The Return to Big City Experience: Evidence from Danish Refugees*

Fabian Eckert	Mads Hejlesen	Conor Walsh
Yale University	Aarhus University	Yale University

First Version: November 2017
This Version: July 2018

Abstract

Using a quasi-random settlement policy for refugees in Denmark between 1986-1998, we provide evidence for steeper returns to experience in big cities. Exploiting exogenous variation in initial placement, we show that the slope of an individual's lifetime wage path depends strongly on placement in the country's capital, Copenhagen. Conditional on observables, settled refugees initially earn similar hourly wages across regions, but those placed in Copenhagen see their wages grow 0.81% faster than others with each year of experience they accumulate. We further show that this premium is driven by greater acquisition of experience at high-wage establishments, and differential sorting across occupations. Estimating a spatial model of earnings dynamics reveals that sorting on unobserved ability within cities plays an important role in explaining observed patterns.

*We thank Giuseppe Moscarini and Rune Vejlin for their continued guidance and support with this project. We also thank Costas Arkolakis, Anna Piil Damm, Gilles Duranton, Philipp Kircher, Samuel Kortum, Costas Meghir, David Nagy, Michael Peters, Esteban Rossi-Hansberg and Fabrizio Zilibotti for helpful comments and Sharon Traiberman for help with the system of occupational codes. Additionally, we are grateful to audiences and discussants at the SAET 2018, SED 2018 and the IEB Summer School 2018. This paper started while Eckert and Walsh were visitors at the Economics Department of Aarhus University, the hospitality of whom we gratefully acknowledge. All errors are our own.

1 Introduction

A large literature has documented a substantial urban wage premium, even after controlling for worker level observables, both for the United States and for a host of other countries around the world.[1] Three main hypotheses have been proposed to explain these higher nominal wage residuals in cities: (i) spatial selection on unobservables, (ii) static advantages of working in bigger cities, and (iii) dynamic advantages of working in bigger cities.[2] Any attempt at quantifying the importance of (ii) and (iii) hinges crucially on being able to control for (i).

In this paper, we pursue a novel approach to control for selection on unobservables in the measurement of the urban wage premium for a specific population. In particular, we exploit a large natural experiment conducted in Denmark between 1986-1998 (see Damm and Dustmann (2014)). Throughout this period, the Danish Government allocated 80,000 newly arriving refugees across its 271 municipalities, in proportion to local populations, and without regard to labor market relevant characteristics of refugees. Further, we find that initial allocations were very persistent, with around 80% of refugees still at their initial location of assignment after 20 years in Denmark. The randomness of the initial assignment and its persistence make this an ideal setting for studying the static and dynamic effect of locations on refugees' wages, mitigating concerns of initial selection on unobservables.

In contrast, recent studies of the urban wage premium have leveraged the panel structure of large European administrative datasets to control for unobserved heterogeneity using person fixed effects.[3] With such fixed effects in place the identification of city effects relies uniquely on agents that choose to relocate across locations. Selection into moving induces a potential source of bias in the resulting estimates of both the static and dynamic city effect. There is ample empirical evidence for the importance of selection in individual migration decisions (e.g. Borjas (1987), McKenzie et al. (2010), Young (2013), Lagakos and Waugh (2013)).

To proceed, we split the spatial economy of Denmark in two parts: the commuting zone of Copenhagen, the country's capital, and the union of all remaining commuting zones. We then document the static and dynamic treatment effects of initial assignment on refugees' hourly wages. We find no significant evidence for a static urban difference for refugees' hourly wages. However, individuals

[1]See Glaeser (2011), Carlsen et al. (2013), Glaeser and Mare (2001), Fu and Ross (2010), Rosenthal and Strange (2004) and Puga (2010) for some recent empirical evidence.

[2]In practice (i) and (iii) are likely complementary and mutually reinforce one another. In particular sorting can be driven by differential agglomeration effects across different latent skill groups as discussed in Behrens et al. (2014).

[3]De La Roca and Puga (2017) use Spanish administrative data to document the existence of both a sizable static and dynamic city size premium for individual earnings which persist after controlling for unobserved heterogeneity using person fixed effects. D'Costa and Overman (2014) in turn run person fixed effect regressions in wage differences on a large panel of British workers and find no evidence for a dynamic premium, but do confirm the existence of a static one. Both papers are in the tradition of Glaeser and Mare (2001), who were the first to explore the urban wage premium in regressions with person fixed effects using US survey data.

initially settled in Copenhagen see their wages grow an extra 0.81% for every year of experience relative to their counterparts settled outside of Copenhagen. Additionally, we find that the treatment effect for labour supply differed markedly among education groups. Those with at least a high school degree saw no difference in employment rates, but those with less education were 4% less likely to join the labour force if assigned to Copenhagen.

Next, we leverage the scope of the Danish micro data to explore the mechanisms underlying the city wage growth premium. We find that a substantial part is driven by two compositional effects. First, the greater density of high-productivity establishments in Copenhagen leads individuals assigned there to accumulate more experience with these types of employers.[4] This experience is more valuable than experience collected at low-productivity establishments, and explains part of the dynamic urban experience premium. Second, we find that natives in Copenhagen work disproportionately in both high-skill and low-skill service occupations, and less so in manual occupations, relative to other local labor markets in Denmark. Likewise we observe marked, differential dynamic sorting patterns across occupations for refugees, depending on the region of initial assignment, which helps explain an additional substantive fraction of the city wage growth premium. We are able to dismiss several other mechanisms, including separate wage trends between locations, the effects of ethnic enclaves and variation in educational take-up across space.

Lastly, we account for the interaction of these mechanisms with unobservable ability in a structural model of earnings dynamics. We explore the possibility that assignment to a big city has differential effects conditional on ability, which manifest in the more frequent matching of high-ability workers with high-productivity establishments, and decreased job finding rates for low-ability workers. We find that this sorting was significant in explaining the observed patterns, accounting for more than 50% of the dynamic experience premium.

Literature on the Urban Wage Premium. A rich literature documents the existence and size of an urban wage premium in the cross-section which persists even after controlling for workers' observable characteristics.[5] In an early contribution, Glaeser and Mare (2001) document that average wages in cities in the U.S. with more than 1 million inhabitants are about 36% higher than in areas below this cutoff. This effect is reduced to about 11% once unobserved heterogeneity is controlled for by including person fixed effects. Glaeser and Mare (2001) are also the first to differentiate between a static and a dynamic component of the urban wage premium, reporting a wage gain of about 12% that is realized after five years in a city, with 8% of this gain accruing in the first year alone.

[4] To define high-productivity establishments we use the full sample of all Danes for any given year to calculate average wages paid at each establishment. Then we split the population of establishments into those whose average wage falls into the top three deciles of the national establishment distribution and the rest. These top decile establishments in each year are what we refer to as high-productivity establishments. The distribution of theses establishments across space is skewed towards Copenhagen.

[5] Rosenthal and Strange (2004) provide an excellent review of this literature. Other important papers in this field are Glaeser (2011), Carlsen et al. (2013), Glaeser and Mare (2001), Fu and Ross (2010), Rosenthal and Strange (2004) and Puga (2010).

In the latest contribution to this literature, De La Roca and Puga (2017) use a large administrative panel of Spanish workers to document the existence of a substantial urban earnings premium even after controlling for unobservable heterogeneity via person fixed effects. In line with our study, De La Roca and Puga (2017) provide evidence that a substantial fraction of this premium is dynamic in nature, and large in magnitude. Our failure to find systematic evidence for a static urban premium in hourly wages is in line with evidence from Combes et al. (2008) and Mion and Naticchioni (2009) who find moderate static premia in French and Italian matched employer-employee data after accounting for firm sorting across space.[6] None of these papers use a natural experiment or other form of exogenous variation to control for unobserved heterogeneity in location decisions.

There is also a nascent literature on the potential sources for a dynamic urban wage premium. Wheeler (2006) provides early evidence that the urban wage growth premium is due to between-job wage growth rather than within-job growth. Baum-Snow and Pavan (2011) use a structural model together with US data to conclude that sorting on unobserved ability within education group, differences in labour market search frictions and distributions of firm-worker match quality contribute little to observed city size wage premia. In a recent contribution Papageorgiou (2017) demonstrates theoretically, and provides corroborating empirical evidence, that larger occupational variety in cities allows workers to learn their comparative advantage faster, engendering faster wage growth in cities. Relative to this literature, our work highlights the importance of differences in the occupation and establishment distributions across space.[7] Our findings suggest that workers in cities are more likely to accumulate experience at higher productivity establishments and to progress into higher paying occupations as their careers advance.

Literature on Exogenous Placement Experiments. We are not the first to employ the exogenous variation associated with the Danish refugee dispersal policy in the economic literature. Damm and Dustmann (2014) are a recent example of an application involving the same data, in this case studying the effect of growing up in a high-crime neighborhood on the likelihood of children to themselves commit crimes. Other papers using the Danish Refugee Dispersal policy include Damm (2005), Damm (2009), Damm and Rosholm (2010), Foged and Peri (2013), and Damm (2014). We owe a substantial debt to this work. Furthermore, a similar experiment in Sweden has been exploited by Aslund and Rooth (2007), Edin et al. (2003) and others. Our paper is the first to use an exogenous placement experiment for the study of the size and nature of the urban wage premium.[8]

Literature on Refugee Assimilation. There is a small literature on the economic integration of

[6] Combes et al. (2008) find a moderate premium of 2% in the French data, while Mion and Naticchioni (2009) find a very small static premium of 0.2% in the Italian data. Given that extant differences in population (density) tend to be large the former will generate sizable differences in a way that the latter estimate will not.

[7] None of these papers uses matched employer employee data, which is what enables us to study the importance of the employer-occupation and employer-firm match, relative to this previous literature.

[8] Other notable papers in the exogenous placement literature are Aslund et al. (2009), Shoag and Carollo (2016), Peters (2017), Chetty et al. (2016), Beaman (2011), Edin et al. (2003), Gould et al. (2004), and Imberman et al. (2012).

refugees to which we contribute.[9] Chiswick (1978) and Borjas (1985) are early examples, and Constant and Massey (2002) more recent. This literature seems to agree that the assimilation process is slow and full economic assimilation for most refugees is unlikely even after a substantial time in the host country. As part of the small literature on the development of best practices for refugee assimilation and integration, Edin et al. (2004) provide an early evaluation of a similar refugee dispersal program implemented in Sweden. They report an overall negative effect on refugee earnings from the dispersal policy.[10]

To the best of our knowledge our paper is the first to document the causal effect of cities on refugees' economic assimilation. We show that the causal treatment effect of being assigned to a city on subsequent wage growth is substantial and is comparable in magnitude to the returns to an additional two years of education over a working life. This suggests that the success of assimilation is a function of the spatial distribution of refugees. Given that refugees sort mainly into big cities in the absence of dispersal policies (see Damm (2009)), this can explain the finding of Edin et al. (2004) that dispersal policies on average lower earnings of refugees.

Clearly, refugees are a singular population, and at first glance their usefulness for studying the sources of the urban wage premium is not clear. They may not know the language, have different preferences and have a need to integrate into the host society. However, the mechanisms we uncover driving the differential return to experience for those allocated to Copenhagen appear to be more general. We would expect the density of high-wage establishments and different occupation distributions to also impact the earnings of natives living in Copenhagen. While we can say less about their quantitative importance for natives, our results suggest that future studies of the urban wage premium should consider these mechanisms as fundamental sources of the dynamic premia that cities bestow.

Plan for the Paper. The remainder of the paper is structured as follows. While the spatial dispersal policy has been used before in the economic literature, Section 2 introduces the policy, our data sources and the sample selection used throughout the paper. Section 3 documents the treatment effect on hourly wages, earnings and labour supply induced by the initial allocation. Section 4 provides evidence of the importance of various observable mechanisms in explaining the dynamic and static treatment effect documented in Section 3. We quantify the importance of sorting on unobserved ability in a structural model of earnings dynamics in Section 5. Section 6 concludes.

[9] A much larger literature is concerned with the effect of refugees on natives' labor market outcomes. Altonji and Card (1991), Borjas (2003), Dustmann et al. (2008), Ottaviano and Peri (2012) and Burstein et al. (2017) are important examples.

[10] In addition, dispersal policies have been used to evaluate the effect of ethnic enclaves on immigrant assimilation and integration. Edin et al. (2003) reports a positive effect of ethnic enclaves for low skilled workers and a negative, yet insignificant effect of ethnic enclaves on high skill immigrants. Damm (2009) largely confirms this and additionally reports, utilizing the Danish experiment, that enclaves decrease the employment likelihood of high skilled workers and increase that of low skilled workers. She concludes with suggestive evidence of information networks that help low skill immigrants in finding employment.

2 Background, Data and Sample Selection

In this Section, we describe the Danish spatial dispersal policy. Next, we discuss the construction of our sample used for estimation before clarifying our treatment of the Danish geography. Lastly, we document the persistence of assigned locations for our sample.

2.1 The Danish Refugee Dispersal Policy

Our description here follows Damm and Dustmann (2014) and Damm (2009), which provide substantially more detail. Prior to 1986, refugees chose municipalities according to their own preference, which resulted in relatively few municipalities housing a large share of the refugee population. In 1986 the Danish government initiated a refugee dispersal policy aimed at distributing refugees across counties and municipalities in proportion to population size. The motivation was to ensure that each municipality housed a fair share of the refugees, rather than only a few localities performing the work of integrating refugees.

Upon arrival in Denmark, refugees were housed in Red Cross reception centers located across Denmark. After being granted asylum, refugees faced no legal impediments to labor market participation. Within 10 days of the asylum decision, refugees were assigned temporary housing in one of 15 counties in Denmark. Each county assigned the refugees to a municipality within the county and helped them find permanent housing.[11] When assigning refugees to a municipality within a county the local county council had access to the birth date, marital status, number of children, and nationality of the refugee.[12] This information was the only information available to the council upon assigning refugees to municipalities, and assignment was random conditional on this information. Importantly, the council did not have information on years of schooling or family income, and the council did not meet with the refugees in person. In the last two columns of Table 4 in Appendix A.1 we compare refugees assigned to Copenhagen and elsewhere, and confirm that years of schooling do not differ by initial assignment after controlling for information available to the council.

The refugees received social assistance for the first 18 months in the assigned municipality as well as participating in Danish language courses. The refugees were encouraged to stay in the initial assigned municipality, but they were not forced to stay. Damm (2005) concludes that the 1986-1998 refugee program indeed succeeded in assigning refugees to municipalities in proportion to local population sizes. Our empirical strategy exploits this initial exogenous variation in two steps. First, we study the effect of initial placement in an urban area on life time wage growth. In a second step, we use the persistence of the initial assignment to explore to what extent we can interpret these

[11] Damm (2005) reports that 90% of refugees was assisted in finding permanent housing.

[12] The council had a tendency to assign families with a large number of children to less populated municipalities as bigger houses were available in these municipalities.

dynamic treatment effects as the return to big city experience.

2.2 Data and Sample Selection

Our analysis relies on administrative data provided by Statistics Denmark. Our core dataset is a matched employer-employee panel covering the entire Danish population from 1986 to 2012, including all refugees from the time of being granted asylum. The dataset includes detailed labor market information such as annual average hourly wages, annual earnings, 4-digit occupation codes and current municipality of residence and work. For refugees it additionally includes country of origin and the year of arrival. A large set of basic characteristics such as gender, age, years of education, and family information on spouse, number of children, and age of youngest and oldest child is also available. Additionally, on the firm side, the dataset contains industry, employment and earnings by establishments, linked to the worker level. A complete description of all variables used can be found in Online Appendix B.1.

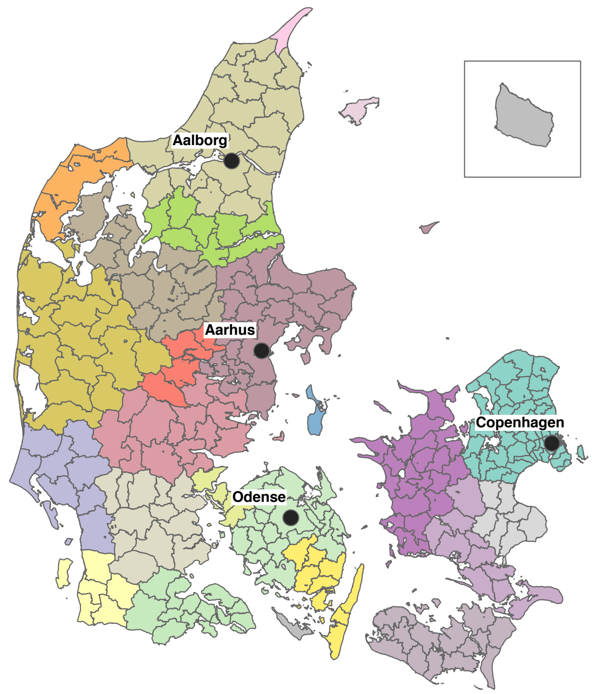
Following previous papers exploiting the same experiment (see e.g. Damm and Dustmann (2014)) we restrict our sample to refugees arriving from Lebanon, Iran, Iraq, Somalia, Sri Lanka, Vietnam, Afghanistan, and Ethiopia between 1986 and 1998. Further, our baseline sample contains only men between the age of 19 and 55.

The first two columns of Table 4 in Appendix A.1 show mean characteristics for refugees and natives respectively. Refugees are on average younger than the cross-section of Danes, have fewer children and are less likely to be married. Another feature of the data is that we are missing education information for around 19% of our sample. This arises from the fact that education information was collected from a survey after arrival for refugees, whereas educational information for Danes is based on administrative records.

2.3 The Geography of Denmark

In many ways, Denmark provides an interesting laboratory to examine how the returns to experience differ in dense urban agglomerations. Denmark has a single large city, Copenhagen, which houses some 2.3 million people in its broader metropolitan area. The remaining 3 million people are located either in three second-tier cities (Aarhus, Aalborg and Odense), smaller towns or rural areas. The second-tier cities are at least an order of magnitude smaller than Copenhagen, with the largest containing around 250,000 people (Aarhus). In our analysis, we exploit this structure in a simple manner: we divide the economy into two locations, the Copenhagen commuting zone and everywhere else. In the remainder of the paper, when we refer to Copenhagen, we refer to the Copenhagen commuting zone. Likewise, when we refer to Non-Copenhagen, we refer to the collection of all municipalities but the Copenhagen commuting zone.

Figure 1: 1986 Commuting Zones in Denmark



Note: This Figure shows the 23 commuting zones constructed by the authors based on 1986 cross-municipality commuting flows constructed from work and residence municipality identifiers in the IDA dataset.

