Occupational sorting and wage gaps of refugees

Christopher F Baum (Boston College and DIW Berlin)
Hans Lööf (Royal Institute of Technology, Stockholm)
Andreas Stephan (Jönköping University and DIW Berlin)
Klaus Zimmermann (UNU-MERIT Maastricht University, CEPR, GLO)

Portuguese Stata Conference, January 2020

Motivation for the study

- In recent years, the world's attention has been drawn to the plight of refugees and economic migrants.
- A major concern for the receiving country is the ability of these individuals to become well-integrated in their new society and become productive members of the workforce.
- We analyze wage income differences between refugee immigrants in Sweden and native workers within occupations using a matched sample of longitudinal data drawn from the entire Swedish labor force over more than a decade.

Motivation for the study

- In recent years, the world's attention has been drawn to the plight of refugees and economic migrants.
- A major concern for the receiving country is the ability of these individuals to become well-integrated in their new society and become productive members of the workforce.
- We analyze wage income differences between refugee immigrants in Sweden and native workers within occupations using a matched sample of longitudinal data drawn from the entire Swedish labor force over more than a decade.

Motivation for the study

- In recent years, the world's attention has been drawn to the plight of refugees and economic migrants.
- A major concern for the receiving country is the ability of these individuals to become well-integrated in their new society and become productive members of the workforce.
- We analyze wage income differences between refugee immigrants in Sweden and native workers within occupations using a matched sample of longitudinal data drawn from the entire Swedish labor force over more than a decade.

Review of literature

- Researchers have found that refugee immigrants (RIs) tend to concentrate in low paid jobs (Colic-Peisker and Tilbury, *J.Ref.Stud.*, 2006)
- RIs tend to earn lower wages than observationally equivalent natives (Dustman, Glitz, Vogel, *EER*, 2010; Dustman, Frattini, Preston, *REStud*, 2012; Llull, *JHR*, 2017)
- These differences tend to abate over time (Connor, *J.Ref.Stud.*, 2010)
- RIs often face discrimination in seeking better-paying jobs in the labor market (Grand and Szulkin, *Labour*, 2002)
- RIs, compared to other immigrants, usually have a worse starting point but better development in the longer term (Chin and Cortes, *Handbook of Migration*, 2015)

Review of literature

- Researchers have found that refugee immigrants (RIs) tend to concentrate in low paid jobs (Colic-Peisker and Tilbury, *J.Ref.Stud.*, 2006)
- RIs tend to earn lower wages than observationally equivalent natives (Dustman, Glitz, Vogel, *EER*, 2010; Dustman, Frattini, Preston, *REStud*, 2012; Llull, *JHR*, 2017)
- These differences tend to abate over time (Connor, *J.Ref.Stud.*, 2010)
- RIs often face discrimination in seeking better-paying jobs in the labor market (Grand and Szulkin, *Labour*, 2002)
- RIs, compared to other immigrants, usually have a worse starting point but better development in the longer term (Chin and Cortes, *Handbook of Migration*, 2015)

Review of literature

- Researchers have found that refugee immigrants (RIs) tend to concentrate in low paid jobs (Colic-Peisker and Tilbury, *J.Ref.Stud.*, 2006)
- RIs tend to earn lower wages than observationally equivalent natives (Dustman, Glitz, Vogel, *EER*, 2010; Dustman, Frattini, Preston, *REStud*, 2012; Llull, *JHR*, 2017)
- These differences tend to abate over time (Connor, *J.Ref.Stud.*, 2010)
- RIs often face discrimination in seeking better-paying jobs in the labor market (Grand and Szulkin, *Labour*, 2002)
- RIs, compared to other immigrants, usually have a worse starting point but better development in the longer term (Chin and Cortes, *Handbook of Migration*, 2015)

