of comparison, the Consumer Price Index is used to control for inflation with 2014 as the basis year.

[13] The Ministry of Local Government and Modernisation stated per mail 2018-04-03 that the reason is missing information concerning learning environment in elementary schools. Data properties, however, do not allow for setting missing values to zero and controlling for elementary school quality. In total, missing values make up about 10% of the variable sample.

Municipal Expenses Per Participant

The variable *expenses* denotes annual municipal gross expenses per participant related to arranging the *introduction programme*. The variable is a proxy for municipal expenses related to the Norwegian language training and social studies. Municipal aggregated numbers from SSB are divided by the municipal number of participants. Furthermore, units are given per 1000 NOK and transformed to per 1 NOK. Negative values make no sense in terms of expenses and such values are therefore omitted.

4.4 Dummy Variables

In order to capture qualitative factors, the benchmark model includes binary variables commonly called dummy variables. The variable has value equal to one if the event is taking place and zero otherwise.

Urban Area

The variable *urban* indicates Norway's four largest urban areas: Oslo, Bergen, Stavanger/Sandnes and Trondheim. These four areas are selected on the basis of the substantial population differential to the fifth largest urban area, Drammen. Table B.1 in appendix B provides an overview over the municipalities included in the urban areas as well as the population from 2014 to 2016. SSB defines an urban area as an area with more than 200 inhabitants with less than 50 metres between the buildings (SSB, 2017-12-19). For the years in question, the geographical borders of the urban areas remained unchanged.

Political Party in Local Governance

The political spectrum in Norway is generally grouped in three: left, central and right. Furthermore, at the municipality level, voters may elect local coalitions called local lists and common lists. The analysis employs a set of three dummy variables. The variable *left* takes the value one for parties on the left side of the political spectrum, while the variable *right* equals one for the parties on the right. Furthermore, the third variable, *list*, takes the value one to control for municipalities registered with local lists and common lists. Hence, the central parties are regarded as the reference category equal to zero in all three dummies. Figure B.2 in appendix B display an overview of local parties in governance in the period 2014-2016.[14] A party is defined

[14] The municipal election in 2015 induced a turnover for some municipalities.

as in governance if the municipality's mayor belongs to the party.

4.5 Descriptive Statistics

Table 4.1 presents descriptive statistics for the dependent variable, explanatory variables and dummy variables. On average, 471 immigrants are employed in each municipality, ranging from five in Fyresdal (2014) and Engerdal (2015) to 51,949 in Oslo (2016). This means that on average, almost every fourth person of the total immigrant population is an employed immigrant who is eligible for the training. The average municipality has an immigrant population of 1,772 and 105 participants enrolled in the training. The average participation rate of immigrants is thus 5.9%.

Table 4.1: Descriptive statistics of the variables.

Variable	Obs	Mean	Std. Dev.	Min	Max
employed immigrants	1,088	470.873	2,797.380	5.000	51,949.000
participants	1,074	104.911	327.639	4.000	5,696.000
total immigrant population	1,110	1,771.698	8,741.757	75.000	163,348.000
prodindex	1,020	3.269	9.123	-18.600	53.000
income	1,110	478,975.600	49,697.000	338,000.000	643,000.000
expenses	935	160,467.600	79,567.840	250.000	1,118,833.000
urban	1,110	0.046	0.209	0.000	1.000
left	1,110	0.463	0.499	0.000	1.000
right	1,110	0.238	0.426	0.000	1.000
list	1,110	0.042	0.201	0.000	1.000

Figure 4.3 box plots the distribution of the population variables. The plot displays four quartiles for each variable per year, each part reflecting 25% of the variable's observations. The box thus represents the two middle quartiles, from the 25th percentile to the 75th percentile. Observations outside the box and tails (whiskers) are called outliers.

All variables reveal a constant nature indicating little time variation within units. On the other hand, the median is visibly lower than the reported means for each variable which is due to outliers above the top whiskers. These outliers, with Oslo as the most extreme case, augment the mean to the extent that most observations are below the mean. The median is therefore below the mean in most of the rectangles, which shows lopsided data that is skewed slightly upwards. The skew is further highlighted by standard deviations that are substantially higher

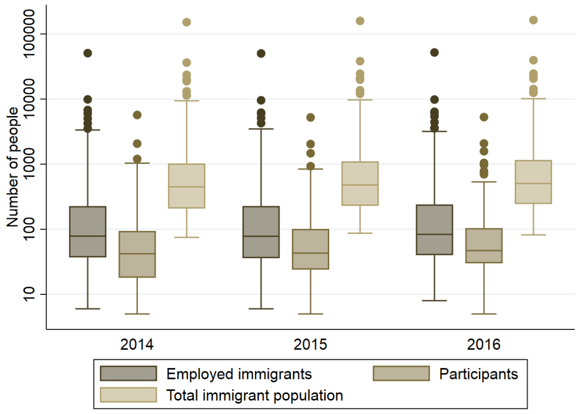


Figure 4.3: Distribution of population variables across 2014-2016.

than the mean.

On average, the median income for a household is NOK 478,976, while the average municipality spends NOK 160,468 per participant in the *introduction programme*. The box plot in Figure 4.4 shows positive outliers for median household income during all years, as well as one negative outlier in 2014. Hence, the skew is marginally upwards throughout the period. On the other hand, municipal expenses are skewed upwards as a result of several outliers.

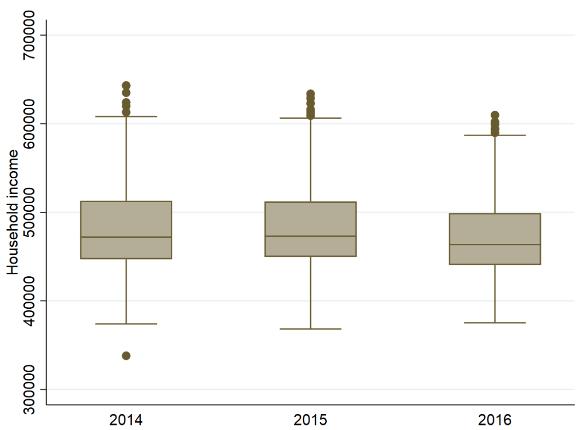


Figure 4.4: Distribution of median household income across 2014-2016.

The average deviation from the national mean in the production index is 3.24% which implies some extreme values. This is confirmed by a maximum observation as high as 53% higher than the national mean. On the contrary, the strongest negative deviation is 18.6%. Figure 4.5 illustrates this and reveals several outliers and a marginal upward skew.

The proxy variable for municipal expenses related to the training displays great variation

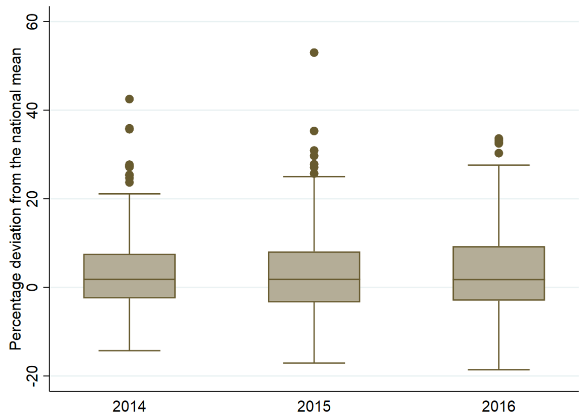


Figure 4.5: Distribution of the production index across 2014-2016.

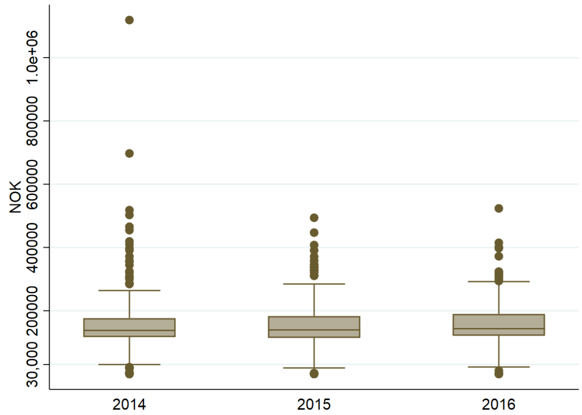


Figure 4.6: Distribution of municipal expenses per participant in programme across 2014-2016.

between municipalities. The reported maximum value is reported as almost seven times higher than the mean value, while the minimum value is as low as 250 NOK. Figure 4.6 displays this and also reveals that the median municipal expenses are lower than the average municipal expenses for all years.

Urban areas make up roughly 5% of the municipalities. There are twice as many left parties as right parties in municipal governance, while list is represented in 4.5% of the municipalities. About 25% of the municipalities are governed by a central party.

Having described the data foundation of the analysis, the thesis proceeds in the next chapter to develop an empiric model that address potential endogeneity in the key independent variable.

5. Model Identification

The analysis aims at estimating a causal effect of the number of participants in Norwegian language training and social studies on the number of employed immigrants. Assuming no self-selection, a naive model is written as follows

$$\begin{gathered}\ln(employment)_{it} = \beta_0 + \beta_1 \ln(participants)_{it} + \alpha_t + u_{it} \\ i = municipality, \quad t = time\end{gathered} \tag{5.1}$$

where $\ln(employment)_{it}$ denotes the logarithm of the number of employed immigrants in municipality i at time t and is the dependent variable in the model. The key independent variable, $\ln(participants)_{it}$, is the logarithm of the number of participants in Norwegian language training and social studies in municipality i at time t. Year fixed effects that capture shocks that may occur on a national level are given as α_t. Technological advances or national policies are examples of such factors that may vary over time and affect immigrant employment at the same time as they are common across municipalities. Lastly, u_{it} is the error term.

Borjas (1994, p. 1684-1685) states that the choice of acquiring proficiency in a foreign language may be endogenous, resulting in inconsistent and biased estimators. Hence, if the underlying assumption of no self-selection by immigrants fails to hold, ordinary least squares (OLS) regression on the naive model produces an inconsistent and biased parameter of interest.

The first assumption in order to have exogeneity in the key independent variable is no self-selection into participation in the training. Such a decision may be correlated with unobserved heterogeneity. For instance, immigrants who are motivated to learn the Norwegian language are likely to make a greater effort during the training and may also be more apt for employment. Thus, there may be characteristic differences between immigrants choosing to enroll in training and immigrants who choose otherwise. Dustmann & Fabbri (2003, p. 698) point out that if such dissimilarities affect employment, the estimated effect of language training might be upward biased.

However, self-selection into participation in the context of this thesis should not matter for several reasons. First, completing the training is required for obtaining permanent residence or a Norwegian citizenship. This rules out self-selection by refugees, and families reuniting with refugees. Second, other immigrants outside EEA regulations and their families planning a longer stay in Norway, will for the same reason have no choice but to undertake the training. This concerns persons stating education or labour as the reason for migration. In principle, however, education and seasonal immigrants are free to choose participation. The intuition here is that international students and seasonal workers tend to reside in Norway for shorter periods of time and are therefore not required to attend the training because they do not apply for a permanent residence.[15] Immigration data from SSB, presented in Table B.1 in appendix B, shows that education and seasonal immigrants account for a majority of the non-refugee immigrants outside EEA regulations. Hence, it is plausible to assume that in practice roughly 60% of such immigrants have a choice of participation. However, it could be argued that education and seasonal immigrants are unlikely to participate. First, programmes taught in English is a major reason for studying in Norway (Jannecke, 2013, p. 21). Second, seasonal jobs, such as within agriculture, are often manual labour intensive and do not require high proficiency in the Norwegian language. The limited need of Norwegian language skills may thus serve as an incentive not to undertake the training.[16]

A third reason for making the case against self-selection into participation is the tendency of immigrants from EEA countries not to enroll in the training. Although the *introduction act* does not cover immigrants arriving according to EEA regulations, they do have the opportunity to participate. However, immigrants from EEA countries make up an insignificant share of participants in training.[17] Moreover, the number of EEA immigrants in training is also a negligible share of the total number of immigrants from this area. This group is in other words inclined not to participate. On the basis of this tendency and the aforementioned arguments, self-selection into participation appears in practice to be exogenous for all immigrant groups.

[15] A permanent residence is required for stays longer than three years.

[16] Statistically, it is hard to argue in any direction due to limited data. The argument is therefore based on reasoning.

[17] See Table B.4 in appendix B for information concerning these numbers.

The analysis assumes this to hold.[18][19]

The second assumption in order to have exogeneity in the key independent variable is no self-selection into certain municipalities subject to local training and labour market conditions. For instance, an immigrant may be prone to settle down in municipalities with a good reputation for immigrant employment or municipalities with a pre-existing immigrant community. If this is the case, endogenous sorting might lead to inconsistent and biased estimators. During the years in question refugees are allocated exogenously. This is done by the IMDi in collaboration with municipal authorities.[20] Furthermore, refugees is the only immigrant type that is allocated exogenously.[21] Considering that a significant share of immigrants therefore potentially self-select, it is reasonable to assume the existence of selective immigrant sorting into certain municipalities.[22]

Self-selection conditional on local circumstances results in endogeneity in the naive model. In order to attain consistent and unbiased results, the naive model first extends by employing the logarithm of the lagged number of participants in training in municipality i at time t-1. This is the key independent variable and given as $\ln(participants)_{it-1}$. Lagging the variable is useful in strengthening the assumption of exogeneity.[23] Furthermore, enrollment in the training prevents immigrants from working in the short term. The lagged variable will capture the effect in a longer term. Second, the model incorporates several regressors controlling for self-selection by immigrants. Thus, a basic OLS model is given as

[18] If education and seasonal immigrants and immigrants from the EEA actually self-select into participation, the assumption of exogenous self-selection is violated. Figure B.1 in appendix B shows that only a minor part of the immigrants would then be suspect to exogenous self-selection.

[19] It should be noted that the government is able to modify eligibility requirements as it observes the effect of participation on employment. Nonetheless, implementing such reforms is time demanding and will not be a problem within the context of this thesis.

[20] To be sure, refugees may choose to allocate themselves. In such cases, the municipality needs to validate the immigrant's choice. Ministry of Justice and Public Security (2015-2016) states nonetheless that more than 90% of refugees are allocated through an agreement between IMDi and the municipalities.

[21] Figure B.1 in appendix B shows refugees to make up on average 18.3% of total immigration between 2014-2016. The estimate is increasing, from 12% in 2014 to 26% in 2016.

[22] This conclusion presupposes rational agents, i.e. that self-allocation takes into account local conditions in the municipality.

[23] There might be a reverse causality between the dependent and the key independent variable in which lower employment results in a higher propensity of participation.

$$\begin{aligned}\ln(employment)_{it} = \beta_0 + \beta_1 \ln(participants)_{it-1} + \beta_2 \ln(totimmpop)_{it} \\ + \beta_3 \ln(prodindex)_{it} + \beta_4 \ln(income)_{it} + \beta_5 \ln(expenses + 1)_{it} \\ + \delta_1 urban + \delta_2 left + \delta_3 right + \delta_4 list + \alpha_t + u_{it}\end{aligned} \tag{5.2}$$

To address possible sorting subject to labour conditions, the model includes control variables for the total immigrant population and the median household income level in municipality i at time t. Both variables are given in logarithmic form, $\ln(totimmpop)_{it}$ and $\ln(income)_{it}$ respectively. In order to estimate a causal effect of participation in the training on employment it is necessary to control for the total immigrant population, as the employment of immigrants is likely to increase with the immigrant population.[24][25][26] Additionally, including the total immigrant population in the municipality enables the model to control for immigrant sorting into municipalities contingent on pre-existing immigrant communities. Empirical evidence is somewhat mixed on the effect of ethnic enclaves. Blom (2012, p. 33) shows that immigrants in Norway prefer not to live in ethnic enclaves, whereas other studies suggest that immigrants tend to self-select into ethnic enclaves (Damm (2009); Edin et al. (2003)). The latter studies also find indications of increased earnings due to increased ethnic enclave. Hence, one may argue that larger ethnic enclaves may attract immigrants through an earnings channel. Moreover, it is worth noting the possibility of a two-way causality between total immigrant population and the number of employed immigrants. A high employment rate of immigrants might serve as a pull-factor resulting in more people immigrating to Norway in pursue of work. Consequently, the total immigrant population increases. Nonetheless, even if $\ln(totimmpop)$ might suffer from simultaneity, it is not decisive for the estimated effect of the main independent variable. On the basis of that argument, in combination with little within

[24] The data excludes asylum seekers and persons with a residence permit for less than six months. Furthermore, by disregarding persons without legal residence in Norway, the data yields a potential underestimation of the real number of immigrants. The number of unrecorded cases might be large as The Norwegian Directorate of Immigration (2014, p. 11) estimated the amount of illegal immigrants in Norway to be between 18 100 and 56 000 in 2014.