Following Tolbert and Sizer (1996), we use commuting flows of all Danish workers between Denmark's 271 municipalities in 1986, together with a hierarchical clustering algorithm to construct 23 Danish commuting zones.[13] These zones are shown in Figure 1, with the Copenhagen commuting zone appearing as the collection of blue municipalities in the east.[14]

2.4 Persistence of Initial Allocation

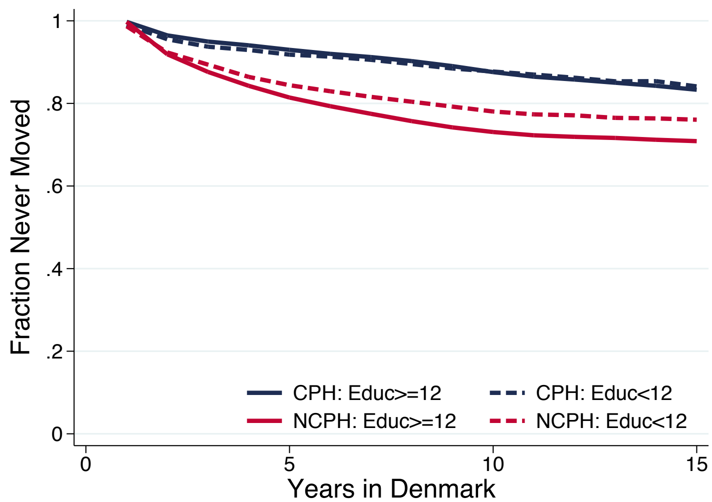
While assignment to a location is an interesting treatment in its own right, it is useful as a setting to study how the returns to working in a big city accumulate over time to the extent that it is persistent. For our exercise, the relevant definition of mobility is at the big city level. In Figure 2 we show the fraction of refugees who have never moved from their initial zone of assignment, which as above is either Copenhagen (CPH) or Non-Copenhagen (NCPH), by years spent in Denmark. After 15 years,

[13] For two municipalities, i and j, we define $similarity(i,j) = \frac{commuters(i,j)}{min(population(i), population(j))}$, where $commuters(i,j)$ equals the number of individuals living in municipality i and working in municipality j and vice versa. We use the average-linkage method to recompute similarity scores between municipalities after merging two municipalities. Following Tolbert and Sizer (1996), we use a cut-off of 0.985 to cut the resulting dendrogram of the hierarchical cluster algorithm, which results in 23 Danish commuting zones.

[14] In Appendix A.10 we additionally show the commuting zones that would result if 1980 or 2000 commuting flows were used for the underlying cluster analysis. We show that our main results are robust to the choice of commuting zone delineation.

78% have never lived in the other zone. We also find that refugees are less likely to leave Copenhagen if placed there, and that this does not differ by initial educational level. However, we do find that refugees with at least a high-school degree are more likely to move to Copenhagen than those with less education.

Figure 2: Persistence of Initial Assignment by Education Groups



Note: Years in Denmark is computed as the number of years since being granted asylum. Educ is Years of education at assignment. Those with missing educational information are categorized as less than 12 years.

We discuss the implications of this persistence for the interpretation of the treatment effects in Section 3.2.

3 The Returns to Big City Experience

In the first part of this Section, we compute the static and dynamic treatment effect of being initially assigned to Copenhagen on three outcome variables: wages, yearly earnings and the extensive margin of labor supply. The second subsection discusses the extent to which the dynamics treatment effect on wages can be interpreted as big city experience premia.

3.1 Static and Dynamic Treatment Effects

3.1.1 Wages and Earnings

We now examine the effect of initial allocation to Copenhagen on the statistical return to experience. Our exercise is conceptually simple: we segment our sample by initial allocation to Copenhagen or not, and construct lifetime wage-experience profiles. Our purpose here is *not* to determine the causal returns to experience, big city or otherwise. Instead, we ask how wage-experience profiles differ across two identical groups of refugees who have been exogenously allocated to different locations.

In a simple linear setting, there are two ways in which these wage-experience profiles could differ as a result of initial allocation. First, there could be different intercepts, such that one group earns uniformly higher hourly wages regardless of experience. We refer to this as the Static Treatment Effect. Second, there could be a difference in slopes, such that individuals in one group see their wages rise faster with experience. We term this the Dynamic Treatment Effect. These two objects are well-identified conditional on the randomness of the dispersal policy, and form the starting point for a deeper analysis of the return to big city experience in the rest of the paper. It is worth stressing that these are conceptually distinct objects from the usual discussion of static and dynamic effects of cities (e.g. Glaeser and Mare (2001)), since they only condition on initial assignment, not current location. They are specific to the context of our experiment; the extent to which they can be interpreted as "city premia" will be explored in the next subsection.

We begin by constructing experience for every refugee in our sample, incrementing years of experience Exp_{it} by 1 if individual i undertook paid employment in the previous year. We then estimate the following equation:

$$y_{it} = \beta_1 Exp_{it} + \beta_2 InitCph_i + \beta_3 \left(InitCph_i \times Exp_{it} \right) + \phi' \boldsymbol{X}_{it} + \epsilon_{it}. \tag{1}$$

Here y_{it} is either the log of the hourly wage in Danish Kroner, or the log of yearly earnings in Danish Kroner, deflated by an index of Danish nominal wage growth. This index removes aggregate trends in the level of wages due to inflation and aggregate productivity growth.[15] $InitCph_i$ is an indicator variable that takes a value of 1 if the refugee is allocated to Copenhagen upon arrival and is 0 otherwise. $\mathbf{X}_{it}$ is a vector of controls that include cohort fixed effects, nationality fixed effects, and the variables which were relevant to the assignment of refugees to Copenhagen or not.

We report the estimation for the specification for hourly wages conditional on working in Column (1) of Table 1. We find there to be no significant difference in initial wages across the two locations. However, each year of experience earns a refugee assigned to Copenhagen an additional 0.81% wage increase compared to the refugees placed outside the metropolitan area.

We then examine the treatment effect on two distinct sub-populations. In Column (2), we report the results for refugees who had at least a high-school degree upon arrival in Denmark. In Column (3), we do the same for refugees with either less than a high-school degree, or missing education

[15] To construct an index of nominal wage growth, we use the entire population of native workers and apply our sample selection criteria from Section 2.2. We take the average hourly wage across this population in each year and divide by the average in 1986. This gives an index number for each year of our sample, and we use this to deflate mean hourly wages and earnings for refugees. As an alternative, we also control for aggregate trends using time fixed effects, and find little difference.

Table 1: Wage- and Earnings-Experience Profiles

	(1) $logwage_{it}$	(2) $logwage_{it}$	(3) $logwage_{it}$	(4) $logearnings_{it}$	(5) $logearnings_{it}$	(6) $logearnings_{it}$
Exp_{it}	0.0228***	0.0250***	0.0211***	0.0759***	0.0784***	0.0750***
	(0.00142)	(0.00159)	(0.00126)	(0.00322)	(0.00370)	(0.00341)
$InitCph_i$	0.000477	0.00858	-0.00892	-0.0725***	-0.0543**	-0.104***
	(0.0101)	(0.00883)	(0.0136)	(0.0184)	(0.0190)	(0.0237)
$InitCph_i \times Exp_{it}$	0.00810***	0.00736***	0.00813***	0.0214***	0.0186***	0.0261***
	(0.00148)	(0.00163)	(0.00134)	(0.00303)	(0.00330)	(0.00312)
No. Kids at Arr.	-0.0199*	0.0143	-0.0213*	-0.0298*	0.0432	-0.0432*
	(0.00824)	(0.0117)	(0.0102)	(0.0138)	(0.0230)	(0.0193)
Married at Arr.	0.0401**	0.0210	0.0285*	0.0892***	0.0409	0.0888***
	(0.0120)	(0.0163)	(0.0121)	(0.0163)	(0.0217)	(0.0177)
Age at Arr.	-0.00143	-0.00228*	-0.00273***	-0.000111	-0.00167	-0.00263
	(0.000932)	(0.000841)	(0.000705)	(0.00220)	(0.00210)	(0.00186)
Observations	97402	57994	39408	107297	63870	43427
R^2	0.056	0.062	0.055	0.155	0.158	0.156
Sample	All	Educ≥12	Educ<12	All	Educ≥12	Educ<12
Nationality FE	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: Specification given in Equation (1), for different sub-samples. Missing education information is coded as Educ<12. Robust standard errors clustered at the level of initial commuting zone. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

information. The differential slope for wages is very similar across subgroups, and we continue to find no significant difference in initial wages.

In the right-hand panel of Table 1, we repeat these regressions with log earnings as the dependent variable. For earnings, a negative initial premium emerges for all three specifications: earnings are initially lower in the city for both of our sub-populations. In terms of magnitude, this negative static premium is equal to about 3.5 years of the dynamic premium for earnings, so that over the years the the dynamic premium comes to dominate. To be consistent with our results for hourly wages, it must be the case that conditional on working, refugees assigned to Copenhagen are initially working less hours. We examine hours worked non-parametrically in Appendix A.3. There we show that yearly hours worked increase significantly with experience for our population, and that an initial gap in hours worked between CPH and NCPH eventually closes.

In Appendix A.4 we repeat these treatment regressions using different spatial units as the basis for analysis. This yields intuitive results. First, we drop the refugees assigned to the three second-tier cities, and compare Copenhagen to the rural areas of Denmark. We continue to find no static effect, but a larger Dynamic Treatment Effect emerges. Concordantly, when we drop Copenhagen and repeat the regression comparing the second-tier cities to the rural areas, we find a much smaller dynamic effect.

Lastly, we examine how departures from linearity affect our conclusions. In Appendix A.5 we repeat our baseline regression non-parametrically, with dummies for 3-year experience bins. In line with a large literature (e.g. Lagakos et al. (Forthcoming)) we find concavity in the returns to experience. However, the treatment effect itself is broadly linear in years of experience, and the size of the effect after 20 years is well captured by the linear model. We continue to find no static effect for wages, and a small negative static effect for earnings.

While wages and earnings are our key outcome of interest, these are only observable for individuals who work. One driver of the dynamic treatment effect on wages and earnings could be differential selection into work across different locations. To study this issue, we turn to an analysis of the treatment effect of initial allocation on the extensive margin of labor supply, before offering interpretations of the dynamic treatment effect on wages and earnings.

3.1.2 Labour Supply

The other key outcome of interest in our study is the extensive margin of labour supply. This both complements and informs the above analysis of treatment effects on lifetime wage paths. In particular, we examine two primary channels through which assignment to a city might affect labour supply.

The first channel is the decision to work at all throughout a lifetime. In Table 2 we regress an indicator variable for never having worked by the last year in our sample (2012) on the assignment variables from specification (1), as well as indicator for assignment to Copenhagen. We run the regression stratifying the sample into those with at least a high-school degree in Column (1), and those with less than a high-school degree in Column (2).

We find that the effect of assignment to a city on the extensive margin differs markedly across education groups. For those with at least high school degree, assignment to Copenhagen did not significantly affect the chance that the refugee had never worked in our sample period. Moreover, the point estimate itself is tiny at 0.2%. However, those without a high-school degree saw both a significant and economically detrimental effect of being assigned to the city, raising their chance of never working by almost 4%.

The second is employment rates conditional on ever joining the labour force. In Table 6 in Appendix A.2 we regress an indicator for current employment on a full set of dummies for time spent in Denmark and the assignment variables, conditional on having worked at least one year in our sample. Employment rates rise steadily throughout our sample period for both groups, in both locations[16].

[16] However, it is worth noting even after a considerable amount of time spent in Denmark, beyond 15 years, employment rates for refugees remain about 12% below those of natives for the same age and gender.

Table 2: Labour Force Participation

	(1) Never Employed	(2) Never Employed
$InitCph_i$	0.00259	0.0371***
	(0.00882)	(0.0111)
Age at Arr.	0.0188***	0.0196***
	(0.000602)	(0.000539)
No. Kids at Arr.	0.0347**	-0.0285**
	(0.0126)	(0.0111)
Married at Arr.	-0.0691***	-0.0295*
	(0.0113)	(0.0140)
Observations	11138	9434
R^2	0.141	0.175
Sample	Educ≥12	Educ<12
Nationality FE	Yes	Yes
Cohort FE	Yes	Yes

Note: $NeverEmployed_i$ is an indicator taking a value of 1 if the individual never took up paid employment between 1986-2012. Missing education information is coded as Educ<12. Robust standard errors clustered at the level of initial commuting zone. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

For the purposes of our study, again we find somewhat different outcomes depending on education at arrival. For the educated, assignment to a city does not seem to have had any impact on employment rates once in the labour force. However, for those with less than a high-school degree, employment rates are 2% *higher* once controlling for assignment relevant variables.

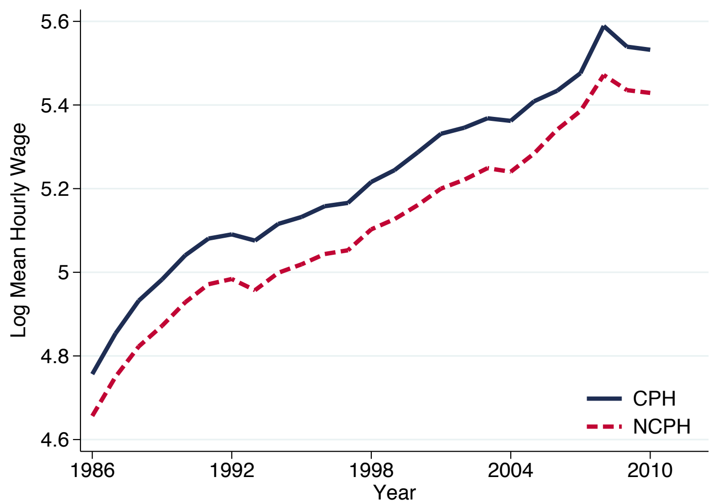
Together with the evidence in Table 2, this suggests that the treatment induced selection into the labour force for the less educated; they were less likely to work at all if assigned to a city, but those that did had greater success in getting and keeping a job. One hypothesis is that the city environment could be particularly hard on those without formal education, and those that do join the labour force are more able than those who do not, compared with those assigned to locations outside the city.

We do not confront this issue directly. Instead, for the remainder of the paper we focus on those with at least a high school degree, which is around 55% of the individuals in our baseline sample. This is not because employment and wage outcomes for the less skilled are unimportant or uninteresting. However, our purpose in this study is to leverage this particular historical setting for insight into the sources of the urban wage premium. Since we can rule out selection into work as a driver of the Dynamic Treatment Effect for those with at least a high school degree, this allows us to say more than we otherwise might. This is also not to say that other forms of selection *within* a city, conditional on working, are not a driver of the Dynamic Treatment Effect on wage and earnings outcomes, an issue we take up below.

3.2 Interpreting Treatment Effects

Here we discuss the interpretation of the findings in the previous section. First, we rule out that differential aggregate wage trends across the two regions are driving our results. If the overall level of wages in Copenhagen were growing faster than other regions, such a trend could show up in β_3 in Equation (1), as the arriving cohorts age. In Figure 3 we plot the average hourly wage earned by all working Danes, from 1986 to 2010, split by each worker's current location in that year. The Copenhagen premium is quite stable in the aggregate, averaging 10 log points between 1986 and 2010. Most importantly, we do not observe a systematic divergence between Copenhagen and Non-Copenhagen for Danes. We infer that Copenhagen is not on a steeper overall growth path than the rest of the country.[17]

Figure 3: Wage Growth Among All Danes



Note: Log mean hourly wage is the log of the mean of all hourly wage observations in that region and year in Denmark. Copenhagen and Non-Copenhagen correspond to our commuting zone definitions discussed above.

We now turn to the extent to which this Dynamic Treatment Effect can be interpreted as the statistical return to experience *actually earned in the city*. The ideal experiment to construct actual city wage-experience profiles would be to randomly allocate individuals across Copenhagen and Non-Copenhagen, and then not allow them to move for the remainder of their life. This would identify, at the level of population averages, both the level effect and the difference in slopes from placing the same populations in different environments, and keeping them there. While undoubtedly extreme, it is important to note that even this experiment would not identify the causal return to an additional year of city experience for a given individual; who gains experience within a city may depend on an

[17] Note that this is not inconsistent with differential returns to experience in Copenhagen; simple models of lifecycle earnings with overlapping cohorts can exhibit differential wage-experience slopes across locations, but a stable aggregate city premium over time as older cohorts are replaced.

unobserved type, as may other determinants of wages which correlate with experience. We return to this in Section 5. However, this experiment would be the ideal first step to determining the aggregate returns to experience, controlling for selection of individuals *across* cities.

Unfortunately, this is not the experiment we observe. Since we are only conditioning on the initial assigned location, using the Dynamic Treatment Effect as a proxy for the return to big city experience in an ideal experiment hinges on two crucial points.

First, under the assumption that the true wage-experience profile is steeper in the city, the extent to which individuals move at all will cause the dynamic treatment effect documented above to underestimate the true dynamic city wage premium. This is because those initially allocated to Non-Copenhagen, but who move to Copenhagen and benefit from the city's dynamic premium, will still be counted in the Non-Copenhagen stratification. Their higher wages drive a wedge between the pure dynamic treatment effect of the initial allocation relative to the true slope difference between locations, which is higher. Similarly, those initially allocated to Copenhagen who leave Copenhagen and earn lower wages with experience will still count in the Copenhagen stratification, pushing downwards the estimated dynamic treatment effect relative to the true wage-experience profile in Copenhagen.

To clarify this point, suppose that after de-meaning all variables, hourly wages of worker i at time t can be written as

$$w_{it} = \gamma^s E_{it}^s + \gamma^b E_{it}^b + \eta_{it}, \tag{2}$$

where E_{it}^s and E_{it}^b are years of experience in small (s) and big cities (b) respectively, and γ^s and γ^b are the causal returns to these experience types, with $\gamma^b > \gamma^s$ by assumption. Moreover, suppose that η_{it} is a structural error that captures other determinants of city-wages (e.g. good firm matches and occupation shifters), with $\mathbb{E}[E_{it}^s \eta_{it}]$ and $\mathbb{E}[E_{it}^b \eta_{it}]$ not necessarily zero. The ideal experiment does not hope to recover measures of γ^s, but instead estimates Equation (2) by OLS on two identical populations who cannot move. To see what this delivers, write $\boldsymbol{\gamma} = [\gamma^s \ \gamma^b]$ as

$$\boldsymbol{\gamma} = \mathbb{E}[E_{it}' E_{it}]^{-1} \left(\mathbb{E}[E_{it}' w_{it}] - \mathbb{E}[E_{it}' \eta_{it}] \right),$$

where $E_{it} = (E_{it}^b \ E_{it}^s)$. Under the conditions of the ideal experiment, with no movement at all between the two regions

$$\mathbb{E}[E_{it}' E_{it}]^{-1} = \begin{bmatrix} 1/(\sigma_E^s)^2 & 0 \\ 0 & 1/(\sigma_E^b)^2 \end{bmatrix},$$

since every individual in the sample will only have experience of one type. Then, for example,

$$\text{plim } \hat{\gamma}^b = \gamma^b + \frac{1}{(\sigma_E^b)^2} \mathbb{E}[E_{it}^b \eta_{it}] \equiv \beta^b.$$

Here β^b is the slope of the population wage-experience profile for individuals living in the big city, and staying there. It consists of both the true causal return to big city experience γ^b, and a term which reflects how this experience covaries with other determinants of wages there. We could estimate the two slope coefficients by two separate OLS regressions, one on the individuals assigned to Copenhagen and one on those assigned to Non-Copenhagen, given they do not move.