Review of literature

- Researchers have found that refugee immigrants (RIs) tend to concentrate in low paid jobs (Colic-Peisker and Tilbury, *J.Ref.Stud.*, 2006)
- RIs tend to earn lower wages than observationally equivalent natives (Dustman, Glitz, Vogel, *EER*, 2010; Dustman, Frattini, Preston, *REStud*, 2012; Llull, *JHR*, 2017)
- These differences tend to abate over time (Connor, *J.Ref.Stud.*, 2010)
- RIs often face discrimination in seeking better-paying jobs in the labor market (Grand and Szulkin, *Labour*, 2002)
- RIs, compared to other immigrants, usually have a worse starting point but better development in the longer term (Chin and Cortes, *Handbook of Migration*, 2015)

Review of literature

- Researchers have found that refugee immigrants (RIs) tend to concentrate in low paid jobs (Colic-Peisker and Tilbury, *J.Ref.Stud.*, 2006)
- RIs tend to earn lower wages than observationally equivalent natives (Dustman, Glitz, Vogel, *EER*, 2010; Dustman, Frattini, Preston, *REStud*, 2012; Llull, *JHR*, 2017)
- These differences tend to abate over time (Connor, *J.Ref.Stud.*, 2010)
- RIs often face discrimination in seeking better-paying jobs in the labor market (Grand and Szulkin, *Labour*, 2002)
- RIs, compared to other immigrants, usually have a worse starting point but better development in the longer term (Chin and Cortes, *Handbook of Migration*, 2015)

- We seek to shed light on refugees' wage performance by analyzing the impact of occupational sorting on the wage gap between RIs and native workers.
- We adopt the occupational classification scheme of skill-biased technical change of Autor, Levy, Murnane (*QJE*, 2003) and Acemoglu and Autor (*Handbook of Labor Economics*, 2011) to compare wages of RIs and matched native workers.
- These authors have noted an increasing wage gap between routine and non-routine tasks, and between cognitive and manual work.
- When combined with low occupational mobility, occupational sorting could have significant economic consequences for RIs' labor market integration.

- We seek to shed light on refugees' wage performance by analyzing the impact of occupational sorting on the wage gap between RIs and native workers.
- We adopt the occupational classification scheme of skill-biased technical change of Autor, Levy, Murnane (*QJE*, 2003) and Acemoglu and Autor (*Handbook of Labor Economics*, 2011) to compare wages of RIs and matched native workers.
- These authors have noted an increasing wage gap between routine and non-routine tasks, and between cognitive and manual work.
- When combined with low occupational mobility, occupational sorting could have significant economic consequences for RIs' labor market integration.

- We seek to shed light on refugees' wage performance by analyzing the impact of occupational sorting on the wage gap between RIs and native workers.
- We adopt the occupational classification scheme of skill-biased technical change of Autor, Levy, Murnane (*QJE*, 2003) and Acemoglu and Autor (*Handbook of Labor Economics*, 2011) to compare wages of RIs and matched native workers.
- These authors have noted an increasing wage gap between routine and non-routine tasks, and between cognitive and manual work.
- When combined with low occupational mobility, occupational sorting could have significant economic consequences for RIs' labor market integration.

- We seek to shed light on refugees' wage performance by analyzing the impact of occupational sorting on the wage gap between RIs and native workers.
- We adopt the occupational classification scheme of skill-biased technical change of Autor, Levy, Murnane (*QJE*, 2003) and Acemoglu and Autor (*Handbook of Labor Economics*, 2011) to compare wages of RIs and matched native workers.
- These authors have noted an increasing wage gap between routine and non-routine tasks, and between cognitive and manual work.
- When combined with low occupational mobility, occupational sorting could have significant economic consequences for RIs' labor market integration.

Description of the data

The data are provided by Statistics Sweden and contain extensive information on all individuals born between 1954–1980, as well as the firms where they are employed.

We distinguish between three refugee groups:

1. Those from European countries, arriving 1990–1996
2. Those from non-European countries during those years
3. Those arriving 1980–1989 from any country

We observe the labor market outcomes of these three refugee cohorts over the period 2003–2013.

Description of the data

The data are provided by Statistics Sweden and contain extensive information on all individuals born between 1954–1980, as well as the firms where they are employed.

We distinguish between three refugee groups:

1. Those from European countries, arriving 1990–1996
2. Those from non-European countries during those years
3. Those arriving 1980–1989 from any country

We observe the labor market outcomes of these three refugee cohorts over the period 2003–2013.