[25] An alternative variable could be the labour force of immigrants. However, the variable is the total immigrant population because the Norwegian language training and social studies not only aims to improve the labour market performance of immigrants, but also their labour market participation. As a result, one would anticipate the labour force of immigrants to increase as more people enroll in the training, resulting in correlated explanatory variables.

[26] To be sure, theory asserts that previous immigrants bear the competitive pressure resulting from immigration as they are often considered perfect substitutes (Bansak et al., 2015, p. 155). However, the positive relationship is assumed to be little affected by this substitutability.

variation as shown in Table B.5 in appendix B, simultaneity issues in $\ln(totimmpop)$ is disregarded.

Furthermore, controlling for the median household income after tax is useful when addressing labour conditional sorting. Evidence from Europe indicates that earnings may serve as a pull factor for immigrants.Ortega & Peri (2012) This is especially the case in areas with few mobility barriers which is the case in this thesis. Immigrants may allocate in hope of higher earnings and high level income municipalities may enjoy a superior status.[27] The income variable captures such sorting.

Another concern is self-selection on the basis of non-labour conditions such as welfare and opportunities. First, an individual's propensity to sort into a municipality may increase if that municipality offers services that are relatively superior compared to that of other municipalities. Several studies show for instance that people are drawn to areas with high quality schools (among others, see Black (1999, p. 595); Kane et al. (2003, p. 135)). In Scandinavia, Büchel & Frick (2005, p. 206) and Hansen & Lofstrom (2003, p. 97) provide evidence that immigrants tend to intensively consume public goods. Borjas (1999) introduces the welfare magnet hypothesis stating that high level welfare may serve as a pull factor for migration.[28] A production index, $prodindex_{it}$, is used to control for local public services that may impact selective immigrant sorting.[29] Second, the municipalities tend to arrange the training differently (Djuve et al., 2017, p. 274).[30] Immigrants might therefore sort into municipalities with high costs per participant in the training as this might be an indicator of training quality. Albeit non-existing data concerning the Norwegian language training and social studies, SSB conducts data on expenses for the *introduction programme*. Wooldridge (2016, p. 279) states that "a proxy variable is something that is related to the unobserved variable that we would like to control for in our analysis". In other words, the approach necessitates a correlation between the proxy and the unobserved variable. Since the introduction programme encompasses the Norwegian language training and social studies, the

[27] High household income increases municipal revenue through taxes which serves as an argument of not explicitly including municipal revenue.

[28] Albeit, De Giorgi & Pellizzari (2009) find a significant, yet small, effect of welfare generosity on migration decisions in Europe, empirical evidence for this hypothesis is somewhat mixed Giulietti & Wahba (2012).

[29] There is a significant correlation between a municipality's public services and its economical constraints (Ministry of Local Government and Modernisation, 2017, p. 49). If the municipality's revenues are high, the local government has financial freedom to invest in public services. By employing the production index as a control, the model takes municipal revenue into account.

[30] See appendix A for information concerning the responsibility of municipalities in arranging the training.

programme expenses serves as an appropriate proxy and is given as $\ln(expenses + 1)_{it}$. In addition, the proxy variable downplays the issue regarding omitted variable bias.

Furthermore, $urban$, $left$, $right$ and $list$ are qualitative urban or political dummies, as explained in Section 4.4. The urban dummy captures the tendency of immigrants to settle down in urban areas to a greater extent than the native population, particularly around the largest cities (NOU 2011:14, 2011, p.354). Moreover, immigrants may hold different attitudes towards the parties which in turn may cause self-selection into certain municipalities and bias. Additionally, it is interesting to analyse the impact of training across urban and rural areas, and across political parties.

Section 10.1 challenges the assumptions of no self-selection by means of an instrumental variable. However, the assumptions made in this chapter are underlined by similar estimates obtained in the instrumental variable approach and the main analysis.

Wooldridge (2016, p. 171-173) points out that using variables on logarithmic form enables analysing elasticities and semi-elasticities, in addition to disregarding different unit measurements in the variables. Consequently, variables take on logarithmic form where fitting. One limitation is that logarithmic form is not possible whenever a variable displays negative values or values equal zero. However, Wooldridge (2016, p. 173) proposes that in cases with non-negative values and relatively few zeros, one can transform the variable on logarithmic form by using $\ln(var + 1)$.[31]

The dependent variable and the key independent variable contain neither negative nor zero values. Thus, the model includes both variables in normal logarithmic form. The interpreted effect is elastic, measuring the percentage increase in employment of immigrants when the number of participants increases by 1%. Likewise, the total immigrant population is given in logarithmic form yielding the same percentage interpretation.

The production index is given in level form. The reason is that the index allows for positive and negative deviations from the national mean (the mean is equal zero). The interpreted effect is therefore semi-elastic, indicating the percentage change in employment of immigrants of a unit increase in the production index, when multiplying the coefficient with 100.

Median household income and the proxy for municipal training expenses per participant

[31] The percentage change interpretation remain similar to normal logarithmic form, $\ln(var)$. On a technical note, $\ln(var + 1)$ cannot be normally distributed. As a result, inference from t and F statistics is not exact. Nonetheless, OLS estimators satisfy asymptotic normality, indicating an approximated normal distribution (Wooldridge, 2016, p. 155). This is the case throughout the analysis.

exhibit positive monetary values, frequently of grand magnitudes. Logarithmic form narrows the variables' range, reducing their sensitivity to extreme values. Additionally, this solves for potential heteroskedasticity or skew in $\ln(income)$ as this variable is strictly positive. The proxy variable, however, displays observations with the value zero and is therefore included in the alternative logarithmic form, $\ln(expenses + 1)$, as proposed earlier.

The chapter presented and discussed a model that corrects for endogeneity concerns and ensures a causal interpretation of participation. The following chapter presents and comments estimates on this causality, and other estimates, obtained with an OLS regression.

6. Basic OLS Analysis

The first part of this chapter presents the basic analysis of an OLS regression on model 5.2 introduced in the previous chapter.[32] The second part examines the validity of the results, focusing on heteroskedasticity and serial correlation.

6.1 Results OLS

Table 6.1 presents seven specifications that aim at assessing a causal relationship between participation in the Norwegian language training and social studies and the employment of immigrants.

The dependent variable remains $\ln(employment)$ throughout the analysis and $\ln(participants)$ is the key independent variable. The first column includes the key independent variable and year fixed effects. Subsequently, the specifications extend with socioeconomic variables. Column (5) adds all control variables, whereas column (6) further adds dummy variables. All regressions are performed in the statistical software Stata.

The specification in column (1) is considered to be naive and should not be given much attention. Including only the number of participants yields a mechanic relationship with immigrant employment, in which the total immigrant population is not controlled for. In this case, the immigrant population in a municipality is assumed not to be correlated with the number of participants in training or the number of employed immigrants.

The magnitude and significance of participation remain robust across specifications. Estimates vary from 0.174 in column (6), to 0.189 in column (4). Given the elasticity interpretation between the dependent and key independent variable, the specifications estimate that a 10% increase in the number of participants increases the number of employed immigrants by between 1.74% and 1.89%. The estimates seem to be precise as they remain significant at the 1% level through all specifications. In spite of high correlation between the number of participants and the total immigrant population, as shown in Table B.3 in appendix B, standard errors remain constant at a low value.

[32] A discussion on panel data and OLS is provided in appendix C.

Similar to the number of participants in training, the socioeconomic variables show robust estimates across every specification. As anticipated, the total immigrant population in a municipality appears to have a positive effect on the number of employed immigrants. The effect is significant at the 1% level in all columns and reveals a seemingly one-to-one relationship between participation and employment, which may be a result of the high correlation value mentioned earlier. Another explanation may be the reverse causality between immigrant employment and the total immigrant population as discussed in Chapter 5. Estimates for the production index and income are positive throughout specifications, yet imprecise due to low significance. On the other hand, municipal expenses is significant at the 1% level and positive. A 10% increase in expenses is expected to increase immigrant employment by 0.26% in column (6).

The dummy coefficients measure the average difference in immigrant employment immigrants between the two groups in question, such as urban and rural areas, that have the same levels of participants, immigrant population, production index, income and expenses. Estimates indicate 6.6% higher immigrant employment in municipalities within Norway's four largest urban areas than in the remaining municipalities.[33] A conclusion can nevertheless not be drawn due to low significance. Among the political dummy variables, political parties on the left wing and parties created with a common list exhibit significant estimates at the 1% and 5% level respectively. A municipality governed by a left party is estimated to have 7.79% higher immigrant employment than municipalities governed by a central party, while the difference between municipalities with governed by a list or central party is predicted to be 15.3%.

6.2 Model Validity

This section examines the validity of the OLS regression using two diagnostic tests. Rejection of the test hypotheses indicates that the basic model estimated by OLS possess

[33] All tables in the thesis report dummy coefficients that illustrate the difference between two qualitative groups. These estimates, however, are only logarithmic approximations that do not point out which group is the base group, that is the group against which comparisons are made. In other words, the estimates simply explain the difference between the two groups, which is 6.4% in the urban dummy case. Wooldridge (2016, p. 212) proposes a formula that generates accurate estimates according to the chosen base group. Hence, $100 * [exp(\hat{\beta_1}) - 1]$ provides the accurate estimate that urban areas have 6.6% higher effect than rural areas. On the other hand, $100 * [exp(-\hat{\beta_1}) - 1]$, gives the accurate estimate that rural areas have 6.2% lower effect than urban areas.

Table 6.1: Results basic analysis: OLS

Variable	(1) ln(employment)	(2) ln(employment)	(3) ln(employment)	(4) ln(employment)	(5) ln(employment)	(6) ln(employment)
ln(participants)$_{-1}$	1.095***	0.180***	0.179***	0.189***	0.179***	0.174***
	(0.025)	(0.019)	(0.019)	(0.020)	(0.020)	(0.020)
ln(totimmpop)		1.000***	1.000***	0.990***	0.996***	0.996***
		(0.018)	(0.019)	(0.020)	(0.020)	(0.020)
prodindex			0.0003	0.001	0.001	0.001
			(0.001)	(0.002)	(0.002)	(0.001)
ln(income)				0.217*	0.175	0.182
				(0.126)	(0.125)	(0.131)
ln(expenses+1)					0.023***	0.026***
					(0.008)	(0.008)
urban						0.064
						(0.054)
left						0.075***
						(0.027)
right						0.035
						(0.034)
list						0.153**
						(0.060)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year 2016	-0.003	0.022	0.021	0.026	0.024	0.024
	(0.050)	(0.022)	(0.021)	(0.022)	(0.022)	(0.022)
Constant	0.575***	-2.414***	-2.421***	-5.221***	-4.952***	-5.063***
	(0.109)	(0.068)	(0.074)	(1.626)	(1.619)	(1.699)
Observations	601	601	601	601	601	601
R^2	0.747	0.960	0.961	0.961	0.963	0.964

Note: Each column represents a separate specification. All regressions include year specific effects. The number of employed immigrants is the dependent variable in all specifications, while the number of participants at time $t-1$ is the key independent variable. Standard errors in parenthesis. *Significant at 10%; **significant at 5%; ***significant at 1%.

statistical weaknesses that should be dealt with in order to attain consistent, unbiased and efficient estimators.

6.2.1 Identifying Heteroskedasticity

Heteroskedasticity implies differing variance across municipalities. As an illustration, assume that the number of private firms within a municipality decreases with superior public services. The error term captures the associated effects because the number of firms also affect the immigrant employment rate. Consequently, the error term variance decreases with the production index, implying biased variance estimators and standard errors.[34]

Results of a general White test is presented in Table C.1 in appendix C. The null hypothesis of homoskedasticity is rejected in specification (2). However, the White test appears to state homoskedasticity in the other specifications. The change in the test outcome might indicate that adding control and dummy variables has a moderating effect on heteroskedasticity in the model. As a result, heteroskedasticity is not an issue.

6.2.2 Identifying Serial Correlation

Serial correlation occurs when residuals are correlated across time and causes misguided inference in which one could be deluded to assert insignificant estimates as significant. Possible serial correlation in data may result from a not randomly drawn sample. The reason is that shocks occurring on a national level presumably affect the municipalities inducing cross-sectional correlation.

Results of a general Arellano-Bond test with one lag is presented in Table C.1 in appendix C. The null hypothesis of no serial correlation is strongly rejected and seems to provide evidence of serial correlation that generates biased results.

The above finding suggests that OLS is not the appropriate estimation method to ascertain the effect of participation in training on immigrant employment. Consequently, the next chapter develops a more advanced framework for the main analysis.

[34] Note that this is merely an example, not necessarily reality.

7. Econometric Framework Main Analysis

The previous chapter concluded with serial correlation in the OLS regression. Furthermore, participation in training may be correlated with unobserved time-invariant characteristics which may induce bias. This chapter presents the framework of the main analysis, which comprises random and fixed effects as well as cluster-robust standard errors. Appendix C provides a thorough discussion on these methods.

Clustering

The main analysis groups municipalities as clusters of observations that vary over time without affecting estimated coefficients. Thus, the approach allows for serial correlation within the clusters and obtains robust inference.[35][36]

Random Effects

The fundamental assumption in random effects estimation is that there is no correlation between any unobserved effects and each of the explanatory variables in the model, which is $Cov(x_{itj}, \eta_i) = 0$ for all i, t and j. This means that random effects is not threatened by the presence of serial correlation. Randoms effects is therefore more efficient than OLS. The reason for this is that random effects transformation, often referred to as generalised least squares (GLS), eliminates the serial correlation, even in cases where the serial correlation is unobserved. Hence, implementing random effects will improve the precision of estimates since the assumption of no correlation holds whether the explanatory variables are constant over time or not. Although panel data requires certain variation within and between municipalities in the sample, the random effects approach will provide desirable estimator properties in large samples with few time periods, which are characteristics that fit well the data in this thesis. Table B.5 in appendix B shows that several variables in the sample exhibit very little within variation. The limited variation is explained by time-invariant variables of a constant nature combined with a short panel over three years. However, the GLS

[35] The analysis presumes identical cross-sectional dependence for every municipality. Hence, combining cluster-robust standard errors and year dummies purges spatial dependence (Cameron & Miller, 2015, p. 7).

[36] According to Cameron & Miller (2015, p. 342), the effective number of clusters in a balanced panel data set is at least 25, an unvexed issue considering the grand number of municipalities in the analysis.

transformation allows for constant explanatory variables and ensures efficient estimators.

Fixed Effects

If the unobserved effect is correlated with any explanatory variable, fixed effects is assumed to be a better approach than random effects. This method eliminates time-invariant unobserved effects by within groups transformation of the original equation. The transformed equation is then estimated with OLS. OLS will then produce consistent and unbiased estimators. Because of the elimination of variation between municipalities, fixed effects estimation requires a certain level of within variation.