The movement of individuals after assignment (see Figure 2) brings us away from this ideal setting. Suppose hence we assign individuals to two groups: those initially allocated to big cities, group B, and those allocated elsewhere, group S, as we do in our treatment regressions. We then construct the variable $\tilde{E}_{it} \equiv E^b_{it} + E^s_{it}$ as total years of experience, irrespective of where these are accumulated, and regress w_{it} on $\tilde{E}_{it}$ for individuals in group B to calculate $\hat{\beta}^B$. This estimator obeys

$$\text{plim } \hat{\beta}^B = \frac{\mathbb{E}[\tilde{E}_{it} w_{it} \mid i \in B]}{\mathbb{E}[\tilde{E}^2_{it} \mid i \in B]} = \frac{\mathbb{E}[(E^b_{it})^2 | i \in B]}{\mathbb{E}[\tilde{E}^2_{it} \mid i \in B]} \frac{\mathbb{E}[E^b_{it} w_{it} \mid i \in B]}{\mathbb{E}[(E^b_{it})^2 | i \in B]} + \frac{\mathbb{E}[(E^s_{it})^2 \mid i \in B]}{\mathbb{E}[\tilde{E}^2_{it} \mid i \in B]} \frac{\mathbb{E}_{it}[E^s_{it} w_{it} \mid i \in B]}{\mathbb{E}[(E^s_{it})^2 \mid i \in B]}.$$

If movement is completely random between areas, then $\frac{\mathbb{E}[E^b_{it} w_{it} | i \in B]}{\mathbb{E}[E^{b2}_{it} | i \in B]}$ is the probability limit of the OLS coefficient on big city experience from conditioning on location for our group B, and including regressors for E^b_{it} and E^s_{it} separately. In this case the above expression simplifies to

$$\text{plim } \hat{\beta}^B = \frac{\mathbb{E}[(E^b_{it})^2 \mid i \in B]}{\mathbb{E}[\tilde{E}^2_{it} \mid i \in B]} \beta^b + \frac{\mathbb{E}[(E^s_{it})^2 \mid i \in B]}{\mathbb{E}[\tilde{E}^2_{it} \mid i \in B]} \beta^s < \beta^b.$$

So movement between the two regions will cause our coefficient in the Dynamic Treatment Effect to be a weighted average of the true wage-experience profiles for our groups, where these weights depend on the extent that the groups move between regions. As seen above in Figure 2, initial assignment is highly persistent. More than 80% of refugees initially assigned to Copenhagen who have been in Denmark for 20 years have never moved out of Copenhagen. Likewise, more than 75% of refugees initially assigned to Non-Copenhagen never move into Copenhagen. Nonetheless, movement away from initial allocation is not zero, and by itself this would mean that interpreting our results as the return to big city experience would understate the true profile.

The second important factor that complicates the interpretation of our Dynamic Treatment Effects as return to big city experience, is selection in *who* moves across locations after assignment. Such selection is likely a complex process, and we do not mean to develop a definitive taxonomy of selection across areas in this section. Instead, here we think about how to interpret the Dynamic Treatment Effect under a simple framework of selection on ability. A classic hypothesis entertained in the urban literature is that workers differ in their ability to learn and take advantage of big city experience, as in De La Roca and Puga (2017). To fix ideas, suppose that workers, indexed by i, differ only on a single dimension of unobserved ability, θ_i, and that the true wage function of worker i at time t is given by

$$w_{it} = \gamma^s E^s_{it} + (\gamma^b + \theta_i) E^b_{it} + \eta_{it} \tag{3}$$

where $E[\theta_i | i \in B] = E[\theta_i | i \in S] = 0$ under the conditions of the natural experiment. Thus workers with higher θ_i will see their wages rise faster with experience if they work in the big city.

Suppose further that no-one leaves Copenhagen, but that some high ability individuals immediately move to Copenhagen on assignment, and stay there.[18] Then we can write

$$\begin{aligned}
\text{plim } \hat{\beta}^B - \hat{\beta}^S &= \frac{\mathbb{E}[\tilde{E}_{it} w_{it} \mid i \in B]}{\mathbb{E}[\tilde{E}^2_{it} \mid i \in B]} - \frac{\mathbb{E}[\tilde{E}_{it} w_{it} \mid i \in S]}{\mathbb{E}[\tilde{E}^2_{it} \mid i \in S]} \\
&= \frac{\mathbb{E}[E^b_{it} w_{it} \mid i \in B]}{\mathbb{E}[(E^b_{it})^2 \mid i \in B]} - \frac{\mathbb{E}[\tilde{E}_{it} w_{it} \mid i \in S]}{\mathbb{E}[\tilde{E}^2_{it} \mid i \in S]} = \beta^b - \frac{\mathbb{E}[\tilde{E}_{it} w_{it} \mid i \in S]}{\mathbb{E}[\tilde{E}^2_{it} \mid i \in S]}
\end{aligned}$$

where the second equality follows from the fact that workers in the B group only have experience in the big city, and the third from the fact that assignment to the B group is random. Let $\varkappa_b$ denote the fraction of individuals in group S who move from s to b at assignment. Also denote by $i \to s$ agents who stay in s, while $i \to b$ denotes agents moving from s to b. Then, to illustrate the second term, use the law of iterated expectations to write

$$\begin{aligned}
\frac{\mathbb{E}[\tilde{E}_{it} w_{it} \mid i \in S]}{\mathbb{E}[\tilde{E}^2_{it} \mid i \in S]} &= \frac{\mathbb{E}[\tilde{E}_{it} w_{it} \mid i \in S, i \to b]}{\mathbb{E}[\tilde{E}^2_{it} \mid i \in S]} \varkappa_b + \frac{\mathbb{E}[\tilde{E}_{it} w_{it} \mid i \in S, i \to s]}{\mathbb{E}[\tilde{E}^2_{it} \mid i \in S]} (1 - \varkappa_b) \\
&= \frac{\mathbb{E}[(E^b_{it})^2 (\gamma^b + \theta_i) + E^b_{it} \eta_i \mid i \in S, i \to b]}{\mathbb{E}[\tilde{E}^2_{it} \mid i \in S]} \varkappa_b + \frac{\mathbb{E}[(E^s_{it})^2 \gamma^s + E^s_{it} \eta_i \mid i \in S, i \to s]}{\mathbb{E}[\tilde{E}^2_{it} \mid i \in S]} (1 - \varkappa_b)
\end{aligned}$$

where we used equation (3) to move from the first to the second line.

Lastly, assume that job-finding rates are identical for all workers (though the we still allow the distribution of the structural error to be correlated with experience type).[19] Under this assumption we

[18]This is of course a stylized depiction of what Figure 2 revealed to be occurring in the data in the years after assignment, but can be generalized to more complex movement dynamics

[19]Formally, assume that:

1. Job-finding rates are identical across areas, which imposes

$$\mathbb{E}[\tilde{E}^2_{it} \mid i \in S] = \mathbb{E}[(E^b_{it})^2 \mid i \in S, i \to b] = \mathbb{E}[(E^s_{it})^2 \mid i \in S, i \to s]$$

2. The distribution of matches, and hence the structural error, is independent of type, but may depend on experience in a place

$$\mathbb{E}[E^b_{it} \eta_i \mid i \in S, i \to b] = \mathbb{E}[E^b_{it} \eta_i]$$
$$\mathbb{E}[E^s_{it} \eta_i \mid i \in S, i \to s] = \mathbb{E}[E^s_{it} \eta_i]$$

These two assumptions are relaxed in the structural model in Section 5.

can show that the difference in estimated slopes reduces to

$$\text{plim } \hat{\beta}^B - \hat{\beta}^S = \beta^b - \left[\varkappa_b(\gamma^b + \mathbb{E}[E^b_{it}\eta_i]) + (1-\varkappa_b)(\gamma^s + \mathbb{E}[E^s_{it}\eta_i]) + \varkappa_b \mathbb{E}[\theta_i | i \in S, i \to b]\right] < \beta^b - \beta^s$$

It is clear from this equation that under this simple one-dimensional selection scenario, there are two reasons why the true differential return to experience is understated by our Dynamic Treatment Effect from Section 3.1.1. The first has already been discussed: independent from selection on unobservables, simple migratory movement from s to b (ie. $\varkappa_b \neq 1$), will cause us to understate the true gap in wage-experience profiles. The second reason is that if $\mathbb{E}[\theta_i | i \in S, i \to b] > 0$, such that high ability individuals in group S move into the big city and see their wages grow faster, their steeper wage profiles will be counted in group S, further raising $\hat{\beta}^S$ and shrinking the gap in slopes between our two groups.

Without doubt, there are other possible specifications one could use to think about the interpretation of the Dynamic Treatment Effect, including sorting on preferences for amenities or comparative advantage, and multi-dimensional types (as in Lindenlaub (2017)). However, this simple story of selection across areas on ability is appealing for two reasons. First, it accords with our evidence on observables in Figure 2. There we find very limited movement out of Copenhagen, and further the movement that does occur is not differentiated by skill. However, we find more movement to Copenhagen, and additionally we find that those with more education are more likely to move.

Second, under these assumptions, we can construct an informative upper bound on the true return to experience differential. Namely, we can reestimate our baseline equation only on individuals who *never move* across the two regions. This serves two purposes. First, since everyone who never moves earns all their experience in one location, this measure does not suffer from the first source of downward bias we identified above. Second, we remove on average high-ability individuals whose wages grow faster with experience after moving into Copenhagen. This leaves on average lower ability individuals in our Non-Copenhagen stratification than compared to the randomly selected sample. Jointly, this has the effect of biasing upwards our proxy of the true wage-experience differential slope.

We report our results in Table 9 in Appendix A.6, under the same specifications as our baseline. We see that across these specifications, the reported coefficient for β_3 is uniformly higher by around 0.4%. This suggests that the true difference in wage-experience profile slopes lies in the range of $0.80\% - 1.2\%$, which is a relatively tight bound, suggesting that our Dynamic Treatment Effect is a close proxy for the return to big city experience under the ideal experiment given the assumptions outlined above.

4 Observable Mechanisms of the Dynamic Treatment Effect

The analysis in the previous Section examined the extent to which we could interpret our Dynamic Treatment Effect as the differential return to experience actually earned in a big city. The objective of this Section is to understand the driving forces underpinning the effect itself. We employ the richness of the Danish register data to consider four hypotheses on observable variables that could be contributing to the measured Dynamic Treatment Effect from above. These are:

1. *More valuable experience at high-wage firms.* As is commonly found for big cities in the urban literature (see Combes et al. (2012)), Copenhagen is home to both relatively more high-productivity firms, and more high-wage paying firms. If experience at these firms is more valuable than experience at other firms, and refugees sort into these firms, then part of the Dynamic Treatment Effect will be explained by this compositional effect.

2. *Distinctive occupation distributions.* Even at a high level of aggregation, the distribution of workers across occupations is not uniform across space (see Appendix A.7). Relative to the rest of the country, natives in Copenhagen are more likely to work in more skill intensive service work, and less likely to work in manual labor jobs and agriculture. If the availability of different types of work differs across our two regions, and consequently refugees sort into different occupations as they gain experience, then this will also contribute to the β_3 in Equation (1).

3. *Differential take-up of education.* It is well documented that the skill premium is higher in cities. It could be that on assignment to Copenhagen, it is relatively more attractive to invest in further education and training for arriving refugees, and so they increase their years of education more than those assigned to Non-Copenhagen.

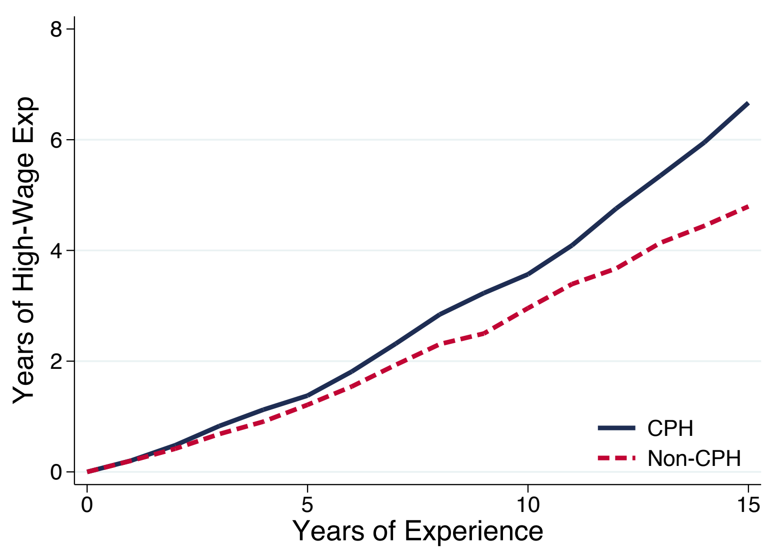
4. *Effects of ethnic enclaves.* Several papers (see Edin et al. (2003)) have examined the effect of being located in an ethnic enclave on employment and wage outcomes, and find generally positive effects, especially for the low-skilled. As Damm (2014) notes, before the policy was introduced, immigrants and refugees clustered in Copenhagen and the other second-tier cities. We could be picking up the dynamic wage effect of colocating with other immigrants of one's nationality, which gives access to informal networks and employment opportunities.

In addition to these observable channels, it is worth noting that dynamic selection on unobservables *within* areas could be driving part of our result. This is conceptually distinct from the selection across areas that complicates our ability to interpret the Dynamic Treatment Effect as the return to big city experience. First, there could be selection into who obtains experience at all. If higher ability individuals are more likely to get jobs in Copenhagen than outside, then this could contribute to the pattern we see. Second, selection on unobservables could interact with the second and third hypotheses above, where higher skilled individuals differentially sort into high-wage firms and high-wage occupations over time. Consideration of this issue requires careful specification of an appropriate structural model, which we save for the last section of the paper. In the current Section however, we start by asking whether, after controlling for observable channels 1-4 listed above, there is any differential return to experience for those allocated to Copenhagen.

We find no support for the third and fourth hypotheses. As seen in Figure 12 in Appendix A.8, and discussed in more detail there, differences in take-up of further education after arrival in Denmark of all different types are non-existent across our two areas. Also in Appendix A.9, we include in our baseline specification controls for the ethnic composition of the municipality of assignment for each refugee, and in particular the stock of co-nationals in the year of assignment.[20] Doing so leaves our baseline results for the Dynamic Treatment effect virtually unchanged. We do see refugees assigned to ethnic enclaves receiving lower wages and earnings on average, but this does not appear to influence the Dynamic Treatment Effect of being assigned to the big city.

The first and second hypothesis speak to a view of the productive advantages of cities that arises from the fact that cities host productive capital and firms, and utilize these through specialized occupations. People who live in big cities do different activities than those outside, and work with fundamentally different production technologies. As they gain experience, this could affect their marginal product by both directly affecting their accumulation of human capital, and indirectly by allowing them to gradually sort into high-productivity firms and capital.

Figure 4: High-Wage Experience



Note: For all observations with a given amount of total experience, we compute the mean number of years of $HighExp_{it}$ and plot against total experience.

[20]It is worth noting that the Copenhagen commuting zone we constructed in Section 2.3 consists of about thirty municipalities, so that there is varitation in these stocks across the municipalities of assignment within the Copenhagen commuting zone.

Jointly, the first and second hypothesis explain much of the Dynamic Treatment Effect. To examine the first, we proceed at the establishment level. We divide all establishments into deciles each year, based on the average wage paid at the establishment. The average wage is computed using all Danish hourly-wage observations at the person-year level (not just refugees), subject to the sample selection procedure outlined in Section 2.2. We then classify an establishment as high-wage if it is in the top three deciles. High-wage establishments tend to be large and disproportionately present in Copenhagen.

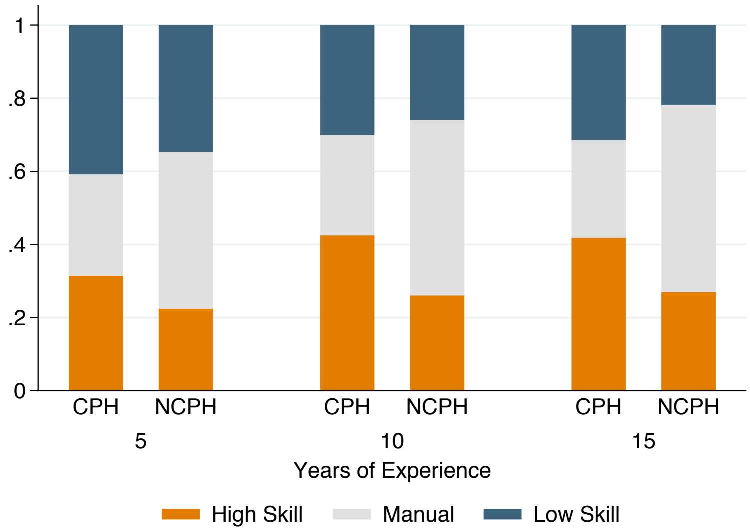
For each refugee, we then divide experience into $HighExp_{it}$ and $OtherExp_{it}$, where these increment by one when a refugee works at a given establishment in the classification in a year. In Figure 4, we plot the average number of years of $HighExp_{it}$ for all refugees with a given number of years of total experience, by initial area of assignment for our preferred sample. We see that those assigned to Copenhagen gradually accumulate more years of $HighExp_{it}$ over time, and 15 years into their work experience have on average 2 more years of $HighExp_{it}$ relative to those not assigned to Copenhagen.

Before examining how this impacts our Dynamic Treatment Effect, we describe how we address the second hypothesis. The datasets we employ contain a detailed description of occupations in Denmark, which categorizes occupations at a four-digit level. Here we work with only the broadest one-digit grouping, which contains nine different categorizations. A list of these can be found in Appendix A.7. To show the differences across areas, we further assign these codes to just three groups: low-skill service work, manual work and high-skill service work. These occupational groups are ordered in terms of the average wage of all Danes working in them, with low-skill workers earning the least and high-skill workers the most. In Figure 5, we show the occupational distribution across areas for refugees with 5, 10 and 15 years of experience.

We see that in both areas, many refugees initially start their careers doing low-skill service work (for example, cleaning, sales and clerical support). However, as they gain experience, refugees placed outside of Copenhagen transition into predominantly manual work. We also see those refugees placed in Copenhagen move more into high-skilled service work, with these representing about 40% of those with 15 years of experience.