Description of the data

The data are provided by Statistics Sweden and contain extensive information on all individuals born between 1954–1980, as well as the firms where they are employed.

We distinguish between three refugee groups:

1. Those from European countries, arriving 1990–1996
2. Those from non-European countries during those years
3. Those arriving 1980–1989 from any country

We observe the labor market outcomes of these three refugee cohorts over the period 2003–2013.

Description of the data

The data are provided by Statistics Sweden and contain extensive information on all individuals born between 1954–1980, as well as the firms where they are employed.

We distinguish between three refugee groups:

1. Those from European countries, arriving 1990–1996
2. Those from non-European countries during those years
3. Those arriving 1980–1989 from any country

We observe the labor market outcomes of these three refugee cohorts over the period 2003–2013.

Description of the data

The data are provided by Statistics Sweden and contain extensive information on all individuals born between 1954–1980, as well as the firms where they are employed.

We distinguish between three refugee groups:

1. Those from European countries, arriving 1990–1996
2. Those from non-European countries during those years
3. Those arriving 1980–1989 from any country

We observe the labor market outcomes of these three refugee cohorts over the period 2003–2013.

Description of the data

The data are provided by Statistics Sweden and contain extensive information on all individuals born between 1954–1980, as well as the firms where they are employed.

We distinguish between three refugee groups:

1. Those from European countries, arriving 1990–1996
2. Those from non-European countries during those years
3. Those arriving 1980–1989 from any country

We observe the labor market outcomes of these three refugee cohorts over the period 2003–2013.

We create two cohorts of native workers. The first cohort contains randomly selected native-born workers of the same age, with the sample size chosen to provide a manageable subsample. Non-refugee immigrants are excluded from this sample. This group serves as the baseline sample in our analysis.

In order to compare the wage earnings of refugee workers with those of native-born workers, we define a second cohort of native-born workers who are matched with refugees on a set of similar characteristics.

We create two cohorts of native workers. The first cohort contains randomly selected native-born workers of the same age, with the sample size chosen to provide a manageable subsample. Non-refugee immigrants are excluded from this sample. This group serves as the baseline sample in our analysis.

In order to compare the wage earnings of refugee workers with those of native-born workers, we define a second cohort of native-born workers who are matched with refugees on a set of similar characteristics.

As a methodological innovation, we have altered our original empirical strategy in which this matched cohort was produced by propensity score matching (PSM). Instead, this cohort is formed by applying the Coarsened Exact Matching (CEM) method (Blackwell, Iacus, King, Porro, *Stata Journal* 2009) as an alternative, taking account of the vigorous critiques in King and Nielsen (*Political Analysis*, 2019) of the PSM technique.

In the CEM method, categorical variables are matched exactly for native-born and all cohorts of refugee workers. Continuous variables are coarsened, or binned, into a categorical form. Matches are defined by individuals occupying the same bin.

As a methodological innovation, we have altered our original empirical strategy in which this matched cohort was produced by propensity score matching (PSM). Instead, this cohort is formed by applying the Coarsened Exact Matching (CEM) method (Blackwell, Iacus, King, Porro, *Stata Journal* 2009) as an alternative, taking account of the vigorous critiques in King and Nielsen (*Political Analysis*, 2019) of the PSM technique.

In the CEM method, categorical variables are matched exactly for native-born and all cohorts of refugee workers. Continuous variables are coarsened, or binned, into a categorical form. Matches are defined by individuals occupying the same bin.

In our application, we applied CEM to birth year, gender, marital status, 6 categories of education, children 0-3 years old, children 4-6 years old, and 5 regional categories. Only birth year was coarsened in the process.

After matching, the following empirical analysis is executed on the original data. The coarsened or binned data are only used for the purpose of matching.

In our application, we applied CEM to birth year, gender, marital status, 6 categories of education, children 0-3 years old, children 4-6 years old, and 5 regional categories. Only birth year was coarsened in the process.

After matching, the following empirical analysis is executed on the original data. The coarsened or binned data are only used for the purpose of matching.