The fixed effects approach chiefly counteracts self-selection into municipalities by eliminating unobserved heterogeneity between the municipalities. The underlying assumption is that such factors and associated effects are constant over time during the period of study Akresh (2007, p. 875). However, fixed effects also proves to be efficient against self-selection into participation and omitted variable bias. First, participation in training may be correlated with unobserved time-invariant characteristics. Fixed effects would then omit the effect of such time-invariant characteristics and makes it possible to assess the net effect of participation in the training on employment. Second, as opposed to several studies, this thesis excludes potentially relevant control variables. For instance, education level as used in Hayfron (2001) is not available at the municipality level. Tertiary production within the municipality is another example. The challenges are easier understood by decomposing the error term in a unit specific component and an idiosyncratic component as shown in appendix C. The unit specific component contains unobserved, yet relevant, factors that may only vary between units. It is possible that local conditions, such as tertiary production or climate, affect immigrants' selection into specific municipalities. If aforementioned municipality specific heterogeneity is not controlled for as a variable in the model, such effects will be captured by the error term and induce correlation with the included variables. Yao & van Ours (2015, p. 76) point out that excluding relevant variables may give an upward bias in the estimated effects of language skills. The fixed effects approach counteracts this problem by eliminating unobserved heterogeneity between the municipalities.

This chapter has concisely established the set-up for the main analysis. The ensuing chapter presents results of random and fixed effects estimations that omit issues concerning identification threats and serial correlation.

8. Results Main Analysis

This chapter presents results from the main analysis, containing random and fixed effects. Both approaches are extensions of the basic model introduced in Chapter 5. All model specifications aim to assess the causal effect of the number of participants in Norwegian language training and social studies on immigrant employment. Subsequent to the presentation of results, Section 8.3 compares findings to that of previous literature. The main analysis replicates all specifications from the basic analysis performed with OLS. All regressions are performed in Stata with cluster-robust standard errors.

8.1 Results Random Effects

Similar to the OLS regression executed in Section 6.1, random effects produces consistent and unbiased estimators in the case of exogenous explanatory variables. On the other hand, random effects is more efficient than OLS due to the elimination of any serial correlation by transforming the associated error component. Table 8.1 presents results from random effects estimation. The following discussion emphasises column (6).

Every specification predicts a positive, yet small, impact of participation on immigrant employment significant at the 1% level. Similar to the OLS results, the estimate is positive for all specifications. The first specification is estimated with the largest effect of 0.298. When accounting for socioeconomic conditions, estimates prove to be highly stable, ranging between 0.057 and 0.061. Columns (2)-(3) and (5)-(6) are estimated with a difference of only 0.001. Albeit Table B.3 in appendix B display a seemingly high correlation between the population variables, such stable results might indicate that there are no issues with multicollinearity. The highest estimate is found when controlling for all explanatory variables except expenses and dummies in column (4). The lowest estimate is, by contrast, found in the two last specifications that include all explanatory variables in column (5) and dummies in column (6). The latter column is the main specification and estimates an increment in immigrant employment by 0.57% when participation increases 10%. Including determinants of immigrant employment appears to have no particular controlling effect, as illustrated by the

Table 8.1: Results main analysis: random effects

Variable	(1) ln(employment)	(2) ln(employment)	(3) ln(employment)	(4) ln(employment)	(5) ln(employment)	(6) ln(employment)
ln(participants)$_{-1}$	0.298***	0.058***	0.058***	0.061***	0.057***	0.057***
	(0.040)	(0.021)	(0.020)	(0.022)	(0.022)	(0.022)
ln(totimmpop)		1.086***	1.083***	1.078***	1.080***	1.068***
		(0.020)	(0.020)	(0.023)	(0.023)	(0.024)
prodindex			-0.001	-0.001	-0.001	-0.001
			(0.002)	(0.002)	(0.002)	(0.002)
ln(income)				0.149	0.138	0.149
				(0.182)	(0.182)	(0.191)
ln(expenses+1)					0.012*	0.013*
					(0.007)	(0.007)
urban						0.088
						(0.063)
left						0.098**
						(0.043)
right						0.032
						(0.052)
list						0.148**
						(0.065)
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year 2016	0.056***	0.019***	0.019***	0.023***	0.022***	0.023***
	(0.008)	(0.007)	(0.007)	(0.008)	(0.008)	(0.008)
Constant	3.613***	-2.503***	-2.485***	-4.409**	-4.402*	-4.540*
	(0.143)	(0.088)	(0.087)	(2.330)	(2.317)	(2.450)
Observations	601	601	601	601	601	601
R^2 within	0.264	0.292	0.294	0.296	0.298	0.300

Note: Each column represents a separate specification. All regressions include year specific effects. The number of employed immigrants is the dependent variable in all specifications, while the number of participants at time $t-1$ is the key independent variable. Cluster-robust standard errors in parenthesis. *Significant at 10%; **significant at 5%; ***significant at 1%.

arbitrary estimate changes. This pattern is analogous to the OLS regression. Moreover, it is not evident that the GLS transformation has a moderating effect on precision, as the standard errors remain similar to the ones found by OLS estimation. Stable standard errors might indicate that the challenges related to serial correlation, as discussed in Section 6.2.2, do not pose a grand problem in the analysis.

Furthermore, total immigrant population is predicted to affect immigrant employment positively at the 1% significance level. Estimates suggest that an immigrant population growth of 10% raises the number of employed immigrants rises by 10.68%. Hence, the percentage change in immigrant employment is more-than-proportionate and can be said to be elastic. The coefficients are nonetheless alike to the OLS estimates and might reflect the two-way causality between immigrant employment and immigrant population. The proxy variable for municipal expenses is also estimated to be a driver of immigrant employment. As the municipality invests 10% more money in the training, the number of employed immigrants is estimated to increase 0.13% at the 10% significance level. The finding suggests a positive relationship between the quality of training and employment.[37]

Estimates of the two remaining socioeconomic variables, the production index and household income, are not significantly different from zero. Similar to the OLS analysis, $left$ and $list$ are the only significant dummies. The estimates predict that municipalities governed by a left or list political party respectively portray 10.30% and 15.95% higher immigrant employment than the central municipalities.

8.2 Results Fixed Effects

In contrast to OLS and random effects, fixed effects produces consistent and unbiased estimators in spite of correlation between explanatory variables and the error component. The within transformation eliminates all municipality specific effects as discussed in Chapter 7 and disregards the assumption of strictly exogenous variables. Table 8.2 presents fixed effects results.[38] The discussions emphasises column (5), which corresponds to column (6) in the discussion on OLS and random effects. The dummy variables are omitted from the analysis because they exhibit no variation over the three years in question. This does not challenge the

[37] The statement presumes that higher expenses is an indicator for training quality, as mentioned in Chapter 5.
[38] Any references to the main analysis later in the thesis points to the fixed effects parameter estimates.

current analysis, however, because by using fixed effects it is possible to control for the same characteristics as with the dummies.[39]

Fixed effects estimation predicts a positive relationship between the number of participants in the training and immigrant employment, all else equal. Adding socioeconomic variables moves point estimates from 0.041 to 0.068, but results remain economically and statistically similar. The specification in the first column portrays an effect of 0.41, which is the lowest parameter estimate among the specifications. The main specification in column (5), by contrast, displays an estimated increase of 0.68% in the number of employed immigrants for a 10% increase in participation. In effect, the magnitude is economically small, illustrated in terms of standard deviations. An increase of one standard deviation in participation brings forth an increase of 0.008 of a standard deviation in employment. The highest estimate is found in the fourth column, which excludes expenses from the model and predicts an upsurge of 0.71% in participation.

Albeit of small magnitude, estimates suggest that as more immigrants participate in the training, the number of employed immigrants eligible for training increases. By obtaining proficiency in the Norwegian language, immigrants are likely to be more productive at work at the same time as they are expected to integrate into the workplace environment to a greater extent. Furthermore, Norwegian language training and social studies simplifies the job searching process for the immigrant. However, since the effect is smaller than a one-to-one ratio, the relationship between participation in the training and immigrant employment is non-elastic. In other words, the number of employed immigrants increases proportionally less than the number of participants in the training. For equivalent levels on participation and employment, predictions are that roughly 7 immigrants becomes employed if 100 persons enroll in the training. Fixed effects yields lower estimates than OLS, hence confirming the assertion of Dustmann & Fabbri (2003) from Chapter 5 that OLS estimates may be upward biased. Furthermore, the estimates are approximately the same as random effects estimates. However, in the main specification, point estimates are larger with fixed effects than with random effects.

Adding new regressors causes standard errors to rise marginally from the first column to the second, but remain unchanged throughout the other columns. Although the naive model in

[39] The robustness analysis will however examine urban and political indicators.

Table 8.2: Results main analysis: fixed effects

Variable	(1) ln(employment)	(2) ln(employment)	(3) ln(employment)	(4) ln(employment)	(5) ln(employment)
ln(participants)$_{-1}$	0.041**	0.058**	0.058**	0.071***	0.068***
	(0.019)	(0.024)	(0.024)	(0.024)	(0.024)
ln(totimmpop)		-0.221	-0.218	-0.286	-0.304*
		(0.205)	(0.202)	(0.173)	(0.174)
prodindex			-0.0001	-0.0002	-0.00003
			(0.001)	(0.002)	(0.002)
ln(income)				1.106**	1.127**
				(0.494)	(0.490)
ln(expenses+1)					0.011***
					(0.004)
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Year 2016	0.081***	0.092***	0.092***	0.120***	0.120***
	(0.006)	(0.011)	(0.010)	(0.014)	(0.014)
Constant	4.705***	6.089***	6.075***	-8.017*	-8.300
	(0.074)	(1.287)	(1.276)	(6.681)	(6.640)
Observations	601	601	601	601	601
R^2 within	0.441	0.446	0.446	0.463	0.469

Note: Each column represents a separate specification. All regressions include year specific effects. The number of employed immigrants is the dependent variable in all specifications, while the number of participants at time $t-1$ is the key independent variable. Cluster-robust standard errors in parenthesis. *Significant at 10%; **significant at 5%; ***significant at 1%. Fixed effects estimation omits all dummy variables because of collinearity.

column (1) fails to control for socioeconomic conditions, the estimated effect of the training is not greatly different from the other specifications. Estimates are statistically distinguishable from zero at the 5% significance level in columns (1) to (3), whereas estimates in columns (4) and (5) are significant at the 1% level. In spite of eliminating time-invariant characteristics between municipalities and depending strongly on within variation, fixed effects estimation produces standard errors that are equal to those of random effects and OLS in size. Moreover, results remain stable and precisely estimated although the analysis exploits only two time periods and limited time variation within municipalities. This may suggest that there is enough variation in participation over time within municipalities.

Unlike participation in training, total immigrant population estimates display higher standard errors when the estimation method changes from random effects to fixed effects. Fixed effects estimation also causes estimates to be of low magnitude for most specifications and lower than predictions with OLS and random effects. In contrast to the anticipated positive impact of the total immigrant population discussed in Chapter 5, the main regression predicts a negative estimate significant at the 10% level. As mentioned in the discussion on the key independent variable estimates, the reduction in magnitude and precision may be a result of little within variation, which is illustrated in Table B.5 in appendix B. In fact, one would be right to expect little variation. Populations usually grow slowly, especially on a year-to-year basis, and the fixed effects omits time-invariant factors. Thus, estimates are not bewildering. Furthermore, correlation matrix B.3 displays a high correlation between the population variables. Although high correlation was not decisive for the stability of the key independent variable, it might be a driver behind the negative coefficient of the total immigrant population.

In the same way as the total immigrant population varies little over time, the production index should neither change much and estimates are negative. However, the predicted effect is of insignificant magnitude and low t-values invalidates any conclusion drawn from this variable.

On the other hand, variables that are more apt to vary across time maintain their strong impact. First, household income is estimated to have a positive effect on immigrant employment. The finding coincides well with the discussion of Chapter 5. Moreover, the main regression predicts a more-than-proportionate increase in employment of 11.27% for a 10% increment in income, which is significant at the 5% level. Second, the proxy variable for municipal expenses per participant is significant at the 1% level. A 10% rise in expenses increases the number of employed immigrants by 0.11%. Albeit small in magnitude, the result seems to affirm the hypothesis of a positive relationship between training expenses and employment, which might indicate higher training quality the higher expenses.

8.3 Results Compared to Previous Literature

This section discusses the above findings in light of relevant literature. It is important to bear in mind that the results may deviate from earlier studies for two major reasons. First, this thesis examines a unique data set in a different country than most previous studies, and second, the identification method and modelling are not necessarily equal to that of former literature. Comparison is done based on fixed effects results due to the convincing outcome of the Hausman test, which is presented in Section C in appendix C.

Chapter 2 discusses existing literature on language training and shows that the impact of training on labour market outcomes is not clear-cut. Closest to this study, Lochmann et al. (2018) find a negative effect on full-time employment,[40] similar to the negative yet insignificant effect found by Sarvimäki & Hämäläinen (2016). On the contrary, the findings of this thesis suggest a more optimistic view albeit the economic impact remains small. Thus, results are more in line with the positive finding of Sarvimäki & Hämäläinen (2016) on earnings. Akresh (2007) finds, on the other hand, no evidence of higher earnings. Language training is however often found to have a positive impact on language proficiency as shown by Beenstock (1996), Gonzalez (2000) and Hayfron (2001). Similar to this thesis, Hayfron (2001) examines immigrants in Norway but is not able to conclude with a significant relationship between language skills and earnings.

Several reasons could explain the differing findings. Hayfron (2001) exploits data from the 90s based on an individual level survey. Hence, data in the thesis and that of Hayfron contain different arrival cohorts of immigrants that also differ with respect to time of residency in Norway before attending language classes. Another distinction is the immigrants' country of origin. Hayfron's analysis is confined to Third World male immigrants whereas this thesis examines immigrants of all backgrounds. The individual level survey allows for a richer set

[40] The predicted effect varies throughout the model specifications. Lochmann et al. (2018) employ two different bandwidths as well as local linear and parametric estimates. Although the majority of specifications produces negative estimates, it is worth noticing that estimates are positive in two specifications. Hence, the authors conclude that the results 'hint' towards a negative effect.

of qualitative control variables, albeit self-rated proficiency might be a source of under- or overestimation of true language skills. Lochmann et al. (2018) points out that language training might be regarded as an investment for future employment because immigrants might not work full time due to participation in language classes. The negative impact in their study might be a result of not analysing long run effects. This thesis, on the other hand, does account for a possible lagged effect of training and might therefore capture the impact in a longer term. This is also the case with Akresh (2007). However, this thesis differs from Akresh in that it analyses the effect on employment rather than earnings and does so with a larger sample size. In the same way as Sarvimäki & Hämäläinen (2016) are not able to differentiate between part-time and full-time work, this thesis might fail to capture the impact on hours worked.[41] However, results may be dissimilar due to the chosen key independent variable as this thesis examines language training itself instead of integration plans.

Results in this thesis are more comparable to the ones of Beiser & Hou (2000) and Dustmann & Fabbri (2003), although those papers do not directly study language training but general language proficiency. The studies find a strong positive relationship between English fluency and employment in Canada and the UK respectively. Overall, the positive but small effect in this thesis compared to the relevance of English proficiency might be due to the global standing of these languages. Moreover, the Norwegian population is considered relatively proficient in English compared to other populations (Education First, 2013). It might be the case that English is in fact a substitute for Norwegian for immigrants arriving in Norway. Such reasoning is in line with the findings of Yao & van Ours (2015) that find negative, yet insignificant, effects of Dutch deficiency on employment.

Having established a positive relationship amounting to 0.68 of a 10% increase in participation, the remainder of the thesis will show whether this effect differ with respect to municipal characteristics. In addition, it will ascertain the robustness of the results to miscellaneous specifications, including an instrumental variable.

[41] Section 4.1 pointed out that SSB defines employment as paid work for at least one hour in the reference week.

9. Heterogeneity Analysis

Thus far, the main analysis has studied the effect of participation in the training on employment at the municipal level without exploring how municipal characteristics may affect estimates. This chapter presents several sensitivity checks to examine how the main results vary across municipal dimensions. Sections 9.1 and 9.2 address potential different effects across urban and rural areas, and political parties in local governance. The chapter proceeds in Section 9.3 to investigate the impact of participation in training on immigrant employment in municipalities with a high and low share of immigrants from Africa and Asia. In the same manner as Section 9.3, Section 9.4 explores effect-differences between municipalities with a high and low share of male immigrants respectively. Finally, Section 9.5 examines age related differences. All regressions are based on the model specification in column (5) in the fixed effects approach, which is considered to be the main regression of the main analysis and includes all control variables.