We now formally model the contribution of these two channels to explaining the Dynamic Treatment Effect. We reestimate Equation (1) with separate controls for $HighExp_{it}$ and $OtherExp_{it}$, but keep the interaction term on only total experience. Effectively, here we constrain the Dynamic Treatment Effect to only operate on total experience. This allows us to examine how much the *composition* of experience is driving the results above. We also include a fixed effect for whether the firm a refugee is working at is a high-wage firm. This specification is estimated in Column (2) of Table 3 (Column (1) merely repeats Column (2) of Table 1 for ease of comparison), and reduces our baseline coefficient to 0.566%, suggesting that those assigned to Copenhagen move to higher paying establishments increasingly over time. As is intuitive, experience at high-wage establishments is worth more in a

Figure 5: Occupations by Initial Assignment and Years in Denmark



Note: The above shows the distribution of refugees , across occupation groups for those with 5/10/15 years of experience. Sample is restricted to those with at least 12 years of education at arrival.

statistical sense than experience at other establishments.

Table 3: Wages and Experience- Mechanisms

	(1) $logwage_{it}$	(2) $logwage_{it}$	(3) $logwage_{it}$	(4) $logwage_{it}$
Exp_{it}	0.0250***			
	(0.00159)			
$HighExp_{it}$		0.0259***	0.0217***	0.0225***
		(0.00121)	(0.00124)	(0.00138)
$OtherExp_{it}$		0.0206***	0.0186***	0.0157***
		(0.00203)	(0.00140)	(0.00126)
$InitCph_i$	0.00858	-0.00241	0.00597	0.00580
	(0.00883)	(0.00676)	(0.00538)	(0.00533)
$InitCph_i \times Exp_{it}$	0.00736***	0.00566***	0.00354**	0.00278*
	(0.00163)	(0.00148)	(0.00115)	(0.00101)
No. Kids at Arr.	0.0143	0.0184*	0.0108	0.00922
	(0.0117)	(0.00803)	(0.0100)	(0.00788)
Married at Arr.	0.0210	0.0177	0.0171	0.0196
	(0.0163)	(0.0134)	(0.0102)	(0.0108)
Age at Arr.	-0.00228*	-0.00178*	-0.00135*	-0.00170**
	(0.000841)	(0.000715)	(0.000626)	(0.000503)
Observations	57994	57994	48183	44135
R^2	0.062	0.137	0.188	0.224
Sample	Educ≥12	Educ≥12	Educ≥12	Educ≥12
High-Wage Firm FE	No	Yes	Yes	Yes
Occupation FE	No	No	Yes	Yes
Industry FE	No	No	No	Yes

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Specification given in Equation (1), using fixed effects for whether an establishment is classified as "high-wage" or not, separate controls for $HighExp_{it}$ and $OtherExp_{it}$, and occupational fixed effects. Robust standard errors clustered at the level of initial commuting zone. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

To control for differential occupational sorting, we include occupational fixed effects at the 1-digit level in Column (3). This further decreases the coefficient on the Dynamic Treatment Effect to 0.354%. This is line with the descriptive patterns above: refugees assigned to Copenhagen have a higher chance of working in high-skilled jobs, and this advantage widens over time. Lastly, we include fixed effects for 1-digit industries, a list of which can be found in the Online Appendix. This further decreases the coefficient to 0.278%. The qualitative patterns here change little with order of inclusion of these three specifications. We conclude that jointly, these observables account for a significant fraction of the Dynamic Treatment Effect.

It is worth emphasizing that in our specifications, occupations are hierarchical, and ordered by average wage. Including controls for sorting into these groups suggests a different effect to that previously explored in the urban literature. For example, in Papageorgiou (2017), occupations are horizontally differentiated, and workers have different comparative advantages across occupations. The advantage of cities is that they allow a worker to try out more occupations, and learn what she is good at; occupations themselves have no fundamental differences. For our study, the advantage of a big city is that the *quality* of work available is different. Firms in Copenhagen offer more opportunities for work in high-skill service jobs, which pay more, *and* less in manual work. Workers move up this hierarchy over time, and on average end up in higher paying occupations in the city.

5 Quantifying Sorting Within Cities

We can explain a significant fraction of the dynamic urban wage premium for our population through controlling for observables indexed to firm and occupation type. In this section, we go further, and quantify the contribution from sorting across unobserved types *within* cities. This requires the specification of an appropriate structural model of spatial earnings dynamics. In line with the spirit of Section 3.2, we consider only one-dimensional heterogeneity, where individuals differ by an unobserved productivity within a job. Ex-ante, there are at least three reasons why sorting within cities in such a setting may be important.

First, who gets experience at all may differ fundamentally within a city and without, even with identical populations in both locations. Specifically, it may be that employment rates for high ability individuals are higher in the city. This would introduce a correlation between observed experience and the likelihood that an individual is a high ability individual, pushing up the estimated returns to experience. Given our evidence in Section 3.1.2 and Appendix A.2 that overall employment rates are not affected by assignment to the city for this population, it would also have to be the case that low ability individuals had correspondingly lower employment rates inside the city, leaving the overall employment rate unchanged.

Second, there could be a correlation between unobserved type and the amount of experience accumulated specifically at a high-wage firm. If, for example, high-ability individuals are more likely to work at high-wage firms, and there are more high-wage firms in the city, we would misattribute some of the observed dynamic city premium to a firm component.

Lastly, there could be complementarities between worker type and firm type, boosting the wage for high ability individuals when they are working at a high-wage firm. If there is differential sorting into such firms over time within the city, this could appear to raise the premium to experience earned in the big city.

In other words, we aim to distinguish mechanisms of sorting by ability and composition from mechanisms of general human capital accumulation, which may occur faster in a city. To this end, we construct and estimate a structural model of location choice and earnings dynamics that explicitly accounts for unobserved heterogeneity of workers. We estimate the model on our sample of refugees using a maximum likelihood strategy that exploits the quasi-random assignment of individuals across space.

5.1 A Model of Spatial Earnings Dynamics

Our model builds on the framework of Baum-Snow and Pavan (2011). We allow individuals to be of one of two latent ability types unobserved by the econometrician. The spatial dispersal policy is modeled as initial assignment to an area, and the randomness of the assignment guarantees the orthogonality of initial assignment and latent ability.

In the model workers can work either in Copenhagen ($j = CPH$) or outside Copenhagen ($j = NCPH$) and be of either of low or high ability, indexed by $h = \{h_L, h_H\}$. Additionally, a worker can be employed at a firm of either one of two types, low or high productivity, indexed by $f = \{f_L, f_H\}$.

In the model workers work for T years before retiring. After retirement individuals live for an additional $T/2$ years at an income equal to their last annual wage before dying for sure. Throughout their working life workers are either employed (E) or unemployed (UE) in either of the two locations. We denote by x^i_f the years of experience of worker i at firm type f and by $\mathbf{x}^i$ the vector $\{x^i_{f_L}, x^i_{f_H}\}$.

Lastly, we abstract from modeling occupation choice for tractability. While occupations are an important driver of the observed premium, here we focus on how how sorting on unobserved ability contributes to only the observed firm channel in order to obtain sharp results.

5.1.1 Wage Function

The wage of a worker in location j at time t, if employed, is a function of four state variables $(h^i, \mathbf{x}^i, f^i)$.[21] For notational simplicity we suppress the worker index i on these state variables for the remainder of the model section. The log wage process of a worker, conditional on receiving an offer in location j and being employed by a firm of type f, is given by

$$\ln w(h, \mathbf{x}, f) = \bar{\varpi} + \theta_h + \psi_f + \alpha_{h_H, f_H} + \sum_p \beta_1^{h, f_p} x_{f_p} + \beta_2 \left(\sum_p x_{f_p} \right)^2 + u.$$

Here, θ_h denotes the component of the wage that depends on (invariant) latent type, and ψ_f that component which depends on the productivity of the firm. In the empirical implementation we set $\theta_{h_L} = \psi_{f_L} = 0$. Consequently, $\bar{\varpi}$ is the mean wage offer of low latent ability workers from low-productivity establishments controlling for experience. α_{h_H, f_H} denotes a complementarity in the

[21] The formulation and wage function we employ can be rationalised as the outcome of a general equilibrium islands model where refugees are a small fraction of the total population. The formal description of this model is available on request from the authors.

offered wage for high-type workers working at high-type firms. $\beta_1^{h,f}$ denotes the return to low- and high-type experience which we allow to differ across latent ability. β_2 denotes a common return to total experience squared. We also assume there is a shock to match quality u that shifts each period's wage. We assume that u is drawn i.i.d in every period from a type-I extreme value distribution with mean zero and a variance denoted by σ_u.

5.1.2 Labor Market and Location Transitions

For simplicity, we assume that workers in location j only receive job offers from firms in the same location. The arrival rate of job offers from unemployment for any firm type is denoted by $\underline{\lambda}_j^h$, which is indexed by latent ability type and location j. Further let $\pi_j^{h,f}$ denote the fraction of job offers from type f firms, conditional on receiving a job offer from any firm, for individuals of ability h in location j.

Workers run the risk of being exogenously separated from their job with probability δ_j^h, which we allow to differ across locations and latent ability. Further, workers may choose to quit a job, either to move to another location or into voluntary unemployment. They receive job offers while employed with probability $\bar{\lambda}_j^h$ .Workers additionally face the risk of being hit by a reallocation shock at a rate μ_j^h. In this case, the worker is separated from his job and receives a new job offer which is of firm type f with probability $\pi_j^{h,f}$, but does not have to pass through unemployment.[22]

Agents receive preference shocks for each location j in every period denoted by η_j. We assume the shock to be distributed according to an extreme value Type-1 distribution with mean zero and variance σ_η. If an agent decides to move he loses his current job and needs to look for a new job in the destination location. Moving between locations incurs at a utility cost of moving denoted by τ.[23]

5.1.3 Value Functions

We denote the expected value of being employed in location j by $\bar{V}_t^E(j, h, \mathbf{x}, f)$, and of being unemployed by $\bar{V}_t^{UE}(j, h, \mathbf{x})$.[24] Here t indexes the years of labor market participation, employed or unemployed, by a given individual since arrival in Denmark.

All workers receive amenity services from location j which we denote by a_j, which additionally captures differences in the cost of living. We denote unemployment benefits by b.

[22] This shock is necessary to match transition probabilities across firm types. In the data a worker may move from a high-wage firm to a low-wage firm, which is hard to rationalise without forced mobility.

[23] Given that we interpret firms as establishments, the same firm can never be located in two regions and hence moving necessitates a "firm" change. It is conceivable that workers search for job in other other regions and move there upon finding one. However, this is empirically indistinguishable give our data from workers moving to another region and finding a job there immediately upon arrival.

[24] Unemployed workers do not have a current firm type f, which implies a reduction in the state space for unemployed relative to employed workers.

The value of being employed and unemployed conditional on realized location and match specific shocks can then be written as follows:

$$\bar{V}_t^E(j,h,\mathbf{x},f \mid u,\eta_j,\eta_{j'}) = a_j + \ln w(\cdot) + \rho \max_{j,j'}\{\bar{U}_{t+1}^E(j,h,\mathbf{x'},f)+\eta_j, \bar{U}_{t+1}^{UE}(j',h,\mathbf{x'}) - \tau_j + \eta_{j'}\},$$
$$\bar{V}_t^{UE}(j,h,\mathbf{x} \mid \eta_j,\eta_{j'}) = a_j + \ln b + \rho \max_{j,j'}\{\bar{U}_{t+1}^{UE}(j,h,\mathbf{x})+\eta_j, \bar{U}_{t+1}^{UE}(j',h,\mathbf{x}) - \tau_j + \eta_{j'}\}.$$

Here $\bar{U}_t^E$ and $\bar{U}_t^{UE}$ denote the continuation values of being in location j at the beginning of next period, net of idiosyncratic location preferences and conditional on current period labor market status. $\mathbf{x}'$ denotes the updated vector of firm type specific experience of the worker next period. Depending on the current firm type f, $\mathbf{x}'$ will increment one of the components of $\mathbf{x}$ by one. Lastly, ρ is a yearly discount rate. Using the properties of the Gumbel distributed location shock we obtain the following expression for the expected value of being employed conditional on the match specific shock:

$$\begin{aligned}\bar{V}_t^E(j,h,\mathbf{x},f \mid u) &\equiv \mathbb{E}_{\eta_{j'}} V_t^E(j,h,\mathbf{x},f)\\ &= a_j + \ln w(\cdot) + \frac{\rho}{\sigma_\eta}\log\left[\exp\left(\sigma_\eta \bar{U}_{t+1}^E(j,h,\mathbf{x'},f)\right) + \exp\left(\sigma_\eta \bar{U}_{t+1}^{UE}(j',h,\mathbf{x'}) - \sigma_\eta\tau_j\right)\right]\end{aligned}$$

Further, the value of being unemployed is given by

$$\begin{aligned}\bar{V}_t^{UE}(j,h,\mathbf{x}) \equiv \mathbb{E}_{\eta_{j'}} V_t^{UE}(j,h,\mathbf{x}) &= a_j + \ln b\\ &+ \frac{\rho}{\sigma_\eta}\log\left[\exp\left(\sigma_\eta \bar{U}_{t+1}^{UE}(j,h,\mathbf{x})\right) + \exp\left(\sigma_\eta \bar{U}_{t+1}^{UE}(j',h,\mathbf{x}) - \sigma_\eta\tau_j\right)\right]\end{aligned}$$

The expressions for $\bar{U}_t^E(j,h,\mathbf{x},f)$ and $\bar{U}_t^{UE}(j,h,\mathbf{x})$ can be written as follows

$$\begin{aligned}\bar{U}_t^E(j,h,\mathbf{x},f) = {}& \delta_j^h \bar{V}_t^{UE}(j,h,\mathbf{x})\\ &+ (1-\delta_j^h)\Big[\mu_j^h\{\pi_j^{h,f}\mathbb{E}_u \max\{\bar{V}_t^E(j,h,\mathbf{x},f \mid u), \bar{V}_t^{UE}(j,h,\mathbf{x})\}\\ &+ \pi_j^{h,f'}\mathbb{E}_u \max\{\bar{V}_t^E(j,h,\mathbf{x},f' \mid u), \bar{V}_t^{UE}(j,h,\mathbf{x})\}\}\\ &+ (1-\mu_j^h)\big[(1-\bar{\lambda}_j^h\pi_j^{h,f'})\mathbb{E}_u \max\{\bar{V}_t^E(j,h,\mathbf{x},f \mid u), \bar{V}_t^{UE}(j,h,\mathbf{x})\}\\ &+ \bar{\lambda}_j^h\pi_j^{h,f'}\mathbb{E}_{u,u'} \max\{\bar{V}_t^E(j,h,\mathbf{x},f \mid u), \bar{V}_t^{UE}(j,h,\mathbf{x}), \bar{V}_t^E(j,h,\mathbf{x},f' \mid u')\}\big]\Big],\end{aligned} \tag{4}$$

and

$$\bar{U}_t^{UE}(j,h,\mathbf{x}) = (1-\underline{\lambda}_j^h)\bar{V}_t^{UE}(j,h,\mathbf{x}) + \underline{\lambda}_j^h \sum_p \pi_j^{h,f_p}\mathbb{E}_u \max\{\bar{V}_t^E(j,h,\mathbf{x},p \mid u), \bar{V}_t^{UE}(j,h,\mathbf{x})\}. \tag{5}$$

In the expression for $\bar{U}_t^E$, f' denotes the opposite firm type of f, i.e. if f' is low than f denotes high and vice versa. These may seem to be daunting expressions, but each term is intuitive. In $\bar{U}_t^E$, the first term represents the value of being exogenously separated and going to unemployment. The second and third lines capture the value of receiving a reallocation show, while the fourth and fifth the value of getting shock or offers, and getting an offer for a new job, respectively. The distributional assumption on the random component of wages allows a closed form solution for the maximized values in (4) and (5).

5.1.4 Initial Transition into Work

We model separately the process for the initial transition into the labor force. Given the information in Section A.2, it takes some time for refugees to integrate into Danish society to the point of joining the labor force. In particular, the fraction of refugees coded as "Not in the Labour Force" (NILF) starts out very large, as they undertake Danish language classes and other integration programs, and receive social assistance. If we were to try to fit this simply using the job-finding rates in λ_j^h, we would not be able to match this relatively slow transition to looking for work.

Since we do not possess information on the activities of refugees before they join the labour force, we need to take a stand on this transition. We model refugees as starting in an initial state (NILF), after which they receive random draws to escape this state. Once they leave this state, they never return to it. This is consistent with what we find in the data; the chance that a refugee returns to the NILF state at any point in their life before retiring, once leaving it, is less than 1%.

In particular, a refugee in NILF receives job offers randomly at rate ϱ_j^h, where we allow this rate to depend on both unobserved type and location. Conditional on an offer, the chance this offer comes from a high-wage firm is the same as for all other refugees as in $\pi_j^{h,f}$. Lastly, we do not allow a refugee to move while they are in the initial state. As such, the value from being in the NILF state is given by

$$\begin{aligned} W_t = \varrho_j^h \Big[& \pi_j^{h,f_L} \mathbb{E}_u \max\{\bar{V}_t^E(j,h,0,f_L \mid u), a_j + \ln b + \rho W_{t+1}\} \\ & + \pi_j^{h,f_H} \mathbb{E}_u \max\{\bar{V}_t^E(j,h,0,f_H \mid u), a_j + \ln b + \rho W_{t+1}\} \Big] + (1-\varrho_j^h) \left[a_j + \ln b + \rho W_{t+1}\right] \end{aligned}$$

This concludes the description of the model. We now discuss how we build the likelihood function for having observed a panel of labor market histories given the structure and parameters of the model.

5.2 Estimation

The data input in our estimation is the entire longitudinal panel of workers' wages, firm types, and locations throughout 1986-2010.[25] This allows us to construct firm type specific experience counts for each individual in our data at each point in their labor market careers.

Latent ability, h, is the only unobserved state. Following Baum-Snow and Pavan (2011) we use the structure of the model to construct the probability of the observed life-path of every agent in the data conditional on its parameters and the agent's latent type. We denote such a life-path of length T by $\mathbf{Y}^i_T$. In particular, define Y^i_t to be the vector of labour market outcomes at time t which consists of a wage, if observed, the location of the worker and the type of labor market transition that the worker has experienced since the previous period so that $\mathbf{Y}^i_t = \{Y^i_1, \ldots, Y^i_t\}$ as the vector of all labor market observations in an individual's job history up to and including period t. Denote the model implied likelihood of observing a given path $\mathbf{Y}^i_T$ conditional on latent ability type and the entire set of parameters of the model, summarized in the vector θ, by $P(\mathbf{Y}^i_T \mid h; \theta)$.

We can then construct an individual worker's contribution to the overall likelihood function by weighting the contribution from the latent ability conditional likelihood as

$$L(\theta) = \chi_L P(\mathbf{Y}^i_T \mid h_L; \theta) + \chi_H P(\mathbf{Y}^i_T \mid h_H; \theta).$$

Here $\chi_L = 1 - \chi_H$ denotes the fraction of low latent ability workers in our refugee sample. Note that χ_L is a parameter to be estimated. Importantly, this is where the natural experiment interacts with the estimation. The randomness of the initial assignment implies that χ_L does not differ among refugees assigned to Copenhagen and Non-Copenhagen.