The CEM method creates strata for each combination of the ‘coarsened’ variables. In our application, 20,114 strata were defined, and 9,644 of those strata contained individuals in both the native-born and refugee categories. In summary:

Table: Coarsened Exact Matching (CEM) for native and refugee individuals

Refugee	0	1
All	2,603,815	101,453
Matched	99,882	99,882
Unmatched	2,503,933	330

The sample size of roughly 100,000 individuals was then used to generate the first cohort of randomly selected native-born workers of similar size.

Work histories for these five groups of workers produced panel data with the following characteristics:

Table: CEM: Employment, labor market establishment, 2003–2013

		natives	matched natives	European refugees	non-European refugees	pre-1990 refugees
employed	%	84.5	84.3	71.7	59.7	65.0
established	% of emp	88.8	88.2	87.0	74.9	79.4
worker-years	obs	1,079,632	1,079,622	392,528	333,044	320,474

A person is defined as being established on the labor market if monthly wage earnings $\geq$ 0.6 monthly median wage earnings, differentiated by gender, conditional on being employed.

Following Acemoglu and Autor (*Handbook of Labor Economics*, 2011), we classify a worker's occupational task category as:

1. cognitive, non-routine tasks: professionals, managers, technicians
2. cognitive, routine tasks: office and admin support, sales
3. manual, non-routine: personal care, food service, cleaning
4. manual, routine (production, craft, repair, laborers

based on the standardized definitions of occupational tasks.

Following Acemoglu and Autor (*Handbook of Labor Economics*, 2011), we classify a worker's occupational task category as:

1. cognitive, non-routine tasks: professionals, managers, technicians
2. cognitive, routine tasks: office and admin support, sales
3. manual, non-routine: personal care, food service, cleaning
4. manual, routine (production, craft, repair, laborers

based on the standardized definitions of occupational tasks.

Following Acemoglu and Autor (*Handbook of Labor Economics*, 2011), we classify a worker's occupational task category as:

1. cognitive, non-routine tasks: professionals, managers, technicians
2. cognitive, routine tasks: office and admin support, sales
3. manual, non-routine: personal care, food service, cleaning
4. manual, routine (production, craft, repair, laborers

based on the standardized definitions of occupational tasks.

Following Acemoglu and Autor (*Handbook of Labor Economics*, 2011), we classify a worker's occupational task category as:

1. cognitive, non-routine tasks: professionals, managers, technicians
2. cognitive, routine tasks: office and admin support, sales
3. manual, non-routine: personal care, food service, cleaning
4. manual, routine (production, craft, repair, laborers

based on the standardized definitions of occupational tasks.

Following Acemoglu and Autor (*Handbook of Labor Economics*, 2011), we classify a worker's occupational task category as:

1. cognitive, non-routine tasks: professionals, managers, technicians
2. cognitive, routine tasks: office and admin support, sales
3. manual, non-routine: personal care, food service, cleaning
4. manual, routine (production, craft, repair, laborers

based on the standardized definitions of occupational tasks.

Following Acemoglu and Autor (*Handbook of Labor Economics*, 2011), we classify a worker's occupational task category as:

1. cognitive, non-routine tasks: professionals, managers, technicians
2. cognitive, routine tasks: office and admin support, sales
3. manual, non-routine: personal care, food service, cleaning
4. manual, routine (production, craft, repair, laborers

based on the standardized definitions of occupational tasks.

Table: Occupational task classifications

Work tasks	ISCO-88/SSYK 96
Cognitive non-routine	
Professionals	21-24
Managers	12-13
Technicians and Associate professionals	31-34
Cognitive routine	
Office and Administrative Support	41
Sales	42-52
Manual non-routine	
Personal Care, Personal Service, Protective Service	51
Food, Cleaning Service	91
Manual routine	
Production, Craft and Repair	71-74
Operators, Fabricators and Laborers	81-83, 93

Descriptive measures

Over the period 2003–2013, 84% of matched natives were employed, compared to 72% of European refugees, 60% of non-European refugees and 65% of pre-1990 refugees.

The following table illustrates the distribution of workers from each of the five cohorts across the four categories of occupational tasks.

Descriptive measures

Over the period 2003–2013, 84% of matched natives were employed, compared to 72% of European refugees, 60% of non-European refugees and 65% of pre-1990 refugees.