9.1 Urban Indicator

Section 4.4 discussed the tendency of immigrants to settle down in urban areas.[42] Not only do labour markets tend to be more complex in greater cities, but the existence of ethnic enclaves may also be a factor affecting immigrant employment ((Damm, 2009); (Schüller, 2016)). It is therefore interesting to investigate whether urban areas face a different relationship between participation and immigrant employment than lesser urban areas. The definition of an urban area remains the same as described in Section 4.4. This analysis extends the main regression model with an interaction term of the urban area dummy and the key independent variable, $\ln(participants)_{it-1} \times urban$. The interaction term estimates the difference in urban areas compared to rural areas in terms of immigrant employment when participation increases.

Column (1) in Table 9.1 displays parameter estimates for the number of participants in training and the urban indicator. Immigrant employment appears to be more sensitive to changes in the number of participants in rural areas than in urban areas. A 10% increase in

[42] Fixed effects estimation omit the urban area dummy due to its constant nature. The dummy shows no within variation because the defined urban areas remained unchanged during the research years.

participation increases immigrant employment in rural areas with 0.69%. For urban areas, the effect is estimated to $0.69\% - 0.4\% = 0.29\%$. The difference is nonetheless economically small and the interaction term is not statistically different from zero. Hence, one cannot conclude with differing sensitivity across rural and urban areas. Furthermore, due to the exclusion of the urban dummy in fixed effects estimation, the analysis is not able to examine whether average immigrant employment is identical across urban and rural municipalities that have the same levels of participants and other socioeconomic factors controlled for in the analysis.

9.2 Political Indicators

In the same way as the urban area dummy was omitted in the fixed effects approach, the dummies for political party in local governance were also omitted. This extended analysis is neither able to investigate if average immigrant employment is identical across municipalities with varying political parties in local governance. The reason is that fixed effects estimation omits the dummies because of no within variation, similar to the urban dummy.[43] However, it is interesting to assess possible differences in the impact of participation on immigrant employment across municipalities, since political parties may conduct distinctive policies in order to employ immigrants.

Column (2) in Table 9.1 presents the extended analysis for the political party effect. The model expands with the interaction terms $\ln(participants)_{it-1} \times left$, $\ln(participants)_{it-1} \times right$ and $\ln(participants)_{it-1} \times list$. Participation in the training in municipalities governed by a central party appears to have a slightly higher impact on immigrant employment compared to the main results in Section 8.2. The model predicts that a 10% increase in the number of participants in a municipality governed by a central party increases immigrant employment with 0.82$\%$. The small economic effect is significant at the 10% level.

Municipalities with left and right political parties in local governance may anticipate a lower effect of participation in the training on immigrant employment than municipalities governed by central political parties. The estimates are given as 0.060 and 0.071, respectively. Municipalities governed by a coalition of parties, however, are predicted to have a higher

[43] The municipal election held in 2015 ensures certain time variation when the analysis covers three years (2014-2016), yet because of the lagged key independent variable the analysis examines only 2015 and 2016. Hence, the political dummies yield no time variation and are omitted.

effect, equal to 0.097. With regards to estimate precision, on the other hand, none of the political indicators yield significant t-statistics. Consequently, one cannot conclude any differences in effect of participation on immigrant employment between municipalities based on the political party in local governance.

9.3 African and Asian Immigrants

Immigrants may vary in characteristics because they originate from different parts of the world and distinct cultures. Thus, individual heterogeneity affecting employment, such as diligence and time management, may vary across countries of origin. Additionally, attitudes toward immigrants may differ as a result of their country of origin. Findings of SSB indicate particularly low employment for immigrants from Africa and Asia, roughly 20% and 10% lower respectively than their counterparts from other non-Europeans countries (SSB, 2017-11-08). The gap to European immigrants is even greater. It is therefore interesting to examine any effect-differentials across the regions of origin.

Column (3) in Table 9.1 displays how results of fixed effects estimation vary along with the region of origin. The sample is split according to the median share of African and Asian immigrants across municipalities, which is 0.31. The dummy variable $AfricaAsia$ takes the value one for municipalities in which the proportion of immigrants from Africa and Asia surpasses the median threshold of 31%. The interaction term $\ln(participants)_{-1} \times AfricaAsia$ estimates the impact differential across municipalities with a relatively high and low share of African and Asian immigrants.

Estimates suggest returns to participation in the training is higher for municipalities with a relatively high share of African and Asian immigrants. A 10% increase in the number of participants in training in these municipalities is predicted to cause an increment of 0.71% in the number of employed immigrants. On the other hand, the benefit in municipalities with relatively few immigrants from Africa and Asia is 3% lower, equal to 0.68%. However, the impact gap of increased participation on employment is not statistically significantly different from zero which makes it hard to draw a conclusion.

The results suggest further that the number of employed immigrants is 2.3% lower in municipalities with a high proportion of immigrants from Africa and Asia compared to other municipalities. Although the estimate is insignificant, it could be misleading to conclude that there is no evidence of lower immigrant employment in municipalities with many immigrants from Africa and Asia. Because of the inclusion of the interaction $\ln(participants)_{-1} \times AfricaAsia$, the coefficient on $AfricaAsia$ is estimated imprecisely due to high standard errors. This occurs since $AfricaAsia$ and $\ln(participants)_{-1} \times AfricaAsia$ are highly correlated in the sample. Moreover, the coefficient of $AfricaAsia$ measures the employment differential across the two types of municipalities when the number of participants is zero. No municipality in the sample reports zero participants, nor is it interesting to examine the gap at zero participants. Instead, it would be more useful to estimate the differential at the median level of $\ln(participants)_{-1}$ in the sample, which is 3.74. To do this, the interaction term is modified to $\ln(participants - median)_{-1} \times AfricaAsia$. When rerunning the regression at the median level of participants in training, municipalities with a relatively high share of African and Asian immigrants are predicted to have 1.26$\%$ lower immigrant employment than their municipal counterparts. However, in spite of obtaining lower standard errors, the rerun causes the estimate to be insignificant as the t statistic is given as $-0.0126/0.018 = -0.71$. Ergo, within the framework of this analysis, there is in fact no evidence that support differences in immigrant employment across municipalities with different proportions of non-European immigrants.

9.4 Heterogeneity by Sex

In order to examine possible gender gaps in the effect of participation in the training on employment, the analysis splits the sample in two based on the share of male immigrants in the municipality. A dummy variable, $male$, takes the value one for municipalities in which male immigrants make up more than half of the immigrant population. The coefficient on $male$ reflects any disparities in immigrant employment across municipalities with different proportions of immigrant men and women. Moreover, the interaction $\ln(participants)_{-1} \times male$ captures gender differences in the effect of participation in the training on immigrant employment.

Estimates suggest a relatively higher gain of participation in training for municipalities in which female immigrants outnumber male immigrants, compared to contrary municipalities. A municipality with a high share of female immigrants appears to increase immigrant employment by 0.67% for a 10% increase in participants. A municipality with lots of males, au contraire, may anticipate a 1.39% lower increase in employment. However, the estimated gap is not significant and one cannot draw any conclusion concerning gender differences in the returns to participation.

Immigrant employment is predicted to be 9.4% higher in municipalities with a majority of immigrant men compared to municipalities with a predominance of immigrant women. Although the estimate is insignificant, one must be careful with concluding that there is no statistically significant evidence of higher immigrant employment in male-intensive municipalities. Similar to the previous section on African and Asian immigrants, the inclusion of an interaction term increases standard errors. In fact, $\ln(participants)_{-1} \times male$ augments the standard error on $male$ by more than fourfold compared to a regression without the interaction term. The explanation is again high correlation between variables. Rerunning the regression with $\ln(participants - median)_{-1} \times male$ produces an estimated effect of 3.8% which is significant at the 5% level. This indicates that municipalities with relatively many male immigrants have 3.8% more employed immigrants than municipalities with relatively many female immigrants.

9.5 Heterogeneity by Age

Phsychobiological literature has found that the ability to learn a language decreases with increased age (Bleakley & Chin, 2004, p. 482). It is therefore interesting to see how this finding affects integration in the labour market. Data from the IMDi arranges immigrants in several age groups: four of these groups indicate working age as defined by 16-66 years old.[44] This section evaluates the impact of participation in training on immigrant employment across municipalities in which immigrants between 16-29 years old constitute a relatively high or low share of immigrants of working age. Across municipalities, these immigrants make up on average 26% of the immigrant population between 16-66 years old, which is also the median. Hence, the dummy variable $young$ takes the value one for municipalities with a share of young immigrants exceeding the threshold at 26%. As in earlier heterogeneity checks, the model includes an interaction term, $\ln(participants)_{-1} \times young$, which assesses possible gaps in the returns to participation in training on immigrant employment.

[44] The groups of working age are 16-19, 20-29, 30-54 and 55-66 years old.

The regression is found in column (5) in Table 9.1. Parameter estimates indicate that municipalities with relatively many young immigrants of working age benefit more than other municipalities. Albeit this is consistent with psychobiological findings, the difference is in practice negligible. For a 10% increase in the number of participants in training, immigrant employment is predicted to rise 0.71% in the former type of municipality. The increment in the latter kind of municipality is, on the other hand, estimated to be merely 0.8% lower. In addition to an economically imperceptible differential, the estimates are insignificant.

Based on an insignificant coefficient, it appears furthermore to be no evidence of greater immigrant employment in municipalities with a high age demographic among immigrants. However, as the previous section established, estimates should rather be calculated at the median level of participants instead of a zero level. Running a regression with the interaction $\ln(participants - median)_{-1} \times young$ seems nevertheless to support that there is not sufficiently strong evidence to conclude with any employment gaps across municipalities based on immigrant age.

9.6 Discussion of Heterogeneous Effects

The heterogeneity analysis indicates that municipalities with relatively many immigrant men have 3.8% higher immigrant employment compared to municipalities with relatively many immigrant women. The result is in line with the findings of Lochmann et al. (2018, p. 32) and Dustmann & Fabbri (2003, p. 706). However, their estimates suggest a greater gender gap of 13-18%. Males are often the major breadwinner of the family, and the importance of them finding a job will therefore lead to an inelastic supply of labour (Yao & van Ours, 2015, p. 83). This argument is strengthened by looking at the reason for migration of the immigrants in the sample. Males are often first in the family to migrate as labour migrants or refugees. Females, on the contrary, often come as family migrants to reunite with their already-migrated spouses. On the other hand, there is no evidence in this thesis that supports gender gaps in the returns to participation in the training. Beiser & Hou (2000, p. 324) find, however, that female refugees are more likely to enter the labour force due to increased English language proficiency than their male counterparts. The results of this thesis also differ from the findings of Lochmann et al. (2018, p. 18). The authors estimate that females benefit less from language training than males in terms of labour force participation. Although insignificant, the estimate in this thesis is of a negative sign which would indicate a greater benefit of participation for female-dominated municipalities.[45] Likewise, Yao & van Ours (2015) find that female immigrants are more affected by language deficiency than male immigrants and argue that the reason might be that female labour supply is more sensitive to human capital. The reasoning is based on the breadwinner-argument already discussed which suggests that females are more elastic in their labour supply than their male counterparts. Furthermore, in the sample of employed workers eligible for participation in the training, about 57% of males and 24% of females are doing manual work that do not require much language proficiency. Contrarily, industries that require language skills, such as education, business services and public administration, employ 37% of males and roughly 64% of females. In light of this, one could argue that the negative sign on $\ln(participants - median)_{-1} \times male$ makes sense because female immigrants tend to speak Norwegian at work to a greater extent than male immigrants. The estimate is anyhow insignificant and rules out conclusive evidence.

It was further shown that the results are not sensitive across urban or rural dimensions. Similarly, Yao & van Ours (2015) find no relationship between residing in an urban area and language proficiency. In spite of statistically insignificance in this thesis, the coefficient for rural areas was estimated to be larger than for urban areas. Schüller (2016, p. 3) points out that ethnic enclaves are often located geographically unfavourable in terms of active employment opportunities. Moreover, she proposes the residential sorting problem as a possible cause for a negative link between enclave residency and employment probabilities. Such a situation would arise if adversely low-skilled immigrants that struggle to find a job select into ethnic enclaves.

The analysis does not support different impact of participation in the training on employment across municipalities with different political parties in local governance. An explanation might be a common understanding for political parties of the importance of integrating immigrants on the labour market. Additionally, local politics are less ideological at the municipal level and conducted policies tend to vary to a greater extent at the national level.

Moreover, no conclusion can be drawn concerning municipalities with a relatively high or low share of immigrants from Africa and Asia. Nonetheless, the coefficient sign indicates higher gain from participation in municipalities with relatively many African and Asians. On the other hand, immigrant employment is assessed to be at an inferior level in these municipalities.

[45] Notice that the effect differential cannot be explained by different average levels of any included independent variable because these are already controlled for.

Both coefficients seem to reflect a greater need for Norwegian language training for immigrants with increasing cultural and linguistic differences from Norwegian. In the Netherlands, Yao & van Ours (2015) find for instance that female immigrants from non-western countries are more likely to have language problems than female immigrants from western countries.

Finally, inference cannot be drawn concerning municipalities with a high and low proportion of young immigrants of working age. The coefficients for the return differential to participation and the gap in immigrant employment level are not statistically significantly different from zero. Notwithstanding, the positive sign of the coefficient on $young$ is in line with theory stating that learning a language might be easier for immigrants arriving in the host-country at a young age. Access to more detailed age groups might increase significance and Lochmann et al. (2018) find, for instance, that language training appears to be more beneficial for immigrants below 40 years old. Moreover, it is not surprising that the coefficient on $\ln(participants - median)_{-1} \times young$ indicates that municipalities with a higher immigrant age demographic display a higher level of immigrant employment compared to opposite municipalities.

This chapter presented heterogeneity analyses to examine differences in the return to participation in the training on employment across municipal dimensions. The next chapter performs several robustness checks to investigate the validity of the results obtained in the main analysis.

Table 9.1: Heterogeneity analysis

Variable	(1) ln(employment)	(2) ln(employment)	(3) ln(employment)	(4) ln(employment)	(5) ln(employment)
ln(participants)$_{-1}$	0.069***	0.082*	0.068**	0.067***	0.063***
	(0.024)	(0.050)	(0.027)	(0.023)	(0.024)
ln(participants)$_{-1}\times$ urban	-0.040	-	-	-	-
	(0.087)	-	-	-	-
ln(participants)$_{-1}\times$ left	-	-0.220	-	-	-
	-	(0.478)	-	-	-
ln(participants)$_{-1}\times$ right	-	-0.011	-	-	-
	-	(0.058)	-	-	-
ln(participants)$_{-1}\times$ list	-	0.015	-	-	-
	-	(0.167)	-	-	-
AfricaAsia	-	-	-0.023	-	-
	-	-	(0.097)	-	-
ln(participants)$_{-1}\times$ AfricaAsia	-	-	0.003	-	-
	-	-	(0.025)	-	-
male	-	-	-	0.090	-
	-	-	-	(0.082)	-
ln(participants)$_{-1}\times$ male	-	-	-	-0.014	-
	-	-	-	(0.020)	-
young	-	-	-	-	-0.021
	-	-	-	-	(0.059)
ln(participants)$_{-1}\times$ young	-	-	-	-	0.008
	-	-	-	-	(0.012)
Control variables from main specification	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes
Year 2016	0.121***	0.120***	0.121***	0.112***	0.122***
	(0.015)	(0.015)	(0.014)	(0.015)	(0.015)
Constant	-8.237	-8.316	-8.372	-5.101	-8.146
	(6.643)	(6.873)	(6.651)	(6.315)	(6.652)
Observations	601	601	601	596	601
R^2 within	0.469	0.469	0.469	0.469	0.470

Note: Each column represents a separate specification. Specifications (1) and (2) regress the urban and political indicators. Specification (3) estimates the effect for municipalities with a relatively high and low share of immigrants from Africa and Asia. Specification (4) displays results for municipalities with a relatively high and low share of male immigrants. All regressions include the set of control variables from specification (5) in the fixed effects analysis (total immigrant population, production index, household income and municipal expenses related to the training). The number of employed immigrants is the dependent variable in all specifications, while the number of participants at time $t-1$ is the key independent variable. Cluster-robust standard errors in parenthesis. *Significant at 10%; **significant at 5%; ***significant at 1%.