We can then decompose $P(\mathbf{Y}^i_T \mid h; \theta)$ as follows

$$P(\mathbf{Y}^i_T \mid h; \theta) = P(Y^i_1 \mid h; \theta) \prod_{t=2}^{T} P(Y^i_t \mid Y^i_{t-1}, h; \theta).$$

$P(Y^i_t \mid Y^i_{t-1}, h; \theta)$ is given by the model implied probability of observing an individual making a given location- and firm transition between $t-1$ and t as well as the likelihood of the observed wage at time t. All individuals are unemployed at the beginning of the first period, thus the probability of observing the first observation, $P(Y^i_1 \mid h; \theta)$, is given by the probability of making a transition from unemployment to the observed labor market state at the end of period one.

Since types are initially randomly allocated we can construct the log likelihood function simply by

[25] We define firm types as in Section 4: we use the full sample of all Danes for any given year to calculate average wage paid at each establishment. Then we split the population of establishments into those whose average wage falls into the top three deciles of the national establishment distribution and the rest. These establishments in the top three deciles in each year are what we refer to as high-productivity establishments.

summing $L(\theta)$ across all workers i in our sample

$$\mathcal{L}(\theta) \equiv \sum_i \log\left[\chi_L P(\mathbf{Y}_T^i \mid h_L;\theta) + (1-\chi_L)P(\mathbf{Y}_T^i \mid h_H;\theta)\right]. \tag{6}$$

A slight caveat is that we do not observe all workers for the same length of time T. Workers that arrive in later years are not observable for an entire T years. Our maximum likelihood strategy accommodates this conveniently: we simply record the contribution to the likelihood of all workers in all the period we observe them for and hence maximize the sum of the observed data. The Online Appendix contains a detailed derivation of the likelihood function and its components, as well as details on the mechanics of the estimation[26].

5.3 Estimation Results and Model Fit

Tables 14 and 15 in Appendix A.11 shows the parameter estimates of the full estimation. Several points are worth discussing. First, the model estimates a fairly significant separation of high and low types, with the high-type fixed effect for wages coming out at 27 log points. This is emphasized in Figure 14 in Appendix A.12, which shows simulated wage densities for both types vs. the actual data. Given the proportion of high-types is estimated at 48%, they account for almost all the mass in the upper quantiles of the wage distribution in the observed data. We also estimate somewhat lower returns to experience for these high types, suggesting that in the data these types see their wages start out higher, but grow slower.

The job-ladder parameters largely show patterns that accord with intuition. First, as with our results in Section 4, the probability of an offer from a high-wage firm is substantially larger in Copenhagen, with the difference being even more pronounced for a high-ability worker, suggesting that the city is operating to boost high-type workers up the job ladder. Second, job finding rates from unemployment are somewhat lower in the city, particularly for low ability workers. Lastly, the reallocation shock appears to occur more frequently outside of Copenhagen.

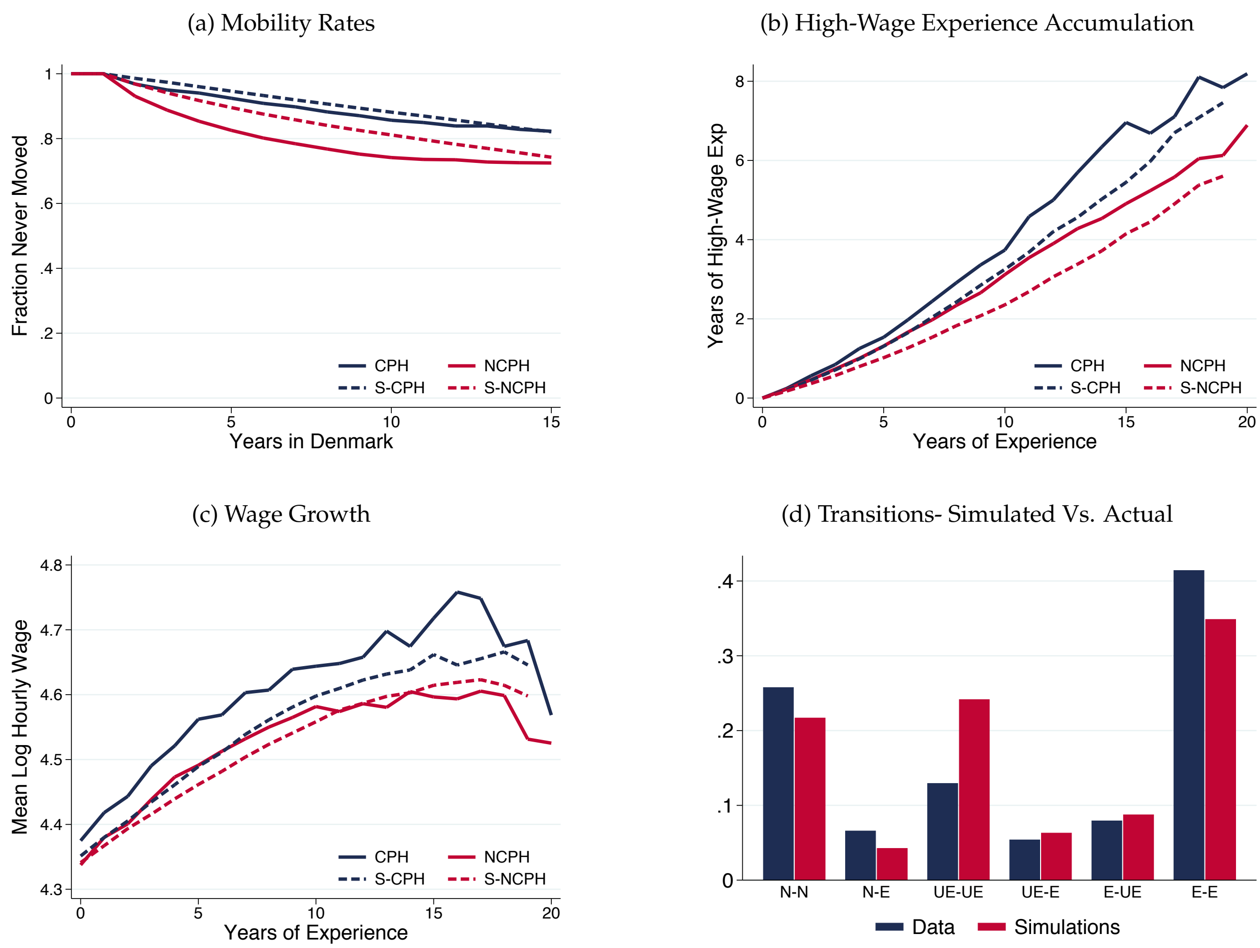
The model does a reasonable job of fitting the important patterns of the data. In Figure 6 we simulate 10^5 agents using the estimated parameters. Figure 6(a) and 6(b) show lifetime moving and high-wage experience accumulation profiles vs. the actual data, which are largely in line.

Figure 6(c) shows the dynamic wage-experience profiles by initial allocation in the simulated and actual data. The broad pattern is correct, though the size of the dynamic premium is less in our simulated model than in the actual data. The reason is that we have not modeled occupational choice in the structural section, and as found in Section 4 this explains a significant fraction of the Dynamic Treatment Effect.

[26]The Online Appendix is available upon request from the authors.

Figure 6(d) shows the state transitions in both simulated and actual data (N standing for the NILF state developed in Section 5.1.4, E for employed and U for Unemployed). These are fitted quite well, with the exception that the model somewhat misses the chance of getting out of unemployment, suggesting our job finding rates are too low.

Figure 6: Model Fit



Note: The model is simulated for 10^5 agents, and simulated data is compared to actual data.

5.4 Structural Decomposition of the Dynamic Wage Premium

We now come to the central result of this last Section. Here we quantify the contribution of sorting on unobserved ability in generating the Dynamic Treatment Effect through the lens of the structural model of earnings dynamics.

To do so, we consider three different restrictions of the full estimated model. These are:

1. *Mobility*. We shut down job-ladder differences across type, to study the contribution of the fact that the city has differential employment effects on high and low ability workers compared

with the outside. To do so, we force job finding rates λ_j^h, job destruction rates δ_j^h, high-wage sorting fractions π_j^h and reallocation shocks μ_j^h to be the same across worker types for a given j, by taking the mean of these parameters (weighted by χ_L).

2. *Returns to Experience*. Here we force the returns to experience for different firms to be the same across worker types, again by taking a simple average. These returns could interact with the mobility parameters above, since we find in Table A.11 that the dynamic returns to working at a high-wage firm are higher than at a low-wage firm, and there are more of these in the city.

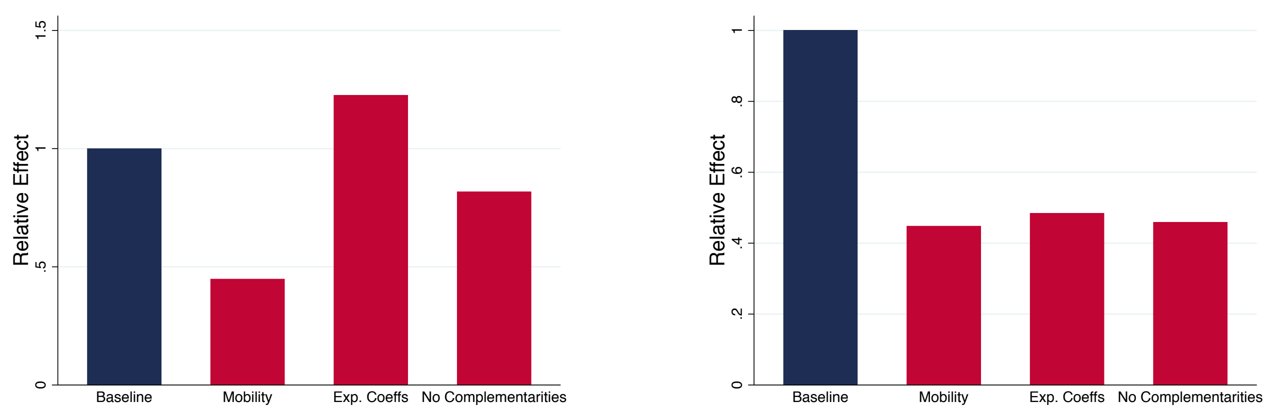
3. *Complementarities*. Lastly, we force the estimated complementarity between high-types and high-wage firms, α_{h_H,f_H}, to be zero.

We then simulate data for 10^5 agents and estimate the baseline regression in (1), conditioning only initial location. We report the Dynamic Treatment Effect of assignment to Copenhagen relative to the Dynamic Treatment Effect in the full structural model in Figure 7. In the left panel we do each of these decompositions in the order above, one at a time, and in the right panel we do them sequentially in the order above.

Figure 7: Dynamic Premium Decomposition

(a) Single Decompositions

(b) Sequential Decomposition



Note: The model is restricted to actual data simulated for 10^5 agents, and simulated data is compared to actual data.

We find shutting down mobility differences across types to be the most important channel, explaining around 54% of the baseline treatment effect in the full model. Once we do that, in Panel (b), shutting down the other two channels does almost nothing further. This is intuitive, since these two channels are dependent on the differential job ladder structure to operate. Differential returns to experience across types and complementarities will only show up as part of the premium if the city sorts high and low ability workers into different firms than would occur outside.

For our population, the city effect is operating not only by allowing all workers to gain more experience at high-wage firms than they would outside, but in particular is allowing high-ability workers to sort into good firms, and stay there once they have arrived. In our reduced form results in Section (3), this suggests that at least part of the explanatory power of conditioning on firm type is coming from the type of the workers. In our setting, in the big city it is more likely that a worker observed at a high-wage firm is themselves of high ability. Since these sorting patterns happen dynamically, this shows up as a differential return to experience between Copenhagen and Non-Copenhagen.

6 Conclusion

We find that exogenous assignment of refugees to a big city had a significant impact on a refugees's lifetime wage path. Initially, those assigned to Copenhagen and elsewhere earn similar hourly wages, but as refugees assigned to the city gain experience a gap opens up. This gap rises to a 16% premium after 20 years of experience. We find that controlling for the kind of establishment that composes an individual's experience, as well as occupational sorting across time, explains much of this gap. A structural analysis of selection and sorting into different establishments within the big city suggests that worker type matters for these patterns, and that the effect of the city for this population is at least partially explained by the higher prevalence of high-ability workers in high-paying firms.

References

ALTONJI, J. G. AND D. CARD (1991): "The effects of immigration on the labor market outcomes of less-skilled natives," in *Immigration, trade, and the labor market*, University of Chicago Press, 201–234.

ASLUND, O., J. OSTH, AND Y. ZENOU (2009): "How Important is Access to Jobs? Old Question-Improved Answer," *Journal of Economic Geography*, 10, 389–422.

ASLUND, O. AND D.-O. ROOTH (2007): "Do When and Where Matter? Initial Labour Market Conditions and Immigrant Earnings," *The Economic Journal*, 117, 422–448.

BAUM-SNOW, N. AND R. PAVAN (2011): "Understanding the City Size Wage Gap," *The Review of Economic Studies*, 79, 88–127.

BEAMAN, L. A. (2011): "Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the US," *The Review of Economic Studies*, 79, 128–161.

BEHRENS, K., G. DURANTON, AND F. ROBERT-NICOUD (2014): "Productive cities: Sorting, selection, and agglomeration," *Journal of Political Economy*, 122, 507–553.

BORJAS, G. J. (1985): "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants," *Journal of Labor Economics*, 3, 463–489.

——— (1987): "Self-Selection and the Earnings of Immigrants," *The American Economic Review*, 531–553.

——— (2003): "The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *The Quarterly Journal of Economics*, 118, 1335–1374.

BURSTEIN, A., G. HANSON, L. TIAN, AND J. VOGEL (2017): "Tradability and the Labor-Market Impact of Immigration: Theory and Evidence from the US," *NBER Working Paper*.

CARLSEN, F., J. RATTSØ, AND H. E. STOKKE (2013): "Education, Experience and Dynamic Urban Wage Premium," *mimeo*.

CHETTY, R., N. HENDREN, AND L. F. KATZ (2016): "The Effects of Exposure to Better Neighborhoods on Children: New Evidence from the Moving to Opportunity Experiment," *The American Economic Review*, 106, 855–902.

CHISWICK, B. R. (1978): "The Effect of Americanization on the Earnings of Foreign-Born Men," *Journal of Political Economy*, 86, 897–921.

COMBES, P.-P., G. DURANTON, AND L. GOBILLON (2008): "Spatial Wage Disparities: Sorting Matters!" *Journal of Urban Economics*, 63, 723–742.

COMBES, P.-P., G. DURANTON, L. GOBILLON, D. PUGA, AND S. ROUX (2012): "The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection," *Econometrica*, 80, 2543–2594.

CONSTANT, A. AND D. S. MASSEY (2002): "Return Migration by German Guestworkers: Neoclassical Versus New Economic Theories," *International Migration*, 40, 5–38.

DAMM, A. P. (2005): "The Danish Dispersal Policy on Refugee Immigrants 1986-1998: A Natural Experiment?" Tech. rep.

——— (2009): "Determinants of Recent Immigrants Location Choices: Quasi-experimental Evidence," *Journal of Population Economics*, 22, 145–174.

——— (2014): "Neighborhood Quality and Labor Market Outcomes: Evidence from Quasi-random Neighborhood Assignment of Immigrants," *Journal of Urban Economics*, 79, 139–166.

DAMM, A. P. AND C. DUSTMANN (2014): "Does Growing Up in a High Crime Neighborhood Affect Youth Criminal Behavior?" *The American Economic Review*, 104, 1806–1832.

DAMM, A. P. AND M. ROSHOLM (2010): "Employment Effects of Spatial Dispersal of Refugees," *Review of Economics of the Household*, 8, 105–146.

D'COSTA, S. AND H. G. OVERMAN (2014): "The Urban Wage Growth Premium: Sorting or Learning?" *Regional Science and Urban Economics*, 48, 168–179.

DE LA ROCA, J. AND D. PUGA (2017): "Learning by Working in Big Cities," *The Review of Economic Studies*, 84, 106–142.

DUSTMANN, C., A. GLITZ, AND T. FRATTINI (2008): "The Labour Market Impact of Immigration," *Oxford Review of Economic Policy*, 24, 477–494.

EDIN, P., P. FREDRIKSSON, AND O. ASLUND (2003): "Ethnic Enclaves and the Economic Success of Immigrants - Evidence from a Natural Experiment," *The Quarterly Journal of Economics*, 118, 329–357.

EDIN, P.-A., P. FREDRIKSSON, AND O. ÅSLUND (2004): "Settlement policies and the economic success of immigrants," *Journal of Population Economics*, 17, 133–155.

FOGED, M. AND G. PERI (2013): "Immigrants' and Native Workers: New Analysis on Longitudinal Data," Tech. rep., National Bureau of Economic Research.

FU, S. AND S. L. ROSS (2010): "Wage Premia in Employment Clusters: Agglomeration or Worker Heterogeneity?" *mimeo*.

GLAESER, E. (2011): *Triumph of the City: How Our Greatest Invention Makes Us Richer, Smarter, Greener, Healthier, and Happier*, Penguin.

GLAESER, E. L. AND D. C. MARE (2001): "Cities and Skills," *Journal of Labor Economics*, 19, 316–342.

GOULD, E. D., V. LAVY, AND M. D. PASERMAN (2004): "Immigrating to Opportunity: Estimating the Effect of School Quality Using a Natural Experiment on Ethiopians in Israel," *The Quarterly Journal of Economics*, 119, 489–526.

IMBERMAN, S. A., A. D. KUGLER, AND B. I. SACERDOTE (2012): "Katrina's children: Evidence on the Structure of Peer Effects from Hurricane Evacuees," *The American Economic Review*, 102, 2048–2082.

LAGAKOS, D., B. MOLL, T. PORZIO, N. QIAN, AND T. SCHOELLMAN (Forthcoming): "Life-Cycle Wage Growth Across Countries," *Journal of Political Economy*.

LAGAKOS, D. AND M. E. WAUGH (2013): "Selection, Agriculture, and Cross-country Productivity Differences," *American Economic Review*, 103, 948–80.

LINDENLAUB, I. (2017): "Sorting Multidimensional Types: Theory and Application," *The Review of Economic Studies*, 84, 718–789.

LUND, C. G. AND R. M. VEJLIN (2015): "Documenting and Improving the Hourly Wage Measure in the Danish IDA Database," .

MCKENZIE, D., S. STILLMAN, AND J. GIBSON (2010): "How Important is Selection? Experimental vs. Non-experimental Measures of the Income Gains from Migration," *Journal of the European Economic Association*, 8, 913–945.

MION, G. AND P. NATICCHIONI (2009): "The spatial sorting and matching of skills and firms," *Canadian Journal of Economics/Revue canadienne d'économique*, 42, 28–55.

OTTAVIANO, G. I. AND G. PERI (2012): "Rethinking the Effect of Immigration on Wages," *Journal of the European economic association*, 10, 152–197.

PAPAGEORGIOU, T. (2017): "Worker sorting and agglomeration economies," *mimeo*.