The following table illustrates the distribution of workers from each of the five cohorts across the four categories of occupational tasks.

Table: Share of workers from population group j in occupational task category k, 2003–2013

	natives	matched natives	European refugees	non-European refugees	pre-1990 refugees
cognitive non-routine	0.519	0.487	0.201	0.269	0.344
cognitive routine	0.121	0.124	0.091	0.087	0.085
manual non-routine	0.151	0.151	0.287	0.378	0.324
manual routine	0.209	0.238	0.420	0.267	0.247
worker-years	753,561	735,772	238,621	138,942	153,932

The table is based on employed persons established on the labor market.

About 49% of native workers are employed with cognitive, non-routine tasks, and about 24% working with manual, routine tasks. In contrast, 42% of European refugees work with manual, routine tasks.

Most non-European refugees (38%) work with manual, non-routine tasks, compared to 15% of matched natives. Interestingly, non-European refugees are more likely to work with cognitive, non-routine tasks (27%) than their European counterparts (20%).

About 49% of native workers are employed with cognitive, non-routine tasks, and about 24% working with manual, routine tasks. In contrast, 42% of European refugees work with manual, routine tasks.

Most non-European refugees (38%) work with manual, non-routine tasks, compared to 15% of matched natives. Interestingly, non-European refugees are more likely to work with cognitive, non-routine tasks (27%) than their European counterparts (20%).

Table: Means for population groups in full sample, 2003–2013

	natives	matched natives	European refugees	non-European refugees	pre-1990 refugees
experience	13.721	14.220	9.605	9.290	11.098
female	0.475	0.385	0.475	0.371	0.395
age	41.575	42.961	41.937	42.224	43.632
married	0.491	0.328	0.270	0.237	0.255
kids age 0-3	0.152	0.124	0.131	0.184	0.139
kids age 4-6	0.146	0.135	0.130	0.180	0.135
educ primary	0.081	0.169	0.123	0.167	0.155
educ secondary	0.495	0.490	0.578	0.416	0.464
educ tertiary	0.193	0.189	0.170	0.207	0.166
educ bachelor	0.114	0.057	0.050	0.069	0.088
educ master	0.105	0.085	0.074	0.128	0.114
educ doctoral	0.012	0.009	0.005	0.011	0.014

These results show that native-born workers have considerably more years of experience than refugee immigrants.

European refugees are more likely to be female than are matched native-born workers.

European refugees are more likely to have a secondary education than the matched native cohort.

We now focus on the cognitive non-routine task category, typically containing the highest paying jobs.

These results show that native-born workers have considerably more years of experience than refugee immigrants.

European refugees are more likely to be female than are matched native-born workers.

European refugees are more likely to have a secondary education than the matched native cohort.

We now focus on the cognitive non-routine task category, typically containing the highest paying jobs.

These results show that native-born workers have considerably more years of experience than refugee immigrants.

European refugees are more likely to be female than are matched native-born workers.

European refugees are more likely to have a secondary education than the matched native cohort.

We now focus on the cognitive non-routine task category, typically containing the highest paying jobs.

These results show that native-born workers have considerably more years of experience than refugee immigrants.

European refugees are more likely to be female than are matched native-born workers.

European refugees are more likely to have a secondary education than the matched native cohort.

We now focus on the cognitive non-routine task category, typically containing the highest paying jobs.

Table: Means for population groups in cognitive non-routine occupational task category, 2003–2013

	natives	matched natives	European refugees	non-European refugees	pre-1990 refugees
experience	13.823	14.573	9.964	9.831	11.384
female	0.497	0.361	0.544	0.395	0.410
age	41.775	43.654	41.117	42.412	42.915
married	0.431	0.287	0.305	0.223	0.255
kids age 0-3	0.181	0.139	0.169	0.194	0.163
kids age 4-6	0.168	0.146	0.142	0.171	0.142
educ primary	0.028	0.071	0.013	0.024	0.026
educ secondary	0.261	0.323	0.181	0.120	0.160
educ tertiary	0.287	0.309	0.267	0.238	0.257
educ bachelor	0.206	0.110	0.210	0.197	0.219
educ master	0.195	0.169	0.308	0.387	0.299
educ doctoral	0.022	0.018	0.021	0.034	0.039

These results show that native-born workers in this task category have considerably more years of experience than refugee immigrants.