10. Robustness Analysis

As several control variables were included in the regression model, Section 8.2 displayed what appeared to be robust estimates for participation in the training on employment. Nonetheless, Chiswick & Miller (2014, p. 9) point out the limited value of a model that is not generalisable across dimensions. Up to now, the analysis has disregarded any missing observations in the data set. Chapter 4 revealed the presence of missing data in certain variables and Section 10.2 presents a full treatment of the potential issue of those missing observations. Moreover, Section 10.3 examines the robustness of the main analysis when employing a non-lagged key independent variable. The chapter then proceeds to substitute the dependent variable of the main analysis to account another age group in Section 10.4. First, however, an instrumental variable approach addresses the possible endogeneity in the key independent variable in Section 10.1. Unless stated otherwise, the regressions are performed with fixed effects estimation with cluster-robust standard errors.

10.1 Instrumental Variable

Chapter 5 discussed the foundation of the main analysis and argued that identification issues presented a major concern to the naive model. In particular, the challenge was self-selection of immigrants into participation or specific municipalities. If the key independent variable was correlated with unobserved heterogeneity, simple OLS would produce inconsistent and biased estimators. Immigrants' motivation for learning Norwegian was used as an example of individual characteristics that are hard to measure and could bias participation. Nonetheless, it was argued that participation was in practice exogenous and several explanatory variables was included in the model to deal with selective municipal sorting. The argumentation hinged upon assumptions concerning the self-selective behaviour of education and seasonal immigrants from outside the EEA and immigrants migrating according to EEA regulations. Operating as an incentive that counteracts participation, the former group might not experience a particular need to learn Norwegian. The latter group is not inclined to participate as shown by numbers from SSB. However, if these arguments fail to hold, roughly 3 of 10 immigrants participate exogenously.[46] This section examines robustness of the main analysis by assuming participation is correlated with unobserved heterogeneity in the error term. Consequently, model 5.2 takes the form

$$\begin{aligned}\ln(employment)_{it} &= \beta_0 + \beta_1 \ln(participants)_{it-1} + \beta_2 \ln(totimmpop)_{it} \\ &\quad + \beta_3 \ln(prodindex)_{it} + \beta_4 \ln(income)_{it} + \beta_5 \ln(expenses + 1)_{it} \\ &\quad + \beta_6 unobschar_{it} + \delta_1 urban + \delta_2 left + \delta_3 right + \delta_4 list + \alpha_t + u_{it}\end{aligned} \tag{10.1}$$

In this case, $\ln(participants)_{it}$ is not independent from $unobschar_{it}$, which is unobserved and thus captured by the error term. In order to obtain consistent and unbiased estimators, one would need to correct for self-selection into training. The analysis employs an instrumental variable to address the problem. Briefly stated, the instrumental variable method, often written 2SLS (two-stage least squares), uses predicted values in the original equation based on a first stage regression for the key independent variable. Hence, the idea is that the instrument works on behalf of the endogenous variable and eliminates heterogeneity bias. Following the two-assumption approach presented in Wooldridge (2016, ch. 15), the instrument must satisfy relevancy and exogeneity given as

$$Cov(z, \ln(participants)_{-1}) \neq 0 \tag{10.2}$$

$$Cov(z, u) = 0 \tag{10.3}$$

First, the instrument, z, ought to be correlated with the number of participants in training, referred to as instrument relevance. This correlation is a measure of the strength of the instrument where a high correlation implies low variance and a strong instrument. Second, the instrument must be uncorrelated with unobserved heterogeneity in the error term. Moreover, it

[46] The estimate is calculated on the basis of Figure B.1 in appendix B and is given as an approximate average across 2014-2016. In particular, the number is volatile to the share of families that reunites with refugees and labour immigrants, which is unobserved in the data. Consequently, the estimated number is an interval. If the assumptions made in Chapter 5 fail to hold, between 20% and 34% of immigrants participate exogenously in 2014. In 2015, the interval is 27-43%. Finally, in 2016, exogenous participation is estimated to 36-52%. The increasing numbers are due to a boost in the number of refugees.

should only have an indirect effect on the number of employed immigrants, via the number of participants. This is often referred to as instrument exogeneity and enables the use of z.

The instrumental variable method is often used in studies on language proficiency for immigrants. However, the instruments are often given at the individual level such as age at arrival, country of birth and dummy for child in school living at home (Bleakley & Chin (2004); Yao & van Ours (2015); Budría et al. (2017)). Given the available data in this analysis, the aforementioned instruments are not feasible. Other instruments dealing with endogenous participation in language training include unemployment, social benefits and test results (Hayfron (2001); Lochmann et al. (2018)). Due to the choice of the dependent and key independent variable in this study, as well as the arrangement of the Norwegian language training and social studies, these instruments are not feasible. The limited benefit of former studies and instruments makes the search of a viable instrument tedious. However, the number of allocated immigrants in municipality i appears to be a valid instrument. For this to be credible, the number of allocated immigrants must be uncorrelated with the error term in equation 10.1 and immigrant employment, and it must be partially correlated with participation in the reduced form equation. The latter means that after partialling out the effect of the other explanatory variables, allocations and participation are still correlated.

Since allocated immigrants bear refugee status, they must enroll in the training and have no possibility of manipulating participation. This means that the number of allocated immigrants in municipality i seemingly explains much of the variation in the number of participants in the training in municipality i. The instrument also excludes the issue of self-selection into specific municipalities as allocation is done exogenously by the IMDi.[47] To check the requirement of partial correlation, participation is regressed on allocations and all of the other explanatory variables appearing in model 10.1. Hence, one regress the reduced form equation for participation, given as

$$\ln(participants)_{it-1} = \pi_0 + \pi_1 \ln(allocation)_{it} + \pi_2 X_{it} + \pi_1 D_{it} + u_{it} \tag{10.4}$$

[47] One might argue that refugees as individuals are not randomly allocated to municipalities but determined by local economic reasons. However, the analysis takes account of such reasons through a short panel and a set of control variables at the municipal level.

The assignment variable, $\ln(allocations + 1)_{it}$, is the number of allocated immigrants in municipality i. The variable is given in a modified logarithmic form for reasons discussed in Chapter 5. The set of control variables of vector X_{it} and dummy variables of vector D_{it} includes the variables from equation 5.2. Test value for instrument relevancy is reported in Table 10.1. The coefficient of interest is positive as expected, but more important is the t statistic which amounts to 9.09. Stock & Yogo (2005, p.106) propose a rule of thumb concerning the size of the t statistic in an exactly identified multiple regression model.[48] In particular, the t statistic should have an absolute value surpassing 3.2. Clearly, the null hypothesis is rejected and $\ln(allocations + 1)_{it}$ is a relevant instrument.

The above result suggests that if the number of allocated immigrants is uncorrelated with unobserved factors in the error term and immigrant employment, it is a viable instrument for the number of participants in training. Testing the exclusion restriction is, however, not feasible in the case with exact identification which is the case in this analysis. Consequently, the matter of exogeneity must be made on the basis of reasoning. First, allocations appears to be uncorrelated with unobserved heterogeneity, again using motivation as an example. The reason is the exogenous allocation process carried out by the IMDi. Moreover, allocated immigrants are refugees who do not decide on participation themselves. Second, allocations appear to have no direct impact on immigrant employment beyond the one flowing indirectly through participation. Once allocated, the immigrants must enroll in the training thus preventing immediate employment.[49][50]

Column (1) in Table 10.1 reports estimates for 2SLS regression using $\ln(allocations + 1)_{it}$ as an instrumental variable for $\ln(participants)_{it}$.[51] The estimate suggests that municipality i might expect an increase of 2.45% in immigrant employment for a 10% increase in the number of participants in the training. Significant at the 1% level, the estimate is assumed to be precise. This 2SLS estimate is, somewhat surprisingly, higher than its OLS counterpart, reported as 1.74% in column (6) in Table 6.1. The OLS estimate thus appears to be downward biased.

48 A model is exactly identified when there is one instrument for each endogenous explanatory variable.

49 To be sure, the number of allocated immigrants is likely to increase the number of employed immigrants in the long run, but that is beyond the scope of this analysis.

50 While enrolled, the immigrant may partly complete the training at a workplace, but this is not regarded as employment.

51 2SLS-FE regression produced insignificant estimates. The first-stage regression further proved the instrument to be weak in this specification with a t statistic of -2.04. Not only is this test value below the thumb rule threshold of Stock & Yogo (2005), but it is also of a counter-intuitive sign. For these reasons, results are not reported.

Yao & van Ours (2015) claim this is often the case when estimating language effects with 2SLS.[52] Nonetheless, it should be noted that the 95% confidence intervals of OLS and 2SLS are overlapping.[53]

2SLS also estimates the effect to be larger than the fixed effects prediction presented in Table 8.2. More specifically, the predicted effect is 3.6 times higher when estimated with 2SLS. Albeit the larger magnitude, the sign remains significantly positive. This suggests that the fixed effects finding of a positive effect of participation on immigrant employment is robust to the use of an instrumental variable. Furthermore, Table 10.1 reports a Durbin-Wu-Hausman test to examine the endogeneity of $\ln(participants)_{it}$. The test value is insignificant and the null hypothesis is kept, which indicates that participation can be treated as exogenous in the analysis. Hence, the fixed effects estimation appears to be reliable. However, the estimated effect of the total immigrant population is worth mentioning. The coefficient is substantially different compared to the fixed effects estimation and is highly significant. A possible explanation is the high correlation between this variable and the instrumented variable. Wooldridge (2016, p.477) shows that correlation is often reinforced in 2SLS estimation, but this does not alter the major finding that the estimated effect of participation in the main analysis is robust.

10.2 Missing Data and Bottom Coding

In the event of considerable missing data, the assumption of a randomly drawn sample might be violated. Determining the missing data pattern and the missing data mechanism is important because different patterns and mechanisms require different solutions (Little & Rubin, 2002, p. 11). The pattern of missing data in this analysis appears to be of a general form, shown in Figure B.3 in appendix B. Chapter 4 points out that SSB does not register values for municipalities with fewer than four participants which implies that data is not missing at random (NMAR). The underlying argument is that lower values of immigrant employment or participation are more likely to appear as missing in the data. Consequently,

[52] The authors propose the reason to be that the upward bias of endogeneity is overshadowed by the negative bias of measurement error. However, the statement is based on studies taking advantage of self-reported language skills which is a source of measurement error (Dustmann & Van Soest (2002); Dustmann & Fabbri (2003); Bleakley & Chin (2004)).

[53] Another finding worth emphasising is the increase in standard errors in the 2SLS method due to larger variance. Such high standard errors may be caused by a high degree of multicollinearity (Wooldridge, 2016, p. 477), but a large sample size may downplay this issue.

missing data disguise true values that might prove useful for the analysis.

The basic and main analyses perform casewise deletion in line with Akresh (2007), omitting municipalities with completely missing observations for either the dependent variable or the main independent variable during all three research years.[54] Table B.2 in appendix B presents the scope of missing data across all municipalities. Total immigrant population and median household income are complete variables without missing observations. For the two main variables, immigrant employment and participation, the share of missing data is 13.4% and 8.6% respectively. The total share of missing data in the sample is 9.7%.

There might be a tendency of missing values to appear in municipalities with fewer inhabitants. SSB reports underline this argument. Figure 10.1 displays missing and non-missing values for the dependent variable and the key independent variable across municipalities. The map supports the hypothesis that municipalities with fewer inhabitants are more likely to contain missing values than more urban municipalities. In other words, the probability of missing values is higher in rural municipalities than in urban municipalities.

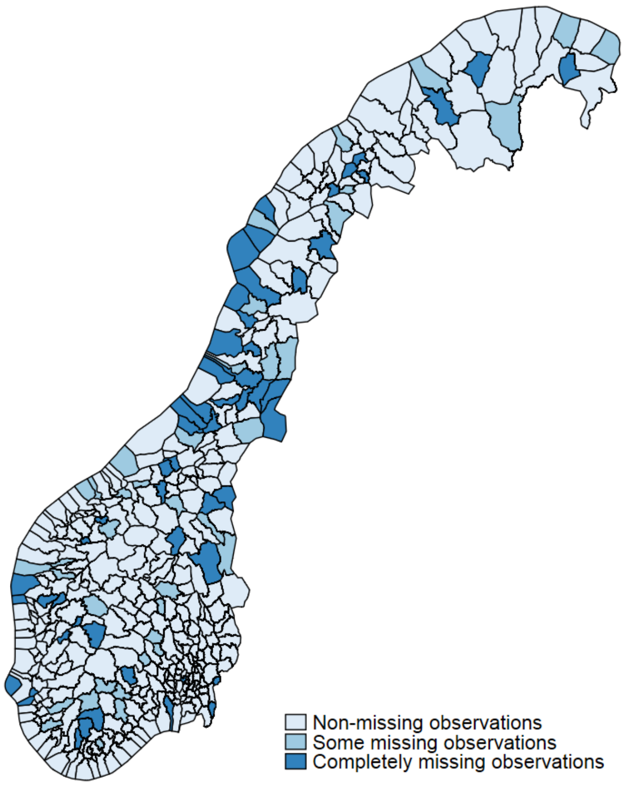


Figure 10.1: Map of municipalities with missing and non-missing observations. Shapefiles are extracted from Geonorge and contain maritime borders of the municipalities.

[54] As these variables are of particular interest, it is not expedient to include municipalities without information concerning these.

In order to address possible issues caused by left-censored data,[55] this section bottom codes missing values inasmuch as SSB data creates a threshold at five participants. Wooldridge (2016, p.548) points out that regressions on a sample containing only the uncensored observations above the threshold yields inconsistent estimators. Thus, the censored analysis imputes missing observations in the key independent variable with values equal to three and a dummy equal to one for bottom coded observations, $bottomcode$. Moreover, controlling for urban areas with a binary variable is anticipated to have a moderating effect on missing data challenges because of the apparent relationship between missing observations and a small population. Missing observations in the dependent variable, on the contrary, remain missing in this analysis.[56]

Column (2) in Table 10.1 provide parameter estimates with bottom coding of missing values in the key independent variable. The bottom code regression seems to provide evidence of robust results in the main analysis. A 10% increase in the number of participants is predicted to increase the number of employed immigrants with 0.6%. The estimate is only 0.88% lower than in the main analysis and significant at the 5% level. Additionally, the magnitude of the socioeconomic variables remain more or less the same as in the main analysis, as well as their significance. The coefficient on $bottomcode$ predict 2.33% higher immigrant employment in bottom coded observations than non-bottom coded observations. The estimate is, however, not significant and there is no evidence of a difference in effect of participation in the training on employment across bottom coded and other observations.

10.3 Static vs. Dynamic Modelling

This section compares regression results of using $\ln(participants)_{it-1}$ and $\ln(participants)_{it}$ as the key independent variable. Results of the former is shown in Table 8.2. Column (3) in Table 10.1 presents results of $\ln(participants)_{it}$ as the key independent variable. Not lagging the key independent variable allows the regression to exploit data from the entire sample period 2014-2016. Moreover, regressing on $\ln(participants)_{it}$ yields a static model that does not capture longer term effects seized by $\ln(participants)_{it-1}$. Compared to the benchmark estimate, the predicted effect of participation is now considerably different. Employing the number of participants at time t as the key independent variable suggests that

[55] Left-censored data refers to unobserved observations below a certain value.
[56] SSB does not inform of any censoring regarding this.

the number of employed immigrants decreases 0.15% when the number of participants in training increases 10%. The negative estimate might be due to the simultaneity issue discussed in chapter 5 because immigrants may not work while being enrolled in the training. This may indicate that lagging the key independent variable as in the benchmark model does solve endogeneity. Notwithstanding, the estimate of the static model is not statistically different from zero. Thus, it appears to be no evidence for a relationship between the number of participants and the number of employed immigrants with $\ln(participants)_{it}$ as the key independent variable.