PETERS, M. (2017): "Can Immigration Cause Local Industrialization? Evidence from Germanys Post-War Population Transfer," in *2017 Meeting Papers*, Society for Economic Dynamics, 352.

PUGA, D. (2010): "The Magnitude and Causes of Agglomeration Economies," *Journal of Regional Science*, 50, 203–219.

ROSENTHAL, S. S. AND W. C. STRANGE (2004): "Evidence on the Nature and Sources of Agglomeration Economies," *Handbook of Regional and Urban Economics*, 4, 2119–2171.

SHOAG, D. AND N. CAROLLO (2016): "The Causal Effect of Place: Evidence from Japanese-American Internment," .

TOLBERT, C. M. AND M. SIZER (1996): "US Commuting Zones and Labor Market Areas: A 1990 Update," *ERS Staff Paper*.

WHEELER, C. H. (2006): "Cities and the Growth of Wages Among Young Workers: Evidence from the NLSY," *Journal of Urban Economics*, 60, 162–184.

YOUNG, A. (2013): "Inequality, the urban-rural gap, and migration," *The Quarterly Journal of Economics*, 128, 1727–1785.

A Appendix

This Appendix contains additional material for the paper. Section A.1 compares the sample of refugees considered in this paper to natives, and further compares refugees between the two assignment regions. In Section A.2 we examine employment rates conditional on joining the labour force by assignment. In Section A.3 we examine yearly hours worked. In A.4 we consider our main specification run on alternative spatial units. Section A.5 re-considers our baseline treatment effects for wages and earnings non-parametrically. In A.6 we run the main wage regression on refugees who never move out of the initially assigned location as this specification provides an upper-bound on the true returns to big city experience. Section A.7 describes the occupational data used throughout the paper, as well as a comparison of the distribution of occupations for natives and refugees by Copenhagen and Non-Copenhagen. Section A.8 provides evidence on take-up of education after arrival to Denmark and we show that it does not differ significantly for refugees assigned to Copenhagen and elsewhere. Section A.9 looks at the effects of ethnic enclaves in our main specification, and Section A.10 shows the results for alternative commuting zone delineations. Sections A.11 and A.12 report the structural model parameters and provide additional information on model fit, respectively.

A.1 Descriptive Statistics for Refugees

In this Section, we compare refugees to a cross-section of Danes, and for refugees only, we compare refugees initially assigned to Copenhagen and elsewhere. Further, we examine to what degree initial years of education for refugees balance between our two regions.

Table 4: Descriptive Statistics - Natives and Refugees

Comparison by:	Group		Assignment (Refugees only)	
	Natives	Refugees	Copenhagen	Non-Copenhagen
Age	36.72	28.24	28.67	28.08
Married	0.47	0.28	0.28	0.28
No. of children	0.68	0.54	0.47	0.57
Age of youngest child	7.43	3.46	3.63	3.40
Age of oldest child	10.01	7.27	7.20	7.30
Missing education	0.00	0.19	0.19	0.19
10 years of education	0.31	0.27	0.23	0.28
12 years of education	0.50	0.34	0.34	0.34
>12 years of education	0.17	0.20	0.23	0.19
Observations	1,335,545	20,493	5,530	14,963

Note: Only men aged between 19-55 are shown. Married is an indicator taking value 1 if the individual is married. No. of children is the number of children in the individual's family. Age of youngest (oldest) child is the age of the youngest (oldest) child in the individual's family. Missing education, 10 years of education, 12 years of education, and ≥ 12 years of education are all indicator variables.

The first two columns of Table 4 shows the comparison of natives and refugees. The group of natives is the 1990 cross-section of Danish males in our matched employer-employee dataset after applying the same sample selection criteria as for our baseline sample described in Section 2.2. We find that the refugees are on average younger when they arrive and enter the labor market than the average Dane in the 1990 cross-section. Accordingly, they are less likely to be married and have less and younger children. As Table 4 shows, for a fraction of refugees it was not possible to determine their level of education when arriving in Denmark. For natives, education level is obtained from national registers with high quality, thus information on education level is only missing in a few cases.

We now compare refugees assigned to Copenhagen and refugees assigned elsewhere at the first year we observe the refugees in our baseline sample. The last two columns of Table 4 shows descriptive statistics for the two groups of refugees. Recall that council officers observed age, nationality, and number of children when assigning refugees to municipalities. We confirm that larger families where assigned to Non-Copenhagen to an extent, as it was easier to house large families outside Copenhagen.

As educational level of refugees where not observed by council officers, we now check whether there is a significant difference in years of education between refugees assigned to Copenhagen and Non-Copenhagen upon arrival to Denmark.

Table 5: Regression of Initial Years of Education on Assignment Variables

	Years of Educ	Years of Educ
$InitCPH_i$	0.164***	0.0980
	(0.0491)	(0.0571)
Married	0.213***	0.145*
	(0.0617)	(0.0691)
No. of children	-0.121***	-0.0549*
	(0.0218)	(0.0247)
Age	0.431***	0.233***
	(0.0191)	(0.0243)
Observations	11812	7386
Sample	All	Educ≥12
Year Fixed Effects	Yes	Yes
Nationality Fixed Effects	Yes	Yes

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: $InitCph_i$ is an indicator taking value 1 if the individual was initially assigned to Copenhagen. Married is an indicator taking value 1 if the individual is married. No. of children is the number of children in the individual's family. Age is the age of the individual. Individuals with missing education information are dropped from the regression. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

We regress years of education on the assignment variables to control for variation in years of education explained by how council officers assigned refugees based on their available information, and a dummy for whether the refugee where assigned to Copenhagen or not. Table 5 shows the result of the regression[27]. Even after accounting for the information available to the council officers, we find a significant difference of 0.164 years of education for refugees assigned to Copenhagen. However, we do not regard this difference as evidence of systematic deviation from our quasi-randomization assumption, since it amounts to an average difference of two months of accumulated education for those assigned to CPH and NCPH.

A.2 Treatment and Employment Rates

In this Section we assess the effect of assignment to Copenhagen on employment rates throughout the sample period, conditional on having joined the labour force. This complements the analysis in Table 2, and together provides a more complete picture of the employment outcomes of refugees across space. In Table 6 we regress an indicator for being employed on the assignment variables, and 3-year dummy bins for time spent in Denmark. We restrict the sample to those who work for at least one year in our data.

[27] We perform this regression only for refugees whose information was not missing in the first year they appear in the dataset, which is larger fraction of the sample than the 19% for whom we never observe educational information from Table 4. We test and reject the possibility that this missing fraction differs between CPH and NCPH.

Table 6: Employment Rates by Initial Assignment

	(1)	(2)
	$logwage_{it}$	$logwage_{it}$
$InitCph_i$	0.00682	0.0198***
	(0.00357)	(0.00504)
3<= YearsinDenmark <6	0.251***	0.231***
	(0.00411)	(0.00488)
6<= YearsinDenmark <9	0.378***	0.318***
	(0.00418)	(0.00501)
9<= YearsinDenmark <12	0.492***	0.396***
	(0.00429)	(0.00520)
12<= YearsinDenmark <15	0.574***	0.466***
	(0.00445)	(0.00545)
15<= YearsinDenmark <18	0.607***	0.516***
	(0.00480)	(0.00593)
18<= YearsinDenmark <21	0.644***	0.545***
	(0.00610)	(0.00770)
No. Kids at Arr.	-0.00192	0.0101
	(0.00561)	(0.00552)
Married at Arr.	0.0266***	0.00972
	(0.00530)	(0.00672)
Age at Arr.	-0.00489***	-0.00929***
	(0.000282)	(0.000322)
Observations	116096	81740
R^2	0.278	0.253
Sample	Educ>=12	Educ<12
Nationality FE	Yes	Yes
Cohort FE	Yes	Yes

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

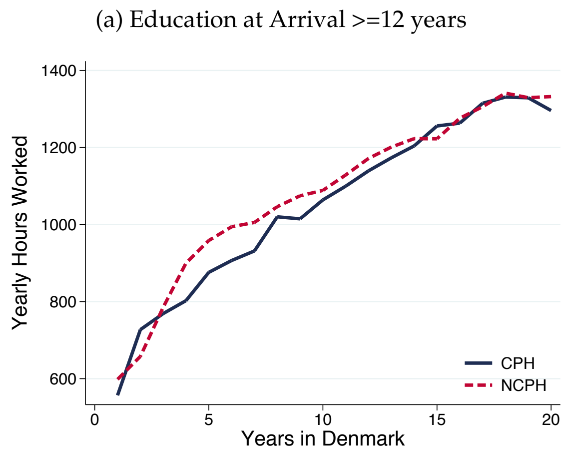
Note: $InitCph_i$ is an indicator taking value 1 if the individual was initially assigned to Copenhagen. Years in Denmark is years since arrival in Denmark, binned into three year bins. No. Kids at Arr. is the number of children in the individual's family at arrival. Married at Arr. is an indicator taking value 1 if the individual was married at arrival. Age at Arr. is the age of the individual upon arriving in Denmark. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

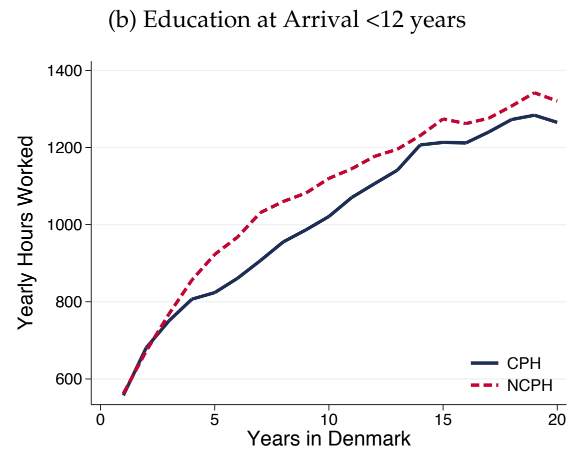
In the first column we report the results for those with at least a high-school degree. We see that assignment to Copenhagen had no significant impact on average employment rates over the lifetime of the refugees for this group. This is not true for those without a high-school degree; controlling for the assignment relevant characteristics, and conditional on joining the labour force, employment rates are 2% higher if initially assigned to Copenhagen.

A.3 Hours Worked

In this Section, we present evidence of assignment to the city on hours worked, conditional on working. Here we plot average hours worked conditional on working that year, by years in Denmark. For our preferred sample, it is true that initially refugees work less hours in Copenhagen, and eventually catch up to those assigned outside Copenhagen. This is consistent with our result that there is a static negative earnings premium from being assigned to Copenhagen.

Figure 8: Mean Yearly Hours Worked





Note: Hours worked are the estimated total hours worked in that year for each individual. Here we take the mean across all observations in each sample, conditional on employment.

A.4 Alternative Spatial Units

Here we consider our baseline treatment regressions from Table 1 utilizing different spatial units as the focus of analysis. In Table 7 we report the results when we leave out the three second-tier cities of Aarhus, Aalborg and Odense, effectively obtaining the differential treatment of being assigned to Copenhagen versus a rural area. We find a moderately larger Dynamic Treatment Effect for hourly wages and earnings than in the baseline, as might be expected.

Table 7: Wage-Experience Profiles, No Small Cities

	(1) $logwage_{it}$	(2) $logwage_{it}$	(3) $logwage_{it}$	(4) $logearnings_{it}$	(5) $logearnings_{it}$	(6) $logearnings_{it}$
Exp_{it}	0.0213***	0.0231***	0.0204***	0.0733***	0.0741***	0.0751***
	(0.00165)	(0.00186)	(0.00127)	(0.00443)	(0.00449)	(0.00450)
$InitCph_i$	-0.0156	-0.00688	-0.0266	-0.0932**	-0.0855***	-0.109*
	(0.00927)	(0.00815)	(0.0153)	(0.0254)	(0.0212)	(0.0380)
$InitCph_i \times Exp_{it}$	0.00991***	0.00960***	0.00928***	0.0239***	0.0226***	0.0269***
	(0.00121)	(0.00144)	(0.00102)	(0.00378)	(0.00379)	(0.00413)
No. Kids at Arr.	-0.0108	0.0283*	-0.0159	-0.0279	0.0487*	-0.0489
	(0.00902)	(0.0103)	(0.0142)	(0.0144)	(0.0205)	(0.0254)
Married at Arr.	0.0342*	0.00987	0.0279	0.0828***	0.0297	0.0868**
	(0.0125)	(0.0142)	(0.0165)	(0.0207)	(0.0260)	(0.0229)
Age	-0.00173	-0.00275*	-0.00312**	0.000117	-0.00113	-0.00331
	(0.00136)	(0.00120)	(0.000997)	(0.00294)	(0.00265)	(0.00249)
Observations	67128	40828	26300	73889	44957	28932
R^2	0.054	0.060	0.052	0.158	0.159	0.162
Sample	All	Educ≥12	Educ<12	All	Educ≥12	Educ<12
Nationality FE	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: Specification given in Equation (1), estimated leaving out individuals assigned to Aarhus, Aalborg or Odense, with the same sub-samples as those in Table 1. Robust standard errors clustered at the level of initial commuting zone. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Similarly, in Table 8 we leave out the refugees assigned to Copenhagen, comparing those assigned to a small city to rural area. We code $InitSmallCity_i$ as an indicator if the refugee was assigned to any of the three commuting zones that include Aalborg, Aarhus or Odense. The dynamic premium is smaller (and not significant), again according with intuition.

Table 8: Wage-Experience Profiles, No Copenhagen

	(1) $logwage_{it}$	(2) $logwage_{it}$	(3) $logwage_{it}$	(4) $logearnings_{it}$	(5) $logearnings_{it}$	(6) $logearnings_{it}$
Exp_{it}	0.0220***	0.0237***	0.0204***	0.0760***	0.0770***	0.0759***
	(0.00127)	(0.00156)	(0.00116)	(0.00371)	(0.00387)	(0.00425)
$InitSmallCity_i$	-0.0385*	-0.0418**	-0.0325	-0.0489	-0.0762**	-0.0100
	(0.0166)	(0.0120)	(0.0219)	(0.0292)	(0.0251)	(0.0423)
$InitSmallCity_i \times Exp_{it}$	0.00466	0.00590*	0.00304	0.00585	0.00962	0.00126
	(0.00252)	(0.00277)	(0.00246)	(0.00456)	(0.00491)	(0.00577)
No. Kids at Arr.	-0.0265**	0.00166	-0.0262	-0.0352	0.0261	-0.0415
	(0.00867)	(0.0103)	(0.0126)	(0.0171)	(0.0266)	(0.0251)
Married at Arr.	0.0534***	0.0387*	0.0358*	0.107***	0.0677**	0.0904**
	(0.0105)	(0.0155)	(0.0153)	(0.0153)	(0.0181)	(0.0242)
Age	-0.00244**	-0.00313***	-0.00333***	-0.00263	-0.00408**	-0.00420*
	(0.000695)	(0.000757)	(0.000645)	(0.00130)	(0.00127)	(0.00177)
Observations	75615	43625	31990	83062	47931	35131
R^2	0.054	0.058	0.058	0.150	0.152	0.154
Sample	All	Educ≥12	Educ<12	All	Educ≥12	Educ<12
Nationality FE	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes

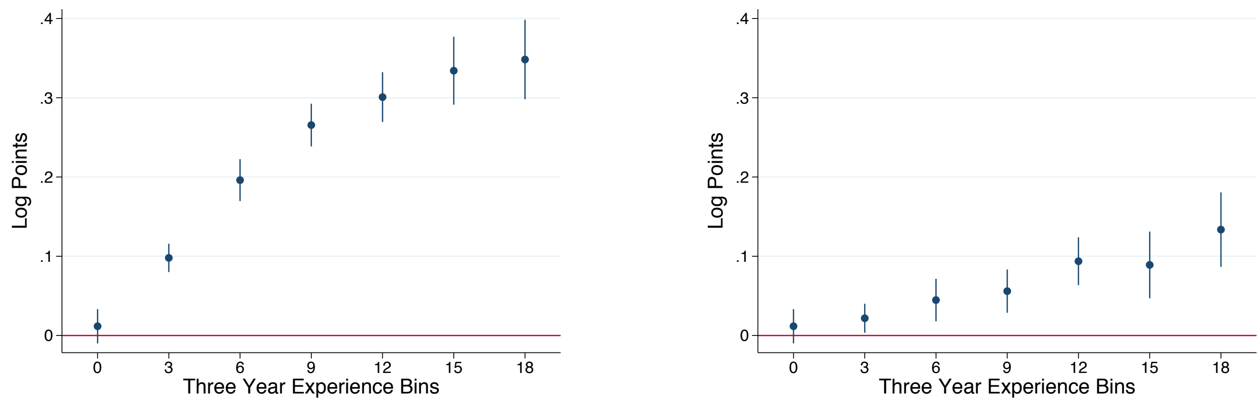
Note: Specification given in Equation (1), estimated leaving out individuals assigned to Copenhagen, with the same sub-samples as those in Table 1. Robust standard errors clustered at the level of initial commuting zone. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

A.5 Non-Parametric Treatment Effects

In this Section we re-examine the evidence in 3.1.1 non-parametrically. We re-estimate equation (1) for hourly wages using dummies for 3-year bins for experience, and the interaction of these bins with being initially placed in Copenhagen. Figure 9 shows the results. In line with most studies, we find concavity in the effects of raw experience on wages. We continue to find no significant difference for initial wages for those placed in Copenhagen, and the size of the Dynamic Treatment Effect after 18-20 years is in line with that implied by the linear regressions.

Figure 9: Non-Parametric Coefficients for Wages

(a) Coefficients for Experience (b) Interaction Terms with $InitCph_i$

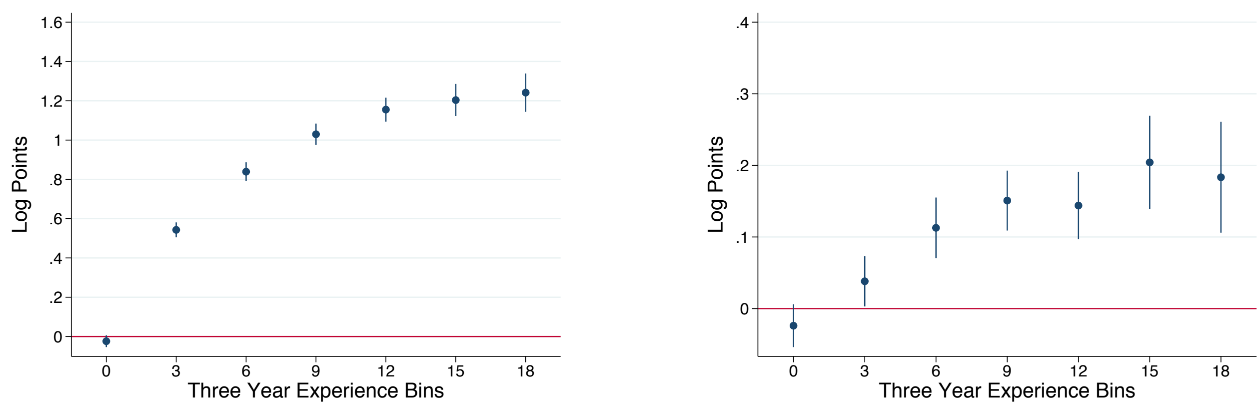


Note: Specification given in Equation (1), with 3-year dummies for experience for experience. 95% confidence intervals shown.