Both European and non-European refugees in this task category are more likely to be female than are matched native-born workers.

Non-European refugees are more likely to have preschool children than are those in the other cohorts. European refugees are more likely to hold postgraduate degrees than the matched native cohort.

Additional comparisons in the full paper show that refugees are more likely to work in very large firms, and are overrepresented in high-tech manufacturing and underrepresented in high-tech knowledge intensive services.

These results show that native-born workers in this task category have considerably more years of experience than refugee immigrants.

Both European and non-European refugees in this task category are more likely to be female than are matched native-born workers.

Non-European refugees are more likely to have preschool children than are those in the other cohorts. European refugees are more likely to hold postgraduate degrees than the matched native cohort.

Additional comparisons in the full paper show that refugees are more likely to work in very large firms, and are overrepresented in high-tech manufacturing and underrepresented in high-tech knowledge intensive services.

These results show that native-born workers in this task category have considerably more years of experience than refugee immigrants.

Both European and non-European refugees in this task category are more likely to be female than are matched native-born workers.

Non-European refugees are more likely to have preschool children than are those in the other cohorts. European refugees are more likely to hold postgraduate degrees than the matched native cohort.

Additional comparisons in the full paper show that refugees are more likely to work in very large firms, and are overrepresented in high-tech manufacturing and underrepresented in high-tech knowledge intensive services.

These results show that native-born workers in this task category have considerably more years of experience than refugee immigrants.

Both European and non-European refugees in this task category are more likely to be female than are matched native-born workers.

Non-European refugees are more likely to have preschool children than are those in the other cohorts. European refugees are more likely to hold postgraduate degrees than the matched native cohort.

Additional comparisons in the full paper show that refugees are more likely to work in very large firms, and are overrepresented in high-tech manufacturing and underrepresented in high-tech knowledge intensive services.

This same categorization of tasks can be used to compute normalized wage earnings for each cohort and occupational task category. Individuals' wage earnings are normalized by median wage earnings in the respective year for those individuals established on the labor market.

Table: Normalized wage earnings for population group j in occupational task category k, 2003–2013

	natives	matched natives	European refugees	non-European refugees	pre-1990 refugees
cognitive non-routine	1.443	1.570	1.250	1.342	1.381
cognitive routine	0.991	1.003	0.960	0.982	1.005
manual non-routine	0.874	0.881	0.865	0.923	0.930
manual routine	1.118	1.122	1.059	1.036	1.079
worker-years	753,561	735,772	238,621	138,942	153,932

In the cognitive non-routine task category, matched native workers have wages 57% higher than the median wage in all occupations, while European refugees have 25% higher wages and non-European refugees have 34% higher wages.

The lowest wage earnings are those of workers performing manual non-routine tasks, with their wages ranging from 86%–93% of median wage earnings across occupations. Refugees are much more likely to be employed in that task category relative to native-born workers.

In the cognitive routine category, similar wage earnings are found in all five cohorts.

In the cognitive non-routine task category, matched native workers have wages 57% higher than the median wage in all occupations, while European refugees have 25% higher wages and non-European refugees have 34% higher wages.

The lowest wage earnings are those of workers performing manual non-routine tasks, with their wages ranging from 86%–93% of median wage earnings across occupations. Refugees are much more likely to be employed in that task category relative to native-born workers.

In the cognitive routine category, similar wage earnings are found in all five cohorts.

In the cognitive non-routine task category, matched native workers have wages 57% higher than the median wage in all occupations, while European refugees have 25% higher wages and non-European refugees have 34% higher wages.

The lowest wage earnings are those of workers performing manual non-routine tasks, with their wages ranging from 86%–93% of median wage earnings across occupations. Refugees are much more likely to be employed in that task category relative to native-born workers.

In the cognitive routine category, similar wage earnings are found in all five cohorts.

Empirical approach

In our empirical analysis, we first consider the probability that a person is employed in one of the occupational task categories using a multinomial logit framework.