Concerning the explanatory variables, results are generally consistent with predicted outcomes in the main analysis. However, despite similar coefficient signs, significance and magnitude change slightly. The exception is total immigrant population that is positive, yet insignificant, in contrast to the significantly negative estimate in the main analysis.

Contrarily to the main analysis, the static model includes the political dummies in the fixed effects estimation. Extending the analysis to three years allows to take into account the municipal election in 2015 and potential political turnover. As a consequence, the dummies possess within variation and are not omitted. However, none of the coefficients are statistically different from zero.

10.4 Different Dependent Age Group

The main analysis examines the causal impact of participation in the training on employment for immigrants covered by the *introduction act*. Hence, the dependent variable, $\ln(employment)_{it}$, contains an immigrant group that serves as an approximate fit for the age requirements for eligibility. Section 4.1 explains that the chosen group is immigrants between 15 and 74 years old. However, since this group is not an exact match, it is interesting to check the robustness of the results for another approximately matched group. Data from SSB also provide employment information among immigrants 20-66 years of age. Hence, this particular robustness regression analyses the effect of Norwegian language training and social studies on the employment of immigrants between 20 and 66 years old. This variable contains less missing observations and the regression in column (4) in Table 10.1 takes advantage of 414 municipalities.

Estimated to 0.94% for a 10% increase in participation, the predicted effect of participation is higher than its counterpart in Table 8.2. The difference is nevertheless appropriately small to confirm the robustness of the main analysis estimates at the 1% significance level. It seems therefore to be of little importance whether the chosen age group is 20-66 or 15-74 years old. Control variables are further generally similar to the main regression without the exception of $\ln(expenses + 1)$ that turns negative and insignificant compared to the main analysis estimate.

10.5 Robustness Summarised

Estimates from the main analysis are shown to be robust through four different robustness checks. First, the instrumental variable, given as the number of allocated immigrants, underlines the positive relationship between participation in training and immigrant employment. The parameter estimates are highly significant and appear precise. Furthermore, the endogeneity test fails to reject the null hypothesis of an endogenous key independent variable. This provides evidence that the main analysis does correct for unobserved heterogeneity. Second, bottom coding of missing values indicates that missing data do not pose a threat to the validity of the main analysis. On the contrary, it is shown that the bottom coded analysis accentuates the sign and magnitude of the key independent variable estimate. Third, the static model regression seemingly justifies the lagged key independent variable in the main analysis, which addresses possible simultaneity in the variable. The negative estimate in the static model can be compared to the lack of evidence on employment probabilities in the study by Lochmann et al. (2018). It is not surprising to find that when immigrants enroll in the training, the likelihood of them becoming employed is negative in the short term. Finally, performing the main regression with an alternative age group as the dependent variable confirms the findings of the main analysis.

Table 10.1: Robustness analysis

Variable	2SLS	FE		
	(1)	(2)	(3)	(4)
	ln(employment)	ln(employment)	ln(employment)	ln(employment)
ln(participants)	-	-	-0.015	-
	-	-	(0.021)	-
ln(participants)$_{-1}$	0.245***	0.060**	-	0.094***
	(0.060)	(0.033)	-	(0.025)
ln(totimmpop)	0.932***	-0.291*	0.218	-0.319
	(0.054)	(0.171)	(0.117)	(0.203)
prodindex	0.0002	0.0004	-0.002**	0.001
	(0.002)	(0.002)	(0.001)	(0.002
ln(income)	0.324	1.138**	0.699*	1.318
	(0.216)	(0.486)	(0.374)	(0.740)
ln(expenses+1)	0.021*	0.011***	0.001	-0.012
	(0.012)	(0.004)	(0.004)	(0.013)
bottomcode	-	0.023	-	-
	-	(0.038)	-	-
urban	0.036	-	-	-
	(0.071)	-	-	-
left	0.065	-	-0.029	-
	(0.041)	-	(0.018)	-
right	0.042	-	-0.009	-
	(0.048)	-	(0.018)	-
list	0.135*	-	0.041	-
	(0.074)	-	(0.054)	-
Year fixed effects	Yes	Yes	Yes	Yes
Constant	-6.752**	-8.523	-5.648	-10.636
	(2.745)	(6.556)	(5.230)	(9.091)
Test for a weak instrument: t test	9.090***	-	-	-
Endogeneity test: Durbin-Wu-Hausman	1.701	-	-	-
Observations	595	611	873	621
R^2 within	0.963	0.468	0.433	0.394

Note: Each column represents a separate specification. Columns (2)-(4) are estimated with fixed effects, whereas column (1) is estimated with 2SLS. The number of employed immigrants is the dependent variable in all specifications, while the number of participants at time $t-1$ is the key independent variable. The number of participants at time t is the key independent variable in the third column. Cluster-robust standard errors in parenthesis. *Significant at 10%; **significant at 5%; ***significant at 1%.

11. Conclusive Remarks

This thesis examines how participation in the Norwegian language training and social studies affect immigrant employment. In particular, the thesis studies the effect on the number of employed immigrants in a municipality when the number of participants increases in that municipality. Through a fixed effects estimation with an array of socioeconomic controls, the main analysis corrects for endogeneity in participation which is often a challenge in such studies. The research question is answered by exploiting panel data at the municipal level between 2014 and 2016.

Across various models and specifications, the study identifies a positive relationship between the number of participants in training and the number of employed immigrants in a municipality. The best estimates predict the number of employed immigrants to increase by 0.68% for a 10% increment in the number of participants. Although the impact is precisely estimated, it is economically small. There is no evidence of a gap in returns to participation across municipalities with a relatively high share of African and Asian immigrants, male immigrants or young immigrants of working age. Neither is there support of differences across urban and rural municipalities or across municipalities with varying political parties in local governance. The finding is robust throughout several extensions. A static model underlines the importance of lagging the key independent variable in the main analysis. Age cohorts and bottom coding of missing values further validate the results. Finally, the study displays how results are robust not only to fixed effects but also to the inclusion of an instrumental variable given as the number of allocated immigrants.

Future studies on the topic are advised to obtain data on an individual level to prevent issues of measurement. In fact, there is a mismatch of age and region of origin for the dependent and key independent variable concerning eligibility of participation. These issues work in opposite directions which makes the overall impact ambiguous. This is particularly relevant due to the attenuation bias that it might cause. Further works should clarify this issue. It is also a drawback that the data set does not provide information on the number of training hours received as this may vary among participants. Similarly, undeclared work is a challenge worth considering in future research as this may result in downwards estimation bias. The research is easily extended to other labour market measurements as earnings and hours worked, and it should aim to include more time periods to avoid little within variation in the variables.

Nonetheless, the finding of this thesis is important because it establishes a positive relationship between language training and employment within a unique framework. For policy makers, it is critical to obtain knowledge about the efficacy of policy tools. The results imply that language training can be a useful tool in order to prepare immigrants for labour market participation.

References

Akresh, I. R. (2007). US immigrants' labor market adjustment: Additional human capital investment and earnings growth. *Demography*, *44*, 865–881.

Bansak, C., Simpson, N. B. & Zavodny, M. (2015). *The economics of immigration [kindle]*. Routledge. https://www.amazon.com/Economics-Immigration-Cynthia-Bansak-ebook/dp/B00WMS8QE0/ref=mt_kindle?_encoding=UTF8&me=. (Accessed: 2018-01-09)

Beenstock, M. (1996). The acquisition of language skills by immigrants: The case of Hebrew in Israel. *International migration*, *34*, 3–30.

Beiser, M. & Hou, F. (2000). Gender differences in language acquisition and employment consequences among Southeast Asian refugees in Canada. *Canadian Public Policy/Analyse de Politiques*, 311–330.

Black, S. E. (1999). Do better schools matter? Parental valuation of elementary education. *The Quarterly Journal of Economics*, *114*, 577–599.

Bleakley, H. & Chin, A. (2004). Language skills and earnings: Evidence from childhood immigrants. *Review of Economics and statistics*, *86*, 481–496.

Blom, S. (2012). Innvandreres bostedspreferanser-årsak til innvandrertett bosetting. *SSB-rapport*, *44*.

Borjas, G. J. (1994). The economics of immigration. *Journal of Economics Litterature*, *32*, 1667-1717.

Borjas, G. J. (1999). Immigration and welfare magnets. *Journal of labor economics*, *17*(4), 607–637.

Borjas, G. J. (2014). *Immigration economics [kindle]*. Harvard University Press. doi:10.4159/harvard.9780674369900.

Büchel, F. & Frick, J. R. (2005). Immigrants' economic performance across Europe – does immigration policy matter? *Population Research and Policy Review*, *24*, 175–212.

Budría, S., de Ibarreta, C. M. & Swedberg, P. (2017). The impact of host language proficiency across the immigrants' earning distribution in Spain. *IZA Journal of Development and Migration*, *7*(1), 12.

Cameron, A. C. & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317–372.

Chiswick, B. R. (1978). The effect of Americanization on the earnings of foreign-born men. *Journal of Political Economy*, *86*, 897–921.

Chiswick, B. R. (1991). Speaking, reading, and earnings among low-skilled immigrants. *Journal of Labor Economics*, *9*, 149–170.

Chiswick, B. R. & Miller, P. W. (2014). International migration and the economics of language. *IZA Discussion Paper Series*(7880).

Damm, A. P. (2009). Ethnic enclaves and immigrant labor market outcomes: Quasi-experimental evidence. *Journal of Labor Economics*, *27*, 281–314.

De Giorgi, G. & Pellizzari, M. (2009). Welfare migration in Europe. *Labour Economics*, *16*, 353–363.

Djuve, A. B., Kavli, H. C., Sterri, E. B. & Bråten, B. (2017). *Introduksjonsprogram og norskopplæring. Hva virker - for hvem?* `https://www.fafo.no/index.php/zoo-publikasjoner/fafo-rapporter/item/introduksjonsprogram-og-norskopplaering`. Fafo. (Accessed: 2018-09-02)

Dustmann, C. & Fabbri, F. (2003). Language proficiency and labour market performance of immigrants in the UK. *The Economic Journal*, *113*, 695–717.

Dustmann, C. & Van Soest, A. (2002). Language and the earnings of immigrants. *ILR Review*, *55*(3), 473–492.

Edin, P.-A., Fredriksson, P. & Åslund, O. (2003). Ethnic enclaves and the economic success of immigrants—evidence from a natural experiment. *The Quarterly Journal of Economics*, *118*, 329–357.

Education First. (2013). *EF English Proficiency Index.* https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/innvandrere-i-og-utenfor-arbeidsmarkedet. NIFU. (Accessed: 2018-07-31)

Giulietti, C. & Wahba, J. (2012). Welfare migration. *IZA Discussion Paper Series*(6450).

Gonzalez, A. (2000). The acquisition and labor market value of four English skills: new evidence from NALS. *Contemporary Economic Policy*, *18*, 259–269.

Hagelund, K., Nordbø, E. W. & Wulfsberg, F. (2011). *Effekter av arbeidsinnvandring.* https://www.norges-bank.no/globalassets/upload/.../staff.../2011/staffmemo_1211.pdf. Norges Bank. (Accessed: 2018-01-12)

Hansen, J. & Lofstrom, M. (2003). Immigrant assimilation and welfare participation: Do immigrants assimilate into or out of welfare? *Journal of Human Resources*, *38*, 74–98.

Hayfron, J. E. (2001). Language training, language proficiency and earnings of immigrants in Norway. *Applied Economics*, *33*, 1971–1979.

Introduksjonsloven. (2003). *Lov om introduksjonsordning og norskopplæring for nyankomne innvandrere (introduksjonsloven) 2003-07-04.* https://lovdata.no/dokument/NL/lov/2003-07-04-80#KAPITTEL_4. (Accessed: 2018-03-12)

Jannecke, W.-J. (2013). *Utenlandske studenter i Norge.* https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/innvandrere-i-og-utenfor-arbeidsmarkedet. NIFU. (Accessed: 2018-07-31)

Kane, T. J., Staiger, D. O., Samms, G., Hill, E. W. & Weimer, D. L. (2003). School accountability ratings and housing values. *Brookings-Wharton Papers on urban Affairs*, 83–137.

Kompetanse Norge. (2017-08-31). *Godkjente tilbydere av norskopplæring).* https://www.kompetansenorge.no/Norsk-og-samfunnskunnskap/Godkjente-tilbydere-av-norskopplaring/. (Accessed: 2018-03-28)

Little, R. J. & Rubin, D. B. (2002). *Statistical Analysis with Missing Data* (Vol. 2nd). John Wiley & Sons, Inc.

Lochmann, A., Rapoport, H. & Speciale, B. (2018). The effect of language training on immigrants' economic integration: Empirical evidence from france. *IZA Discussion Paper Series*(11331).

Ministry of Children and Equality. (2010-2011). *Endringer i introduksjonsloven og statsborgerloven (prop. 79 l).* https://www.regjeringen.no/no/dokumenter/prop-79-l-20102011/id636689/sec1. (Accessed: 2018-01-27)

Ministry of Education and Research. (2015-2016). *Fra utenforskap til ny sjanse (meld. st. 16 2015-2016).* https://www.regjeringen.no/no/aktuelt/fra-utenforskap-til-ny-sjanse/id2476577/. (Accessed: 2018-03-12)

Ministry of Justice and Public Security. (2015-2016). *Fra mottak til arbeidsliv - en effektiv integreringspolitikk (meld. st. 30 2015-2016).* https://www.regjeringen.no/en/dokumenter/meld.-st.-30-20152016/id2499847/sec1. (Accessed: 2018-01-12)

Ministry of Justice and Public Security. (2016). *Rundskriv til lov om introduksjonsordning og norskopplæring for nyankomne innvandrere (introduksjonsloven).* https://www.regjeringen.no/no/dokumenter/g012016/id2511944/. (Accessed: 2018-01-18)

Ministry of Local Government and Modernisation. (2017). *Rapport fra det tekniske beregningsutvalg for kommunal og fylkeskommunal økonomi.* https://www.regjeringen.no/no/dokumenter/rapport-fra-det-tekniske-beregningsutvalget-for-kommunal-og-fylkeskommunal-okonomi---november-2017/id2580410/. (Accessed: 2018-03-31)

NOU 2011:14. (2011). *Bedre integrering - mål, strategier, tiltak.* https://www.regjeringen.no/no/dokumenter/nou-2011-14/id647388/. (Accessed: 2018-01-12)

NOU 2017:2. (2017). *Integrasjon og tillit - langsiktige konsekvenser av høy innvandring.* https://www.regjeringen.no/en/dokumenter/nou-2017-2/id2536701/. (Accessed: 2018-04-19)

Ortega, F. & Peri, G. (2012). The Role of Income and Immigration Policies in Attracting International Migrants. *IZA Discussion Paper Series*(6655).

Sarvimäki, M. & Hämäläinen, K. (2016). Integrating immigrants: The impact of restructuring active labor market programs. *Journal of Labor Economics*, *34*, 479–508.

Schüller, S. (2016). *Ethnic enclaves and immigrant economic integration.* 287 doi: 10.15185/izawol.287. IZA World of Labor.

SSB. (2014-04-24). *Immigrants and norwegian-born to immigrant parents, 1 january 2014.* https://www.ssb.no/befolkning/statistikker/innvbef/aar/2014-04-24#content. (Accessed: 2018-03-30)

SSB. (2017-03-30). *60 per cent of immigrants are employed.* https://www.ssb.no/en/arbeid-og-lonn/artikler-og-publikasjoner/60-per-cent-of-immigrants-are-employed. (Accessed: 2018-03-30)

SSB. (2017-11-08). *Innvandrere i og utenfor arbeidsmarkedet.* https://www.ssb.no/arbeid-og-lonn/artikler-og-publikasjoner/innvandrere-i-og-utenfor-arbeidsmarkedet. (Accessed: 2018-08-26)

SSB. (2017-11-23). *Norwegian for adult immigrants.* https://www.ssb.no/en/nopplinnv. (Accessed: 2018-03-18)

SSB. (2017-12-19). *Population and land area in urban settlements.* https://www.ssb.no/en/beftett/. (Accessed: 2018-03-22)

Stock, J. & Yogo, M. (2005). *Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments* (Vol. 6). Cambridge University Press.