In Figure 10 we report the same regression for yearly earnings. We continue to find a small negative initial premium, but this is no longer statistically significant.

Figure 10: Non-Parametric Coefficients for Earnings

(a) Coefficients for Experience (b) Interaction Terms with $InitCph_i$



Note: Specification given in Equation (1), with 3-year dummies for experience. 95% confidence intervals shown.

A.6 Static and Dynamic Treatment Effects for Stayers Only

In this Section, we replicate our main wage-experience profiles (Table 1) for a sample of refugees who never cross the border of the commuting zone of Copenhagen and elsewhere. Under simple assumption, this sample of refugees provide an upper-bound to true return to big city experience as explained in Section 3.1.

Table 9: Wage-Experience Profiles, Stayers Only

	(1) $logwage_{it}$	(2) $logwage_{it}$	(3) $logwage_{it}$	(4) $logearnings_{it}$	(5) $logearnings_{it}$	(6) $logearnings_{it}$
Exp_{it}	0.0210***	0.0234***	0.0194***	0.0706***	0.0734***	0.0696***
	(0.00153)	(0.00168)	(0.00151)	(0.00329)	(0.00414)	(0.00326)
$InitCph_i$	-0.00122	0.0121	-0.0194	-0.134***	-0.111***	-0.173***
	(0.00946)	(0.00765)	(0.0136)	(0.0190)	(0.0213)	(0.0235)
$InitCph_i \times Exp_{it}$	0.0115***	0.0105***	0.0119***	0.0324***	0.0299***	0.0364***
	(0.00148)	(0.00165)	(0.00155)	(0.00301)	(0.00363)	(0.00294)
No. Kids at Arr.	-0.0125	0.0289	-0.0228	-0.0207	0.0715*	-0.0537*
	(0.0108)	(0.0202)	(0.0126)	(0.0175)	(0.0325)	(0.0196)
Married at Arr.	0.0446**	0.0237	0.0337**	0.0983***	0.0419	0.103***
	(0.0135)	(0.0251)	(0.0109)	(0.0194)	(0.0349)	(0.0172)
Age	-0.00206	-0.00303**	-0.00313***	-0.00265	-0.00487	-0.00421*
	(0.001000)	(0.000937)	(0.000626)	(0.00243)	(0.00265)	(0.00165)
Observations	75066	43429	31637	82553	47732	34821
R^2	0.056	0.062	0.057	0.153	0.156	0.154
Sample	Stayers	Educ≥12	Educ<12	Stayers	Educ≥12	Educ<12
Nationality FE	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: Specification given in Equation (1), with 3-year dummies for experience. 95% confidence intervals shown.

As for the main wage-experience profiles we find no Static Treatment Effect for this sample. The Dynamic Treatment Effect increases from 0.81% to 1.2%. Table 9 shows the results.

A.7 Occupation Codes and Distributions Across Space

The Danish micro-data features detailed 4-digit occupation codes. For the period between 1992 and 2010, the first digit of these codes coincides with International Standard Classification of Occupations (ISCO) 1-digit codes, which we list in the first column of Table 10.

We use the full sample of male Danes to compute average wages by 1-digit occupations for every year between 1992 and 2010. Then we rank occupations according to their average wage for each year. In Column 3 of Table 10, we state the average rank of each occupation across the years 1992-2009. The description of the occupation in Column 1 and the detailed ranking based on the wages in Column 2, naturally suggests a coarser classification of the nine occupations into three comprehensive groups listed in Column 3: high-skill, low-skill and manual professions. The high-skill occupations are also the professions with the highest average wage rank, the low-skill professions those with the lowest and the manual occupation nestled in between with an intermediate average wage rank. We document these coarser wage ranks in Column 4 of Table 10.

Table 10: Description of Occupation Codes and Classifications

ISCO Description (Code)	Wage Rank	Type	Wage Rank
Managers (1)	1	High-Skill	1
Professionals (2)	2	High-Skill	1
Technicians and associate professionals (3)	3	High-Skill	1
Clerical support workers (4)	8	Low-Skill	3
Service and sales workers (5)	9	Low-Skill	3
Skilled agricultural, forestry and fishery workers (6)	5	Manual	2
Craft and related trades workers (7)	4	Manual	2
Plant and machine operators and assemblers (8)	6	Manual	2
Elementary occupations (9)	7	Low-Skill	3

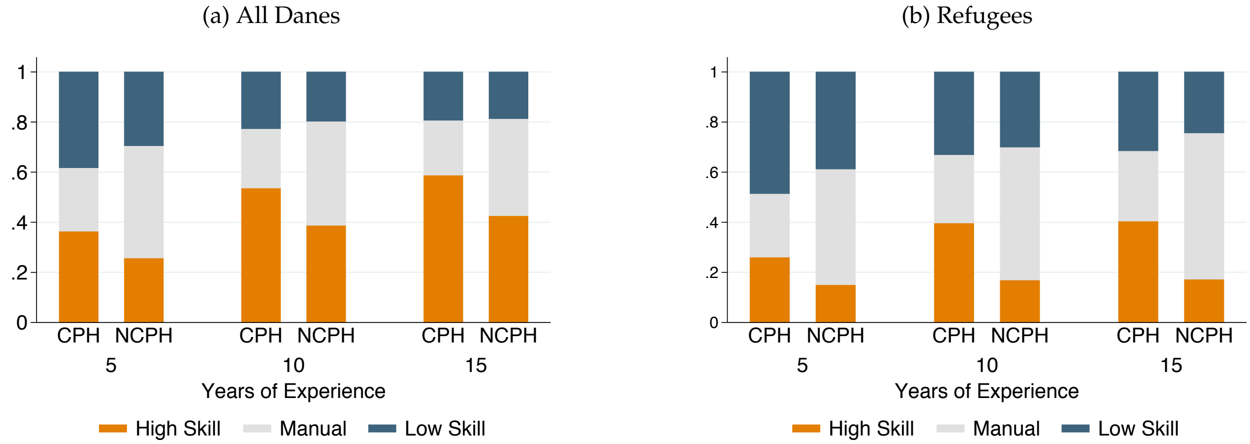
Note: The ISCO Description codes are based on the Danish implementation of the international used ISCO description codes called DISCO. Wage rank is computed as the average yearly rank of average wages of Danes on the 1 digit occs in each year.

In regressions throughout the paper we use the 9 1-digit codes to control for occupation fixed effects. In the estimation of the finite mixture model as well as for some of the graphs in the mechanism section, we employ the coarser ranking.

In the full sample of all Danish men we group workers into the coarse occupation groups outlined in Table 10. Then we categorize workers by their years of labor market experience defined as number of years in which they were associated with an establishment identifier at the point of data collection. The left panel of Figure 11 shows the distribution of Danish men with 5, 10 and 15 years of labor market experience across these occupation groups, by their current location (Copenhagen versus non-Copenhagen). The left panel of Figure 11 shows that in Copenhagen the fraction of people doing manual work is much lower than outside of Copenhagen. The flip-side of this is that outside Copenhagen the fraction of people doing low-skill and high-skill work is persistently lower than in Copenhagen. Additionally, we see that as workers increase their experience they seem to move from low-skill into high-skill employment, leaving the fraction of manual workers in both locations roughly stable over time.

The right panel of Figure 11 shows the identical graph for our male refugee sample. It differs from Figure 5 in that the location specification is not based on initial allocation but on current location for direct comparability with the left hand panel. Like for the native population there are substantially more refugees employed in manual work outside of Copenhagen than in Copenhagen in every year in the data. Differentially from the Danish population, less refugees in both locations work in high-skill occupations regardless of the years of experience. The fraction of low-skill workers is higher for both workers in Copenhagen and outside, however it declines substantially with refugees outside Copenhagen moving mainly into manual occupations, while in Copenhagen refugees seem to make the transition from low-skill into high-skill occupations, whose share increases by almost 20%.

Figure 11: Natives and Refugees by Occupations, Locations and Experience



Note: The left panel shows the distribution of male Danes with 5/10/15 years of experience, across occupation groups within Copenhagen and Non-Copenhagen, respectively. The right panel shows the same data for our preferred refugee sample.

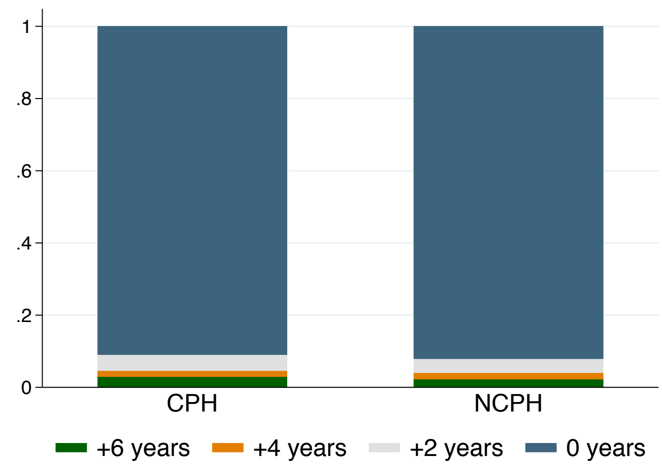
Overall, we see a majority of refugees entering into low-skill service jobs, which include cleaning and sales jobs. From there they move on into higher paying occupations that differ by location: outside Copenhagen manual careers are more common, while within Copenhagen workers predominantly move into various high-skill occupations. Given Figure 2 we know that for the refugee sample, the current location coincides for a majority of workers with their location of initial allocation. This explains the similarity of the right panel of Figure 11 with Figure 5.

A.8 Educational Take-Up

In this Section, we test whether refugees initially assigned to Copenhagen take-up more years of education than refugees assigned elsewhere. Figure 12 shows the fraction of refugees who take-up education after entering our sample by initial assignment, and the amount they take up. We see that the majority of refugees in both regions do not take-up additional years of education. We formally test the differences in educational take-up between the two groups of refugees.

Table 11 shows the result of a t-test of mean differences between the two groups on an indicator variable for taking-up any additional years of education and four separate indicator variable, one for each different amount of additional years of education obtained as in Figure 12. The second column repeats this exercise for only those who had at least a high-school degree upon arrival. In both cases we see only a tiny difference across areas in take-up, equivalent to an extra 0.05 years of schooling in the full sample. This rules out differential educational take-up as driving our results.

Figure 12: Education Take-Up for Refugees by Initial Assignment



Note: Bar plot showing the fraction of refugees who take up no education, 2-3 years of education, 4-5 years of education, and 6 or more years of education, respectively. The left column is based on refugees assigned to Copenhagen and the right column is based on refugees assigned to Non-Copenhagen.

Table 11: T-test of Differences in Take-Up of Additional Years of Education

	CPH	CPH
Years of Additional Education	0.0508*	0.0729**
	(0.0241)	(0.0241)
No Take-Up	-0.0114*	-0.0207***
	(0.00563)	(0.00562)
2 Years of Additional Education	0.00555	0.0122***
	(0.00403)	(0.00322)
4 Years of Additional Education	-0.00109	0.000922
	(0.00267)	(0.00354)
6 Years of Additional Education	0.00699*	0.00759*
	(0.00318)	(0.00315)
Observations	11812	7386
Sample	All	Educ≥12

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Years of Additional Education is the years of additional education at the latest observation of an individual. All other variables are coded as an indicator of whether the individual took up a certain number of years of education. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

A.9 Effects of Ethnic Enclaves

In this Section we test the power of a refugee-specific mechanism in explaining out estimated treatment effects. We re-estimate our main specification in Table 12 including a control for the presence of co-nationals of the refugee at the municipality level. In particular, we compute the stock of foreign-born residents in each municipality every year, for each of the nationalities studied in the paper. Then, for every refugee we construct a variable that records the number of co-nationals residing in their municipality of assignment in their year of arrival, and include this in the regression. This is a simple method for checking whether the presence of co-nationals at a micro municipality-level is driving our treatment effects at the city level, since as Damm (2014) notes, prior to the dispersal policy immigrants and refugees were overwhelmingly clustered in Copenhagen and the other larger cities. Table 12 suggests this is not the case; the estimate for the Dynamic Treatment Effect is virtually the same as that reported in Table 1.

Table 12: Controls for Ethnic Enclaves

	(1) $logwage_{it}$	(2) $logwage_{it}$	(3) $logwage_{it}$	(4) $logearnings_{it}$	(5) $logearnings_{it}$	(6) $logearnings_{it}$
Exp_{it}	0.0231***	0.0252***	0.0213***	0.0766***	0.0789***	0.0756***
	(0.00137)	(0.00154)	(0.00126)	(0.00307)	(0.00357)	(0.00336)
$InitCph_i$	0.00130	0.00964	-0.00870	-0.0704***	-0.0517**	-0.103***
	(0.00883)	(0.00759)	(0.0129)	(0.0171)	(0.0174)	(0.0237)
$InitCph_i \times Exp_{it}$	0.00776***	0.00702***	0.00790***	0.0205***	0.0178***	0.0255***
	(0.00147)	(0.00162)	(0.00132)	(0.00294)	(0.00324)	(0.00301)
No. Kids at Arr.	-0.0206*	0.0112	-0.0208	-0.0316*	0.0356	-0.0418*
	(0.00768)	(0.0105)	(0.0100)	(0.0131)	(0.0214)	(0.0192)
Married at Arr.	0.0370**	0.0187	0.0263*	0.0814***	0.0350	0.0827***
	(0.0123)	(0.0170)	(0.0118)	(0.0169)	(0.0231)	(0.0175)
Age at Arr.	-0.00123	-0.00205*	-0.00259**	0.000418	-0.00105	-0.00220
	(0.000955)	(0.000886)	(0.000691)	(0.00223)	(0.00217)	(0.00180)
$log(EthnicStock_i)$	-0.00958***	-0.0107*	-0.00596***	-0.0239***	-0.0263**	-0.0168**
	(0.00244)	(0.00408)	(0.00146)	(0.00579)	(0.00855)	(0.00515)
Observations	97402	57994	39408	107297	63870	43427
R^2	0.057	0.063	0.055	0.157	0.160	0.157
Sample	All	Educ≥12	Educ<12	All	Educ≥12	Educ<12
Nationality FE	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: $Log(EthnicStock_i)$ is a variable which records for each refugee the number of co-nationals of the residing in their municipality in the year of assignment. Robust standard errors clustered at the level of initial commuting zone. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

We also reestimate these regressions with the ethnic stock *in each year* as a control, instead of year of arrival, addressing concerns that refugees could sort into ethnic enclaves over time, and that this might assist them with employment opportunities. The results are almost identical to the table above.

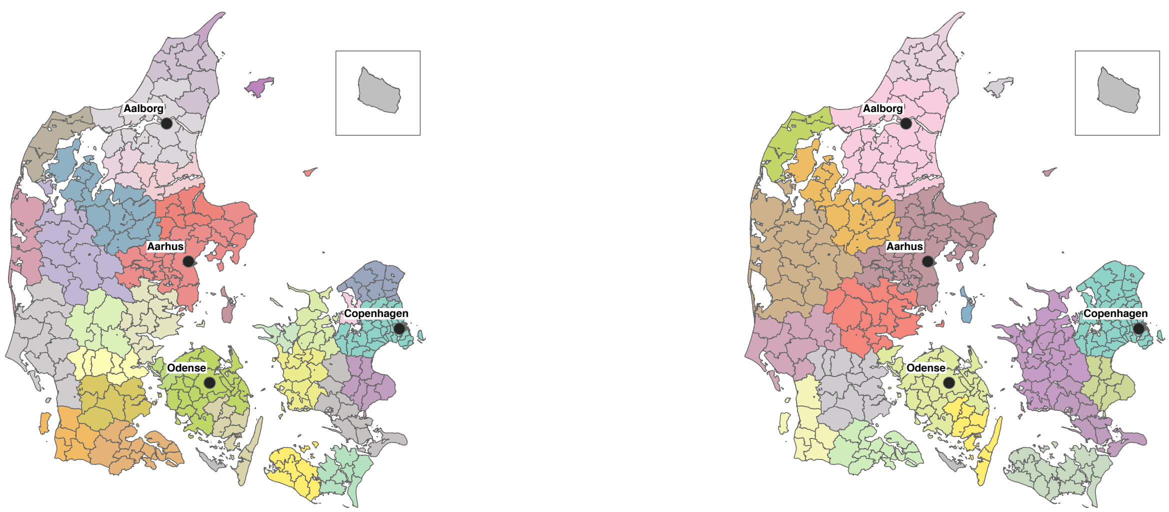
A.10 Alternative Commuting Zone Definition

In this Section we examine our main empirical results using commuting zones based on different years of commuting data. The computed commuting zones change significantly over the sample period, as Denmark becomes more integrated across municipalities resulting in increasing commuting flows. To illustrate this, we plot the computed commuting zones for 1980 and for 2000 in Figure 13.

Figure 13: Alternative Commuting Zones

(a) 1980 Commuting Flows

(b) 2000 Commuting Flows



Note: Commuting Zones constructed based on 1980 (left) and 2000 (right) cross-municipality commuting flows constructed from work and residence identifier in the IDA data set.

We see that the number of commuting zones decreases markedly between 1980 and 2000. However, what matters for our analysis is the delineation between the Copenhagen commuting zone and everywhere else. Beyond 1986, this delineation does not change. As such, we reestimate our main specification using only the 1980 commuting zone. As can be seen, this is a smaller zone which contains more municipalities which correspond to the proper city of Copenhagen. Our results for log hourly wages can be found in Table 13. The reported coefficients are larger in magnitude than our main specification. This accords with intuition; the 1980 commuting zone consists of the urban core of Copenhagen, whereas the 1986 commuting zone includes Northern Zealand, which is not itself a major center of employment.