In the multinomial logit model, one category (manual routine tasks) is specified as the base category. The model estimates coefficients which can then be transformed into average marginal effects (AMEs) of the explanatory variable on the probability of being employed in each category, subject to an adding-up constraint.

Empirical approach

In our empirical analysis, we first consider the probability that a person is employed in one of the occupational task categories using a multinomial logit framework.

In the multinomial logit model, one category (manual routine tasks) is specified as the base category. The model estimates coefficients which can then be transformed into average marginal effects (AMEs) of the explanatory variable on the probability of being employed in each category, subject to an adding-up constraint.

The model computes the probabilities that a worker will be employed in each of the four occupational task categories, taking into account their membership in one of the five cohorts and a number of controls.

The worker-specific characteristics include gender, experience and its square, 6 levels of education, marital status, 4 levels of preschool children, and 6 age categories.

The model computes the probabilities that a worker will be employed in each of the four occupational task categories, taking into account their membership in one of the five cohorts and a number of controls.

The worker-specific characteristics include gender, experience and its square, 6 levels of education, marital status, 4 levels of preschool children, and 6 age categories.

As workers in our sample are linked to firms, we can also include 4 firm size categories and 5 industry categories. These refer to the firm being classified as high-tech, medium high tech, medium low tech, low tech, and knowledge-intensive services (KIS). Finally, 5 regional categories are included, specifying whether the worker is in a metro area, dense close city, rural close city, dense remote or rural remote locale.

Table: Average marginal effects of being employed in occupational task category k, multinomial logit model, relative to random native-born workers

	(1) cogn non-rout	(2) cogn rout	(3) man non-rout	(4) man rout
matched natives	0.001***	-0.003***	-0.002***	-0.006***
	[0.001]	[0.001]	[0.001]	[0.001]
European refugees	-0.168***	-0.026***	0.076***	0.117***
	[0.001]	[0.001]	[0.001]	[0.001]
non-European refugees	-0.195***	-0.029***	0.173***	0.051***
	[0.001]	[0.001]	[0.001]	[0.001]
pre-1990 refugees	-0.125***	-0.033***	0.134***	0.025***
	[0.001]	[0.001]	[0.001]	[0.001]
female	0.014***	0.076***	0.155***	-0.245***
	[0.001]	[0.000]	[0.000]	[0.001]
experience	0.005***	-0.000**	-0.005***	-0.000***
	[0.000]	[0.000]	[0.000]	[0.000]
experience^2	0.000***	-0.000***	-0.000***	-0.000***
	[0.000]	[0.000]	[0.000]	[0.000]

The table presents selected results. All controls included in the model, fit to 1,936,101 worker-years, have significant associations with occupational task categories.

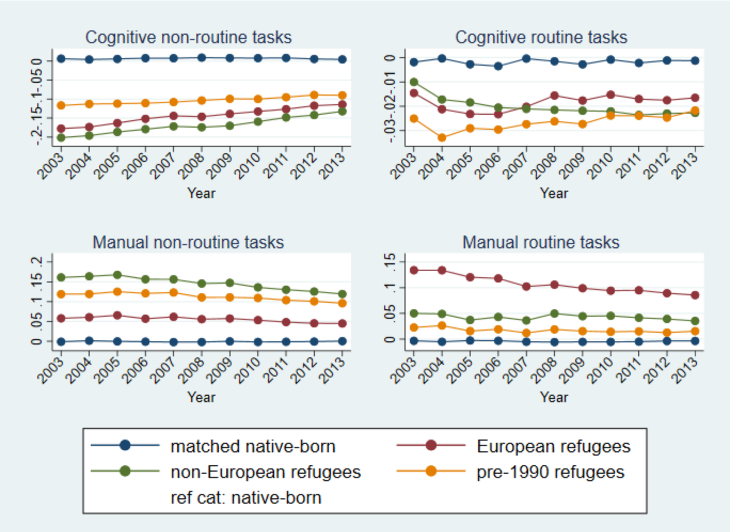
Notably, task category is significantly related to gender, positive for all but the manual routine tasks. Experience and education increase the likelihood of being employed in cognitive, non-routine occupations and decrease the likelihood of manual, routine jobs.