The Norwegian Directorate for Education and Training. (2011). *Det felles europeiske rammeverket for språk.* https://www.udir.no/Upload/Verktoy/5/UDIR_Rammeverk_sept_2011_web.pdf?... (Accessed: 2018-03-19)

The Norwegian Directorate of Immigration. (2014). *Et marginalt problem? Asylsøkere, ulovlig opphold og kriminalitet.* https://www.udi.no/en/statistics-and-analysis/research-and-development-reports/asylsokere-ulovlig

-opphold-og-kriminalitet---et-marginalt-problem/. (Accessed: 2018-03-31)

Wooldridge, J. M. (2016). *Introductory Econometrics: A Modern Approach* (Vol. 6th). Cengage Learning.

Yao, Y. & van Ours, J. C. (2015). Language skills and labor market performance of immigrants in the Netherlands. *Labour Economics*, *34*, 76–85.

A. About the Training

Whereas Chapter 3 briefly presented the Norwegian language training and social studies, this chapter describes the training in a detailed manner.

Right and obligation to participate

The Norwegian language training and social studies distinguishes between the right and the obligation to enroll in the training. According to Introduksjonsloven (2003, §17), the former indicates the right to training free of charge—a right that ceases to apply if the training is not completed within three years. On the other hand, if an immigrant is obliged to participate, she does not have the right to free training and the municipality may thus demand a payment.[57] Additionally, the immigrant is expected to participate in the training.

As of January 1st 2017, immigrants aged 55-67 years old are obliged to participate in the training. Until 2017, immigrants facing an obligation to participate comprised labour and education immigrants between 16 and 55 years of age originating from outside the European Economic Area (EEA), as well as their families. After the reform, however, this group was expanded to 16-67 years old. The reform is nevertheless not relevant in this thesis as it is restricted to 2014-2016.

Responsibility of the Municipalities

The municipalities are accountable for the Norwegian language training and social studies. They may organise the training through a municipal adult education facility, or they may purchase the service of private accredited bidders.[58] Moreover, the *introduction act* allows inter-municipal cooperation in arranging the training. The arrangement of the training is to be done as soon as possible, and no later than three months after the immigrant's settling in the municipality. It is the municipality's duty to initiate the training, but the immigrant may nonetheless request the commencement of training in the absence of a municipal proposal. Immigrants with the obligation to enroll, however, are required to initiate the training.

[57] The cost for the immigrant is related to the execution of the training as well as school equipment.
[58] Kompetanse Norge (2017-08-31) provides a list of such bidders.

Content and Duration

In 2012, the duration of the training was changed from a minimum of 300 to a minimum of 600 hours.[59] The principal reason behind the reform was reports from municipalities claiming that many immigrants needed more than 300 hours of training (Ministry of Children and Equality, 2010-2011, p.3). Immigrants granted residence permits before the reform follow the old arrangement. Labour immigrants from outside EEA still receive 300 hours.

Developed according to the European Centre for Modern Languages of the Council of Europe, the curriculum for the language training defines four levels of language proficiency; A1, A2, B1 and B2 (The Norwegian Directorate for Education and Training, 2011, p.26). Since immigrants have different linguistic and educational backgrounds, 600 hours of training will produce different results. As a consequence, the municipalities arrange the language training in three different streams accounting for individual adaption and progress. Organised on the basis of educational level, the streams comprise participants with little or no schooling, participants with some schooling, and lastly, participants with a good level of general education.[60]

As of September 2013, the Parliament introduced final tests in respectively Norwegian and social studies in order to attain the advisory levels for each stream. The goal is for 90% of the candidates to pass the social studies test. In 2015, 78% passed the test (Ministry of Education and Research, 2015-2016, p.60). The language test is an indicator of proficiency level, as discussed above. Kompetanse Norge is accountable for the execution of the tests. At the time of completion or cancellation[61] of the training, the participant is handed a diploma of participation.

Language Language Training and Social Studies Rather Than Introduction Programme

The *introduction programme* is the second component of the *introduction act* and consists of several measures beyond language training and social studies. However, available data does not allow to control for such measures at the municipal level. Consequently, analysing the Norwegian language training and social studies instead, allows for a more precise assessment

[59] An hour is 45 minutes of class.

[60] As of completion, the advisory level for the first stream is A2 or B1 orally, and A1 or A2 written. The anticipated level for the second stream is A2 or B1 for both speaking and writing, whereas the training estimates an overall level of B1 for the third stream.

[61] Cancellation implies dropping out of commenced training. (Ministry of Justice and Public Security, 2016, p.24)

of the causal effect of language training on the employment of immigrants. The complexity of the introduction programme also presents issues regarding the composition of the measures. While language hours is the only endogenous factor in Norwegian language training and social studies, the highly individualised plans in the introduction programme result in several possible changeable factors as several measures can be increased or decreased, or new measures implemented.

Moreover, the freedom of the municipalities in the *introduction programme* poses a challenge. To participate, the immigrant must be considered in need of qualification. The discernment is made by individual municipal employees and complicates the attempt of a clear group of potential participants. In order to avoid this becoming an issue, particularly rich data must be available, which is not the case.

Based on the discussion above, Norwegian language training and social studies is considered to be the component most suitable for empirical analysis.

B. Data and Descriptive Statistics

This chapter is an extension of the discussion in Chapter 4. First, it discusses choices made about the dependent variable. Then the chapter presents an array of descriptive statistics to back up the discussion and arguments made in the thesis.

Dependent Variable

Data for the dependent variable contains information on two different age groups. The first group encompasses people aged between 15 and 74, whereas the second group covers people from 20 to 66. On one hand, the key independent variable, includes immigrants of all ages, as presented in Section 4.2. Hence, this argues for selecting the first group. The second group, on the other hand, deviates less from the age limits for eligibility of participation. This might indicate that the second group fits better as an approximation. However, SSB (2017-03-30) portrays a far lower employment rate for immigrants for the age groups 15-19 and 66-74 years old in 2016 than other age cohorts. Moreover, in January 2014 only 8% of the immigrant population was older than 60 (SSB, 2014-04-24). The deviation for the group of immigrants 15-74 years old is ergo numerical limited. For this reason, the thesis considers the group of immigrants between 15 and 74 years of age. The robustness check in Section 10.4 concludes however that estimates are similar across the two different age groups.

Furthermore, SSB data categorises immigrants' country of origin in two groups. The first group comprises the EEA, North-America, Australia and New Zealand. The second group contains a wider group of regions, covering European countries outside the EEA, as well as Asia, Africa, South- and Central America, and Oceania except Australia and New Zealand. Immigrants from the EEA are not covered by the *introduction act* unless they are migrating as refugees and SSB numbers show an insignificant inflow of refugees from the EEA during the years in question.[62] However, this group also comprises labour immigrants from North-America, Australia and New Zealand. Labour immigrants from these countries are in fact eligible for participation. Yet, considering that the dependent variable represents the employment of immigrants eligible for training, and keeping in mind the requirements for eligibility, the first group is excluded.

[62] On average, EEA refugees make up only 0.2% of EEA immigrants.

Descriptive Statistics

Table B.1: Norway's four larges urban areas, 2014-2016

Urban area	Municipalities	Population 2014	Population 2015	Population 2016
Oslo	Asker, Bærum, Lier, Lørenskog, Nittedal, Oppegård, Oslo, Røyken, Rælingen, Skedsmo, Ski, Sørum	942 084	958 3783	975 744
Bergen	Bergen	251 281	250 420	252 772
Stavanger/Sandnes	Randaberg, Sandnes, Sola, Stavanger	207 439	210 874	213 313
Trondheim	Trondheim	172 226	175 068	177 617

Note: Data from NSD

Table B.2: Scope of missing data

Variable	Missing values	Share of variable
employed immigrants	171	13.4 %
participants	110	8.6 %
total immigrant population	0	0 %
production index	188	14.7 %
income	0	0 %
expenses	272	21.3 %
Sum	741	9.7 %

Table B.3: Correlation matrix

Variable	ln(employment)	ln(participants)$_{-1}$	ln(totimmpop)	prodindex	ln(revenue)	ln(expenses+1)	urban	left	right	list
ln(employment)	1.000									
ln(participants)$_{-1}$	0.871	1.000								
ln(totimmpop)	0.978	0.85	1.000							
prodindex	-0.298	-0.185	-0.319	1.000						
ln(revenue)	0.26	0.082	0.277	-0.277	1.000					
ln(expenses+1)	0.125	0.144	0.093	-0.056	0.091	1.000				
urban	0.444	0.35	0.446	-0.174	0.353	0.07	1.000			
left	0.1712	0.182	0.149	-0.065	-0.045	0.091	0.012	1.000		
right	0.201	0.122	0.217	-0.047	0.211	0.036	0.174	-0.492	1.000	
list	-0.104	-0.081	-0.116	0.056	-0.158	-0.261	-0.05	-0.206	-0.099	1.000

Table B.4: Participants from the EEA in training

Country	Participants 2014	Participants 2015	Participants 2016
Bulgaria	9	12	12
Croatia	56	41	28
Czech Republic	6	6	4
Estonia	9	11	12
France	12	11	10
Germany	34	30	29
Greece	4	7	8
Hungary	5	9	6
Ireland	6	5	-
Italy	4	4	5
Latvia	15	18	16
Lithuania	21	18	16
Netherlands	11	7	4
Poland	51	50	51
Portugal	7	4	-
Romania	26	23	20
Slovakia	-	-	4
Spain	22	20	13
UK	33	39	36
Sum EEA	331	315	274
EEA share of total participation	0.009	0.008	0.007

Note: Data from SSB. The table excludes countries in the EEA without participants. Data does not distinguish between refugees and other types of immigrants when counting participants.

Table B.5: Between- and within variation

Variable	Variation	Mean	Std. Dev.	Observations
employed immigrants	overall	470.873	2797.38	N = 1088
	between		2772.27	n = 370
	within		50.151	T-bar = 2.941
participants	overall	104.911	327.639	N = 1074
	between		322.36	n = 370
	within		20.743	T-bar = 2.903
total immigrant population	overall	1771.698	8741.757	N = 1110
	between		8745.138	n = 370
	within		280.713	T-bar = 3
prodindex	overall	3.269	9.123	N = 1020
	between		8.355	n = 363
	within		4.459	T-bar = 2.81
income	overall	478975.6	49697	N = 1110
	between		49197.23	n = 370
	within		7453.601	T = 3
expenses	overall	160467.6	79567.84	N = 935
	between		67037.86	n = 340
	within		43773.89	T-bar = 2.75
urban	overall	0.046	0.209	N = 1110
	between		0.209	n = 370
	within		0	T = 3
left	overall	0.463	0.499	N = 1110
	between		0.414	n = 370
	within		0.278	T = 3
right	overall	0.238	0.426	N = 1110
	between		0.361	n = 370
	within		0.226	T = 3
list	overall		0.201	N = 1110
	between		0.176	n = 370
	within		0.098	T = 3

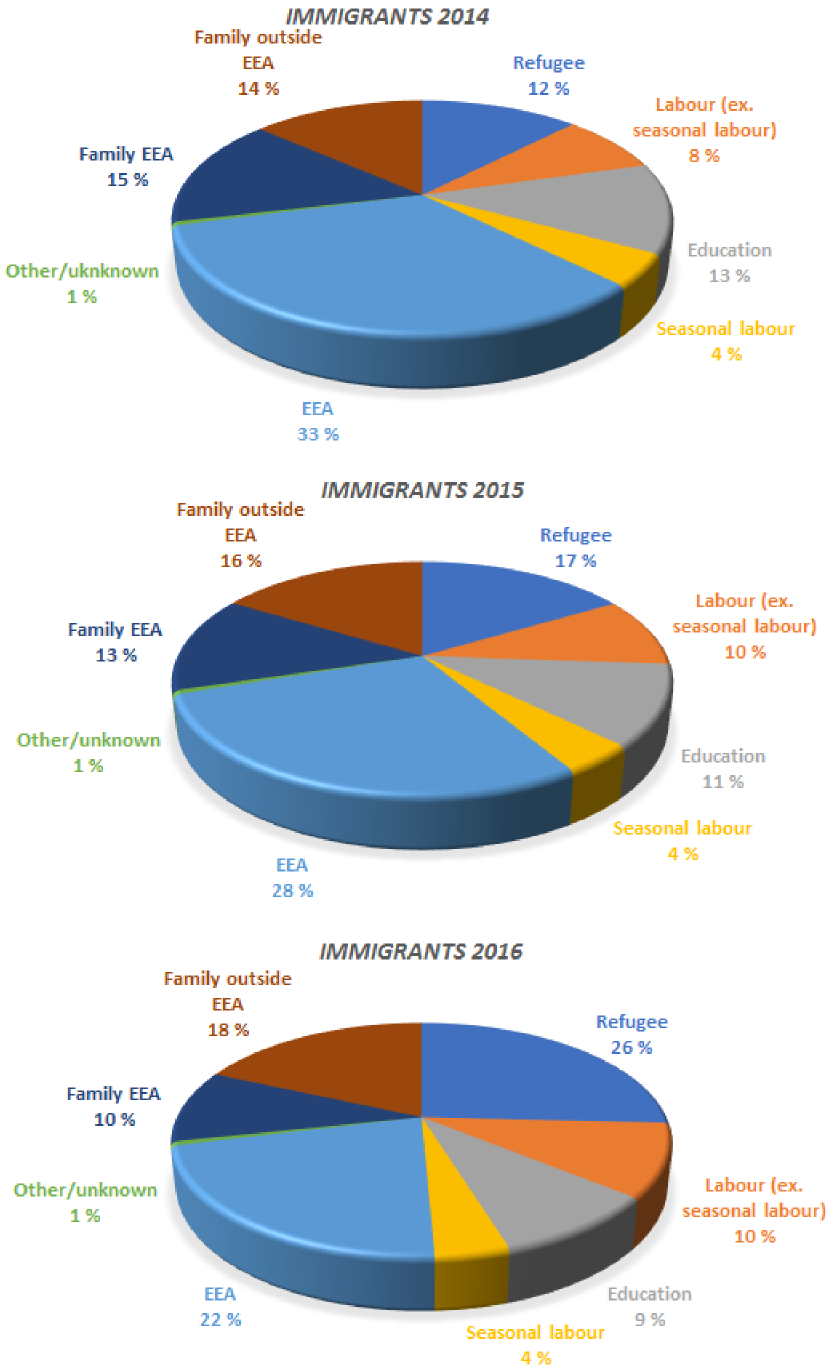


Figure B.1: Immigrants by reason for migration 2014-2016. Data from SSB.

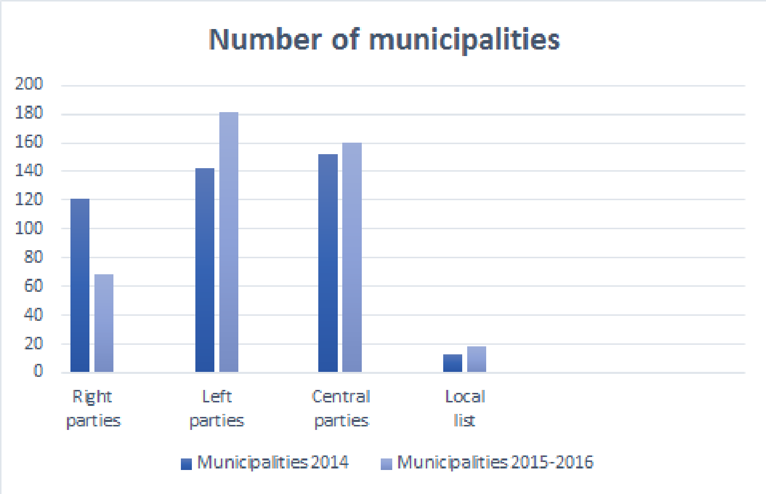


Figure B.2: Political parties in local governance, 2014-2016. Data from NSD.