Table 13: Wage-Experience Profiles, 1980 Commuting Zone

	(1) $logwage_{it}$	(2) $logwage_{it}$	(3) $logwage_{it}$	(4) $logearnings_{it}$	(5) $logearnings_{it}$	(6) $logearnings_{it}$
Exp_{it}	0.0228***	0.0248***	0.0212***	0.0765***	0.0790***	0.0755***
	(0.00123)	(0.00147)	(0.00112)	(0.00273)	(0.00319)	(0.00304)
$InitCph_i$	-0.0101	-0.00415	-0.0161	-0.0870***	-0.0601**	-0.135***
	(0.00954)	(0.00877)	(0.0126)	(0.0144)	(0.0173)	(0.0174)
$InitCph_i \times Exp_{it}$	0.0108***	0.0101***	0.0104***	0.0241***	0.0201***	0.0318***
	(0.00132)	(0.00149)	(0.00126)	(0.00250)	(0.00280)	(0.00273)
No. Kids at Arr.	-0.0199**	0.0139	-0.0211*	-0.0299	0.0424	-0.0427
	(0.00643)	(0.0126)	(0.00964)	(0.0150)	(0.0211)	(0.0239)
Married at Arr.	0.0397***	0.0204	0.0289*	0.0889***	0.0406*	0.0896***
	(0.0106)	(0.0150)	(0.0124)	(0.0136)	(0.0198)	(0.0203)
Age at Arr.	-0.00143	-0.00227*	-0.00276***	-0.000144	-0.00167	-0.00270
	(0.000843)	(0.000833)	(0.000650)	(0.00187)	(0.00193)	(0.00167)
Observations	97402	57994	39408	107297	63870	43427
R^2	0.057	0.062	0.055	0.155	0.158	0.157
Sample	All	Educ≥12	Educ<12	All	Educ≥12	Educ<12
Nationality FE	Yes	Yes	Yes	Yes	Yes	Yes
Cohort FE	Yes	Yes	Yes	Yes	Yes	Yes

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Specification given in Equation (1), using the commuting flows in 1980 to construct commuting zones and initial allocation zones. Robust standard errors clustered at the level of initial commuting zone using 1980 allocation zones. Standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

A.11 Structural Parameter Estimates

In Tables 14 and 15 we present the parameter estimates of our maxium likelihood estimation. We allow all relevant parameters to differ by unobserved type in this specification. The only parameters we fix exogenously are the discount rate, ρ, which we set to 0.98, and the unemployment benefit b, which we set to 3.5. Two parameters, a_{NCPH} and τ_{NCPH}, are normalized, as such a_{NCPH} estimates the difference in amenities between NCPH and CPH and τ_{NCPH} estimates the difference in moving cost.

A.12 Model Fit

In Figure 14 we show the density of wage observations for the actual and simulated data, where the simulated data is split by worker type. Almost all of the mass in the upper tail is accounted for by high ability workers.

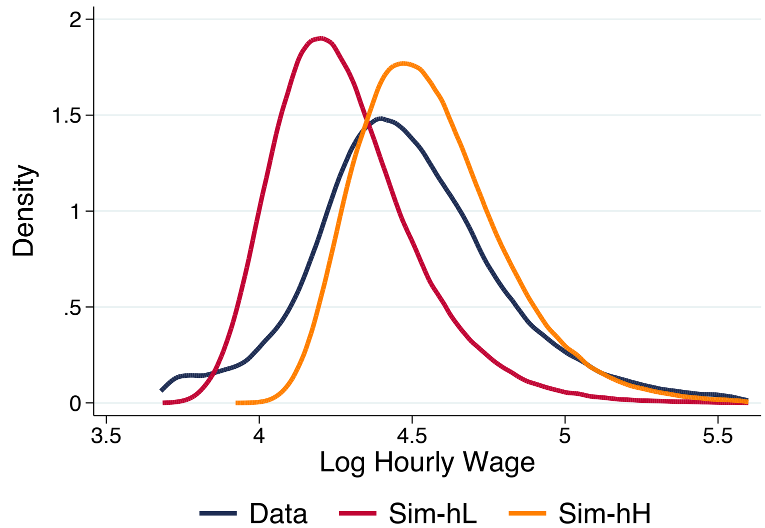
Table 14: Estimates of the Model Parameters- Base and Wage

Base Parameters			Wage Parameters		
Decription	Parameter	Estimate	Description	Parameter	Estimate
Amenity in NCPH	a_{NCPH}	0 -	Base wage	$\bar{w}$	4.075 (0.002)
Amenity in CPH	a_{CPH}	-0.062 (0.003)	High-type fixed effect	θ_{h_H}	0.269 (0.002)
Moving cost to CPH	τ_{NCPH}	1 -	High firm type fixed effect	ψ_{f_H}	0.095 (0.003)
Moving cost to NCPH	τ_{CPH}	6.823 (0.073)	H-H Complementarity	α_{h_H,f_H}	0.089 (0.003)
S.D. of moving cost shock	σ_η	0.967 (0.008)	Return to Exp. for L-type at L-firm	$\beta^{h_L}_{f_L}$	0.026 (0.003)
S.D. of match quality shock	σ_u	5.816 (0.029)	Return to Exp. for H-type at L-firm	$\beta^{h_H}_{f_L}$	0.012 (0.001)
Fraction of L types	χ_L	0.518 (0.009)	Return to Exp. for L-type at H-firm	$\beta^{h_L}_{f_H}$	0.035 (0.001)
			Return to Exp. for H-type at H-firm	$\beta^{h_H}_{f_H}$	0.024 (0.001)
			Quadratic on Experience	β_2	-0.0005 (0.637)

Table 15: Estimates of the Model Parameters- Search Parameters

	Copenhagen		Non-Copenhagen	
Decription	Parameter	Estimate	Parameter	Estimate
Destruction rate for L-type	$\delta^{h_L}_{CPH}$	0.254 (0.008)	$\delta^{h_L}_{NCPH}$	0.283 (0.005)
Destruction rate for H-type	$\delta^{h_H}_{CPH}$	0.194 (0.003)	$\delta^{h_H}_{NCPH}$	0.200 (0.002)
Reallocation shock for L-type	$\mu^{h_L}_{CPH}$	0.295 (0.013)	$\mu^{h_L}_{NCPH}$	0.397 (0.021)
Reallocation shock for H-type	$\mu^{h_H}_{CPH}$	0.210 (0.013)	$\mu^{h_H}_{NCPH}$	0.220 (0.021)
Job-finding rate for L-type	$\underline{\lambda}^{h_L}_{CPH}$	0.201 (0.012)	$\underline{\lambda}^{h_L}_{NCPH}$	0.251 (0.034)
Job-finding rate for H-type	$\underline{\lambda}^{h_H}_{CPH}$	0.240 (0.007)	$\underline{\lambda}^{h_H}_{NCPH}$	0.311 (0.005)
Low-firm offer prob. for L-type	$\pi^{h_L}_{CPH}$	0.762 (0.012)	$\pi^{h_L}_{NCPH}$	0.826 (0.011)
Low-firm offer prob. for H-type	$\pi^{h_H}_{CPH}$	0.682 (0.012)	$\pi^{h_H}_{NCPH}$	0.815 (0.011)

Figure 14: Simulated Vs. Actual Wage Densities



Note: Graph shows density plot of simulated wages vs. actual wages. Simulated wages were obtained by simulating the life-histories of 100,000 agents given equal distribution of types across locations.

B Online Data Appendix

This web-based data appendix contains supplementary material on the construction of the dataset

B.1 Explanation of Variables

- **Age:** Age of individual obtained by DST from the Central Danish Person register (CPR) and kept in IDA-P.
- **Age of youngest child:** In a given year, age of youngest child within the family obtained from DST' register FAMILIE.
- **Age of oldest child:** In a given year, age of youngest child within the family obtained from DST' register FAMILIE.
- **Date of migration:** Date of migration to Denmark recorded by the Danish immigration office.
- **Earnings:** Earnings of the primary November job in IDA-N. Earnings are reported by employers to the Danish tax authorities.
- **Ethnic stock:** The stock of individuals of the same nationality at the same municipality.
- **Gender:** Gender of individual obtained by DST from the Central Danish Person register (CPR) and kept in IDA-P.
- **Hourly wages:** Hourly wage of the November job in IDA-N. Computed from earnings and hours worked of the job. Hours worked is estimated by Statistics Denmark using the amount of mandatory ATP pension payments of the job (also reported by employers to the Danish Tax Authorities). The rate of ATP pension payments vary by hours worked per day and thus offers a way to estimate the number of hours worked for a given job.
- **Industry:** Industry of the workplace obtained from IDA-S. The industry variable takes the following values
  - **A:** Agriculture, forestry, and fishery
  - **B:** Winning and quarrying
  - **C:** Manufacturing
  - **D:** Electricity, gas, steam, and air-conditioning supply
  - **E:** Water supply, sewerage contractors, waste management, and remediation activities
  - **F:** Construction

- **G:** Wholesale and retail trade, repair of motor vehicles and motorcycles
- **H:** Transport and storage
- **I:** Accommodation and food service activities
- **J:** Information and communication
- **K:** Financial and insurance activities
- **L:** Real estate activities
- **M:** Professional, scientific, and technical activities
- **N:** Administrative and support service activities
- **O:** Public administration and defense, compulsory social security
- **P:** Education
- **Q:** Human health and social work activities
- **R:** Arts, entertainment, and recreation
- **S:** Other service activities
- **T:** Activities of household as employers, undifferentiated goods- and services-producing activities of households for own use
- **U:** Activities of extraterritorial organizations and bodies
- **X:** Unknown

- **Initial assignment to the Copenhagen commuting zone ($InitCph_i$):** Boolean variable indicating whether a refugee was initially assigned to the Copenhagen commuting zone or not.
- **Married at arrival:** An indicator, 1, if an individual was married upon arrival to Denmark.
- **Municipality of residence:** A unique identifier for each of the 271 municipalities in Denmark. Source, DST IDA-P (for municipality of residence), IDA-S (for municipality of workplace).
- **Municipality of a workplace:** The municipality of workplace from IDA-S.
- **Nationality/Country of origin:** Nationality of the individual. If the individual is a refugee then it is assumed that the origin country is equal to the birth country.
- **Number of children at arrival (*No. kids at Arr.*):** Number of kids at arrival to Denmark.
- **Occupation:** Occupation information of the November in IDA-N from IDA-P. The occupational codes (DISCO) is a Danish version of ISCO codes (international standard classification of occupations).

- **Spousal identifier:** A person identifier of the spouse of an individual from DST' register FAMILIE.
- **Value added taxes:** Value added taxes in a given year is reported by the firm to the Danish tax authorities. It is computed as total sales minus total purchases.
- **Year of arrival:** We define the year of arrival to Denmark as the first year we observe a refugee in IDA-P.
- **Years in Denmark:** Number of years observed in Denmark after imposing our sample selection.
- **Years of education:** Refugees are interviewed upon arrival to Denmark about their level of education. For natives, years of education is obtained from an administrative educational register.
- **Years of experience (Exp_{it}):** Computed as the number of years worked, i.e. the number of times observed in a primary November job in IDA-N.
- **Years of experience at high-productive establishments ($HighExp_{it}$):** Years of experience at establishments estimated to be in the top three deciles of the average establishment wage distribution.
- **Years of experience at other establishments ($OtherExp_{it}$):** Years of experience at establishments estimated to be outside the top three deciles of the average establishment wage distribution.
- **Workplace:** Workplace identifier from IDA-S. Every workplace is defined as either fictive and non-fictive. A fictive workplace is assigned to an individual if the job can not be assigned to any physical workplace.

B.2 Details on the Construction of the Data Set

We construct our matched employer-employee panel from five sources:

1. Information from IDA, a Danish register-based matched employer-employee dataset constructed by Statistics Denmark.
2. Information on firms' sales and purchases from firm-level VAT data administered by the Danish tax authorities.
3. Between country migration information from Statistics Denmark dataset EPERSONER.
4. Family data including number of children, age of children, etc. from Statistics Denmark dataset FAMILIE.

5. Income data including total yearly labor market earnings from Statistics Denmark dataset INDH.

IDA data. We use three sub-panels within IDA: IDA-P, IDA-N, and IDA-S. IDA-P contains basic characteristics for individuals aged between 15-74 residing legally in Denmark on the 31st of December in a given year. We keep information from IDA-P on gender, age, municipality of residence, years of education.[28] The unit of observation in IDA-P is a person-year. IDA-N contains labor market information constructed using annual tax filings obtained from the Danish tax authorities. IDA-N contains all workers' employment relations. For a given individual, Statistics Denmark defines the job with highest earnings on November 28th of a given year as the individual's primary employment relation. We retain information on an individual's primary employment to construct hourly wage and annual earnings, and the employer's firm identifier. However, we also record the individual's total earnings for the year.[29] The unit of observation in IDAN-N is a person-year. IDA-S contains information on all physical workplaces within a firm in Denmark. Employment which takes place at changing locations is said to take place at a fictitious workplace, which Statistics Denmark do not keep information on.[30] We retain information on industry of the workplace and whether or not the workplace is in the public sector. The unit of observation in the (aggregated) IDA-S is a firm-year.

VAT data. Data on sales and purchases at the firm level are obtained from the Statistics Denmarks panels MOMS and MOMM, which constructed from firm VAT accounts from the Danish Tax Authorities. Firms settle VAT either monthly, quarterly or yearly depending on size of revenue. MOMS covers the period 1995-2000 and contains annual sales and purchases. MOMM is a monthly panel starting in 2001.[31] We aggregate MOMM data to a yearly frequency.[32] The unit of observation is a firm-year.

EPERSONER data. The dataset contains cross-sectional information on all individuals living in Denmark by 1st of January.[33] We retain information on an individual's earliest migration to Denmark. Following Damm and Dustmann (2014), our sample of refugees is all individuals migrating from Lebanon, Iran, Iraq, Somalia, Sri Lanka, Vietnam, Afghanistan, and Ethiopia. The unit of observation is a person-year.

FAMILIE data. Information on an individual's family is gathered every year on January 1st.[34] Family

[28] The information on years of education is an exception; for refugees who have not studied in Denmark, the information on years of education stems from a survey by Statistics Denmark (Damm (2009)).

[29] Annual hours are estimated using information on mandatory pension contributions (ATP). Lund and Vejlin (2015) improves upon Statistics Denmark' estimated annual hours for years 1980-2007 primarily by using additional information on time spent in sickness, leave, etc.

[30] Approximately 3-5% of workers are working at a fictitious workplace.

[31] Statistics Denmark impute VAT for firms who settle VAT either quarterly or yearly.

[32] As all firms settle VAT at least on a yearly frequency then aggregation of MOMM data is not affected by imputation of monthly and/or quarterly VAT.

[33] Recall that the population of IDA-P is all individuals living in Denmark by December 31st. thus in order to merge IDA-P and EPERSONER, we match a IDA-P person-year+1 observation with a matching EPERSONER person-year observation.

[34] As for observations in EPERSONER we have to merge family data onto our other data sources by matching a person-year+1 observation with a person-year observation from FAMILIE.

information is obtained from Danish social-security register (CPR). We retain information on spouse person, which enables us to link individuals in our datasets together into families. Furthermore, we retain number of children, age of oldest child, age of youngest child, and marriage status. The unit of observation is a person-year.

INDH data. Data on total yearly labor market earnings is obtained from Danish tax authorities by Statistics Denmark.

C Online Technical Appendix

The Online Technical Appendix contains full details on the construction and estimation of the structural model of earnings dynamics, and is available on request from the authors.

D Comparability of the Danish Wage Data

In order to highlight the comparability of the Danish administrative data to that used from other countries in other studies, we repeat the fixed effects regressions of De La Roca and Puga (2017) Table 1 here, for the full Danish population between the years 2004 and 2009, as in their sample. Table 16 shows the results for earnings. We find a very similar size elasticity for wage residuals in Column (2), however including person fixed effects does not reduce this nearly as much as in De La Roca and Puga (2017).

Table 16: Replication of De La Roca and Puga (2017) Table 1

	(1) log(earnings)	(2) Wage Residuals	(3) log(earnings)	(4) Wage Residuals
Experience	0.127***		0.0863***	
	(0.000346)		(0.00376)	
Experience Sqd.	-0.00322***		-0.00331***	
	(0.0000136)		(0.0000287)	
Tenure	0.167***		0.173***	
	(0.000305)		(0.000468)	
Tenure Sqd.	-0.00898***		-0.0105***	
	(0.0000206)		(0.0000388)	
V. High Skill Occ.	0.486***		0.0540***	
	(0.00272)		(0.00421)	
High Skill Occ.	0.269***		0.115***	
	(0.00232)		(0.00383)	
Medium Skill Occ.	0.186***		0.0224***	
	(0.00180)		(0.00269)	
Medium Low Skill Occ.	0.0295***		-0.0623***	
	(0.00162)		(0.00207)	
University Educ.	0.190***		0.527***	
	(0.00190)		(0.00758)	
Secondary Educ.	0.0304***		0.0369***	
	(0.00122)		(0.00376)	
Log Area Population.		0.0539***		0.0459***
		(0.00794)		(0.0106)
Constant		-0.625***		-0.523***
		(0.0847)		(0.113)
Observations	2598703	125	2598703	125
R^2	0.382	0.273	0.761	0.132
Worker Fixed Effects	No		Yes	

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In Table 17 we replicate Table 2 for both wages and earnings. We find very similar results for earnings, including the premium for experience in the two biggest cities (which for us is Aarhus and Copenhagen). We find that studying hourly wages, which De La Roca and Puga do not have measures of, attenuates somewhat the value of big-city experience, suggesting at least part of the effect may be coming from permanently increased hours after working in the big city.

Table 17: Replication of De La Roca and Puga (2017) Table 2

	(1) log(earnings)	(2) log(wage)
Exp. 2 Biggest	0.0319***	0.00841***
	(0.00180)	(0.000947)
Exp. 2. Biggest × Exp.	-0.000587***	0.000121**
	(0.0000733)	(0.0000382)
Exp. 3-5 Biggest	0.0251	0.0100
	(0.0129)	(0.00667)
Exp. 3-5 Biggest × Exp.	-0.000699	-0.000148
	(0.000378)	(0.000192)
Exp.	0.0727***	0.0466***
	(0.00379)	(0.00198)
Exp. Sqd.	-0.00286***	-0.00165***
	(0.0000348)	(0.0000183)
Exp. 2 Biggest × Now in 5 Biggest	0.00876***	0.000594
	(0.00108)	(0.000551)
Exp. 2 Biggest × Now in 5 Bigges × Exp.	-0.000608***	-0.000112***
	(0.0000576)	(0.0000298)
Exp. 3-5 Biggest × Now in 5 Biggest	0.0107***	0.00816***
	(0.00300)	(0.00147)
Exp. 3-5 Biggest × Now in 5 Biggest × Exp.	-0.000814***	-0.000533***
	(0.000170)	(0.0000827)
Exp. Out Top 5. × Now in 5 Biggest	0.00803***	-0.00295*
	(0.00219)	(0.00117)
Exp. Out Top 5. × Now in 5 Biggest × Exp.	-0.000162	0.000370***
	(0.000138)	(0.0000729)
Tenure	0.173***	0.000986***
	(0.000468)	(0.000233)
Tenure Sqd.	-0.0105***	-0.0000442*
	(0.0000389)	(0.0000176)
V. High Skill Occ.	0.0524***	0.0375***
	(0.00421)	(0.00226)
High Skill Occ.	0.112***	0.0333***
	(0.00382)	(0.00209)
Medium Skill Occ.	0.0204***	-0.0131***
	(0.00269)	(0.00144)
Medium Low Skill Occ.	-0.0633***	-0.0338***
	(0.00207)	(0.00112)
University Educ.	0.526***	0.542***
	(0.00758)	(0.00434)
Secondary Educ.	0.0441***	0.326***
	(0.00377)	(0.00220)
Observations	2598703	2589002
R^2	0.761	0.760

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$