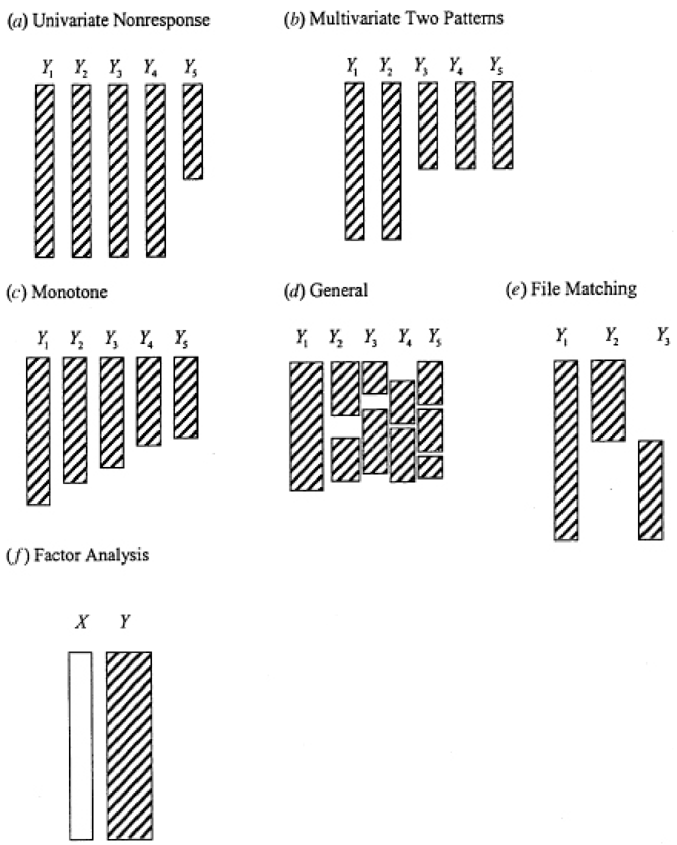


Figure B.3: Missing data patterns. Source: Little & Rubin (2002, p. 5)

C. Econometric Framework

This chapter provides a comprehensive examination of the econometric framework used in Chapter 6 and 7. First, it discusses properties of panel data and OLS, and the validity of OLS in the context of this thesis. Next, the chapter explores the framework of the main analysis which includes cluster-robust standard errors, random effects and fixed effects. The Hausman test referred to in Chapter 8 is also thoroughly explained.

Panel Data and OLS

The available data enables the construction of a balanced panel data set. Such samples, often called longitudinal, contains observations over time for the same cross-sectional units. Units may be individuals, firms, countries and so forth (Wooldridge, 2016, p. 9). In this thesis, data contains information for the same municipalities in different years. The panel data structure allows for examining variation in employment of immigrants within each municipality over time. Simultaneously, it enables the analysis of variation between municipalities.

OLS obtains estimates by minimising the sum of squared residuals. Although panel data is useful for combating endogeneity, a number of assumptions need to hold in order for the OLS estimators to be consistent and unbiased.[63] BLUE estimators require additional suppositions.[64]

Prior to describing the premises for consistency, unbiasedness and efficiency, the chapter turns to the statistical properties of OLS.[65] Assume the following simple model for panel data:

$$y_{it} = \beta X_{it} + u_{it} \quad , \quad i = 1, ..., N \quad , \quad t = 1, ..., T \tag{C.1}$$

The dependent variable on the left hand side is immigrant employment. X_{it} is a vector consisting of the key independent variable and socioeconomic variables, while β expresses the associated vector of coefficients. The subscript i indicates municipalities as cross-sectional units, whereas t denotes yearly time units.

The stochastic error term u_{it}, which also varies over time and between units, can be

[63] An estimator, $\hat{\beta}$, is consistent if the probability converges to the true population value as the sample size increases, $plim\hat{\beta} = \beta$. The estimator is unbiased for β if the expected value of $\hat{\beta}$ is the true value, $E(\hat{\beta}) = \beta$.

[64] An estimator, $\hat{\beta}$, is efficient if it, for a given sample, has the least variance among all the k unbiased estimators, $Var(\hat{\theta}_1) < Var(\hat{\theta}_{k-1})$.

[65] The reader should note that properties are not restricted to a particular sample (Wooldridge, 2016, p. 73).

decomposed as:

$$u_{it} = \eta_i + \varepsilon_{it} \tag{C.2}$$

The decomposed error term encompasses a unit specific component, η_i, and an idiosyncratic component, ε_{it}. The unit specific term captures all unobserved time invariant characteristics between units, while the idiosyncratic term captures all unobserved variation across units as well as time variation.

In pursuance of consistent and unbiased OLS estimators, the following assumptions must hold:

1. The model is linear in the parameters.

2. The sample is randomly drawn.

3. No perfect collinearity in the explanatory variables.

4. Zero conditional mean, thus indicating strictly exogenous explanatory variables, $E(u_{it}|X_{it}) = 0$.

However, BLUE estimators necessitates the following assumptions:

$$E(\varepsilon_{it}\varepsilon_{js}|X_{it}) = \begin{cases} \sigma_\varepsilon^2 & i = j \text{ and } t = s \\ 0 & \text{otherwise} \end{cases} \tag{C.3}$$

$$E(\eta_i\eta_j|X_{it}) = \begin{cases} \sigma_\eta^2 & \text{i=j} \\ 0 & \text{otherwise} \end{cases} \tag{C.4}$$

$$E(\varepsilon_{it}\eta_i|X_{it}) = 0 \tag{C.5}$$

Identity (C.3) ensures that the idiosyncratic error term component has a constant variance, that is homoskedasticity, and no serial correlation. In the same manner, identity (C.4) assumes homoskedasticity across all cross-sectional units and no serial correlation. Finally, identity (C.5) presupposes independence between the two error term components, implying no correlation.

A breach on the above-mentioned suppositions causes the estimators not to be BLUE, but they may be consistent and unbiased all the same. In such an event, OLS exploits variation between the cross-sectional units, and variation within the units across time.

Model Validity

This section presents the diagnostic tests discussed in Chapter 6. Table C.1 reports results from the diagnostic tests concerning three selected specifications from the OLS regressions presented in Table 6.1 in Chapter 6.1. Specification (2) includes the number of participants and total immigrant population. Specification (5) further includes the production index, median income and municipal expenses. Specification (6) includes the full set of control variables, in addition to the dummy for urban area and dummies for political parties. All tests are evaluated at the 5% significance level.

A general White test is performed to detect heteroskedasticity. The test adds squares and cross products of all independent variables in a regression of the squared residuals from OLS estimation. As a result, it uses many degrees of freedom. The null hypothesis of homoskedasticity is rejected in specification (2). On the other hand, the White test appears to state homoskedasticity in the other specifications. Although the White test has a large number of degrees of freedom and it tends to over-reject, it keeps the null hypotheses. An explanation might be the large sample size in this thesis. This might be an indication that adding control and dummy variables has a moderating effect on possible heteroskedasticity in the model.

A general Arellano-Bond test with one lag is executed to identify serial correlation. Although initially proposed for a Generalised Method of Moments, the test is applicable in several contexts such as the panel data OLS of the basic analysis. The Arellano-Bond test strongly rejects the null hypothesis of no serial correlation in all specifications and seems to provide evidence of correlated residuals across time. As a result, OLS estimation is inefficient.

Table C.1: Results diagnostic tests

Test	Property	Test value	Significance	Rejection of H_0
Specification (2)				
Test	Property	Test value	Significance	Rejection of H_0
White	Heteroskedasticity	$\chi^2(8) = 29.84$	$Prob > \chi^2 = 0.0002$	Yes
Arellano-Bond	Serial correlation	$z = 15.73$	$Prob > z = 0.0000$	Yes
Specification (5)				
Test	Property	Test value	Significance	Rejection of H_0
White	Heteroskedasticity	$\chi^2(26) = 37.28$	$Prob > \chi^2 = 0.0706$	No
Arellano-Bond	Serial correlation	$z = 14.16$	$Prob > z = 0.0000$	Yes
Specification (6)				
Test	Property	Test value	Significance	Rejection of H_0
White	Heteroskedasticity	$\chi^2(55) = 64.28$	$Prob > \chi^2 = 0.1834$	No
Arellano-Bond	Serial correlation	$z = 14.09$	$Prob > z = 0.0000$	Yes

Note: Each panel represents a separate specification. The tests are performed on regressions that employ year specific effects except panel d. Chosen significance level is 5%.

Random Effects

This section offers a technical approach to random effects estimation, as discussed in Chapter 7, and is only supplementary. Although the random effects analysis depends on the assumption of strict exogeneity between the error term and each explanatory variable, it should be noted that there is no need for panel data in the case of an uncorrelated unobserved effect as estimators will be consistent by applying single cross-section. However, such a specification easily disregards the issue of serial correlation which section 6.2.2 proved to be present in the OLS analysis. This is easier seen by decomposing the error term as:

$$u_{it} = \eta_i + \varepsilon_i \tag{C.6}$$

The composite error term will be serially correlated because the unobserved effect, η_i, will be present in each time period. However, employing a generalised least squares transformation would solve the issue. In order to take advantage of the decomposed error term and achieve efficient estimators, random effects require constant unconditional variance across time in the idiosyncratic component and no serial correlation in this error. Formally, this is written as

$$E(\varepsilon_{it}^2) = \sigma_\varepsilon^2, t = 1, 2, ..., T \tag{C.7}$$

$$E(\varepsilon_{it}\varepsilon_{is}) = 0, \text{all } t \neq s \tag{C.8}$$

Under these assumptions, the random effects estimator is given as:

$$\boldsymbol{\beta}_{RE} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i^{'}\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X}_i\right)^{-1}\left(\sum_{i=1}^{N} \boldsymbol{X}_i^{'}\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{y}_i\right) \tag{C.9}$$

The vector $\boldsymbol{X_i}$ represents a matrix of explanatory variables, whereas $\boldsymbol{y_i}$ is a vector for the dependent variable. The variance-covariance matrix for each cross unit, $\boldsymbol{\Omega}$, can further be written with constant variance as:

$$\boldsymbol{\Omega} = E(\boldsymbol{u}_i\boldsymbol{u}_i^{\prime}) = \begin{bmatrix} \sigma_\eta^2 + \sigma_\varepsilon^2 & \sigma_\eta^2 & \dots & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 + \sigma_\varepsilon^2 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\eta^2 & \dots & \dots & \sigma_\eta^2 + \sigma_\varepsilon^2 \end{bmatrix} \tag{C.10}$$

The GLS transformation can be shown by defining the transformative parameter as:

$$\theta = 1 - \left[\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_\eta^2}\right]^{1/2} \tag{C.11}$$

The transformation subtracts a fraction of the time average of variables for each cross-sectional observation, given as:

$$y_{it} - \theta\bar{y}_i = \beta_0(1-\theta) + \beta_1(x_{it1} - \theta\bar{x}_{i1}) + \beta_k(x_{itk} - \theta\bar{x}_{ik}) + (u_{it} - \theta\bar{u}_i) \tag{C.12}$$

If the transformative parameter is close to zero, random effects estimates resemble the OLS estimates. On the contrary, a transformative parameter adjacent to one implies estimators similarities between random effects and fixed effects. Chapter 8 showed the random effects estimates for $\ln(participants)_{-1}$ to be closer to those of fixed effects estimation than OLS. Consequently, one could assume the parameter θ is close to one.

Fixed Effects

This section shows explicitly the fixed effects transformation that eliminates correlated unobserved effects. To illustrate this, the benchmark model 5.2 is replicated with a decomposed error term:

$$\ln(employment)_{it} = \beta_0 + \beta_1 X_{it} + \delta_1 D_i + \eta_i + \varepsilon_{it} \tag{C.13}$$

The vector X_{it} includes the key independent variable and all socioeconomic variables, while the vector D_i includes all dummy variables. The unobserved effect is found in the municipality-specific component η_i. The first step is to transform the identity to averages and obtain the cross-sectional equation:

$$\ln(empl\bar{o}yment)_{it} = \beta_0 + \beta_1 \bar{X}_{it} + \delta_1 \bar{D}_i + \eta_i + \bar{\varepsilon_{it}} \tag{C.14}$$

$$\ln(empl\bar{o}yment)_{it} = \frac{\sum_{t=1}^{T} \ln(employment)_{it}}{T} \tag{C.15}$$

$$\bar{X}_{it} = \frac{\sum_{t=1}^{T} X_{it}}{T} \tag{C.16}$$

$$\bar{\varepsilon_{it}} = \frac{\sum_{t=1}^{T} \varepsilon_{it}}{T} \tag{C.17}$$

Executing OLS on equation C.14 generates the between estimator. Stopping at this point would, however, ignore crucial information for how the variables behave over time. In order to obtain the transformed fixed effects equation, the cross-sectional equation is subtracted from the original identity, producing

$$\ln(employment)_{it} - \ln(empl\bar{o}yment)_i = \beta(X_{it} - \bar{X}_i) + \varepsilon_{it} - \bar{\varepsilon}_i \tag{C.18}$$

By eliminating variation between the municipalities, fixed effects transformation gets rid of the unobserved effect and OLS on this equation will produce consistent and unbiased estimators. The dummy variables are also omitted because they yield no time variation during the time period of this thesis.

Hausman Test of the Effects Estimations

Chapter 8 presented results estimated with random and fixed effects. Although estimates remained stable across estimations in terms of the key independent variable, coefficients for the total immigrant population and median household income do change eminently. It is therefore interesting to evaluate which estimation approach is the better choice. Hence, this section executes a Hausman test to easier differentiate between the estimations.

In spite of having certain similar assumptions,[66] fixed effects is generally considered a more plausible method for estimating causality than random effects.[67] The reason is that fixed effects enables arbitrary correlation between explanatory variables and the unit specific error component. On the contrary, fixed effects is useless in cases where the key independent variable is constant across time. Random effects would be appropriate in such a case.

The Hausman test investigates orthogonality between regressors and unobserved effects. The null hypothesis is that the preferred estimation method is random effects versus the alternative hypothesis of fixed effects being preferred. The former implies no correlation between explanatory variables and the unobserved effects, while the latter indicates such a relationship:

$$H_0 : E(\eta_i|X) = 0 \quad vs. \quad H_1 : E(\eta_i|X) \neq 0. \tag{C.19}$$

In the event of failing to reject the null hypothesis, estimates from random and fixed effects estimation are too close to differentiate. An additional reason could be eminently large sampling variation in the fixed effects estimation. Consequently, random effects is the most suitable estimation method. In the case of rejecting the null hypothesis, fixed effects is the preferred specification for the data. However, the Hausman test does not permit the presence of cluster-robust standard errors. An alternative approach would be the Mundlak approach which may be used in the existence of cluster-robust standard errors. First, the procedure averages the panel level of the independent variables. Second, it employs a random effects estimator in a regression of all control variables and panel average variables against the dependent variable.

[66] Appendix 14A in Wooldridge (2016, p. 457-460) presents the many assumptions that make up the foundation for random and fixed effects.

[67] Likewise, random effects is commonly suggested in lieu of OLS due to more efficient estimators.

The null hypothesis is that the variables created in the first step are jointly zero. Rejecting the null hypothesis suggests random effects to be used.

The Mundlak approach strongly rejects the null hypothesis of random effects being the better approach. Table C.2 shows a very significant test value of 129.26. Consequently, fixed effects appears to be the appropriate estimation method.

Table C.2: Hausman test for fixed or random effects

Test	Property	Null hypothesis	Significance	Test value	Rejection of H_0
Hausman (Mundlak approach)	Random or fixed effects	$H_0 : E(\eta_i \vert X) = 0$	$Prob > \chi^2 = 0.0000$	$\chi^2(8)$ = 129.26	Yes