The table presents selected results. All controls included in the model, fit to 1,936,101 worker-years, have significant associations with occupational task categories.

Notably, task category is significantly related to gender, positive for all but the manual routine tasks. Experience and education increase the likelihood of being employed in cognitive, non-routine occupations and decrease the likelihood of manual, routine jobs.

After controlling for observable factors, refugees are less likely to work in cognitive, non-routine and cognitive, routine jobs, and more likely to work in manual, non-routine and routine tasks. Refugees' probability to work in cognitive, non-routine jobs was 15-20% lower in 2003, with most that gap persisting to 2013.

Figure: Marginal effect of cohort on the probability of belonging to occupational category *k*



Modeling wage determination

The second methodological innovation of our study is the use of the correlated random effects approach (Mundlak, *Econometrica*, 1978; Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, 2010) to estimate the determinants of wage earnings over the occupational task categories.

The individual fixed effects model for panel data does not support inclusion of time-invariant regressors such as gender or membership in a cohort. As these variables are key to our study, we seek an alternative approach.

Modeling wage determination

The second methodological innovation of our study is the use of the correlated random effects approach (Mundlak, *Econometrica*, 1978; Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, 2010) to estimate the determinants of wage earnings over the occupational task categories.

The individual fixed effects model for panel data does not support inclusion of time-invariant regressors such as gender or membership in a cohort. As these variables are key to our study, we seek an alternative approach.

The individual random effects model circumvents this restriction, but imposes the very strong assumption that the random effect is uncorrelated with each of the explanatory factors. This assumption is often rejected by a Hausman test vs. the fixed effects estimator.

The CRE approach allows us to consider the impact of time-invariant variables, such as belonging to a certain cohort, without the restrictive assumption of the pure random effects model.

This approach allows the random effects, u_i, to be correlated with the regressors if we model the correlations. As u_i is constant over time, it makes sense for it to be correlated with the time-average of the x_{it}.

The individual random effects model circumvents this restriction, but imposes the very strong assumption that the random effect is uncorrelated with each of the explanatory factors. This assumption is often rejected by a Hausman test vs. the fixed effects estimator.

The CRE approach allows us to consider the impact of time-invariant variables, such as belonging to a certain cohort, without the restrictive assumption of the pure random effects model.

This approach allows the random effects, u_i, to be correlated with the regressors if we model the correlations. As u_i is constant over time, it makes sense for it to be correlated with the time-average of the x_{it}.

The individual random effects model circumvents this restriction, but imposes the very strong assumption that the random effect is uncorrelated with each of the explanatory factors. This assumption is often rejected by a Hausman test vs. the fixed effects estimator.

The CRE approach allows us to consider the impact of time-invariant variables, such as belonging to a certain cohort, without the restrictive assumption of the pure random effects model.

This approach allows the random effects, u_i, to be correlated with the regressors if we model the correlations. As u_i is constant over time, it makes sense for it to be correlated with the time-average of the x_{it}.

The CRE model, as implemented by Schunck (*Stata J*, 2013) and Schunck and Perales (*Stata J*, 2017) can be written as

$$y_{it} = \beta_0 + \beta_w x_{it} + \beta_2 c_i + \pi \bar{x}_i + [\mu_i + \epsilon_{it}] \tag{1}$$

where the c_i variables are time-invariant regressors and the $\bar{x}_i$ variables are time-averages of the regressors for each individual.

The equation still has a composite error term, but by adding the time-averaged values of the regressors, we control for the correlation between u_i and x_{it}.

The β coefficients are identical to the fixed effects (or 'within') coefficients on those regressors, as adding the time-averages is equivalent to centering the regressors via the Frisch–Waugh–Lovell theorem.

The equation still has a composite error term, but by adding the time-averaged values of the regressors, we control for the correlation between u_i and x_{it}.

The β coefficients are identical to the fixed effects (or 'within') coefficients on those regressors, as adding the time-averages is equivalent to centering the regressors via the Frisch–Waugh–Lovell theorem.

The β_w coefficients are then measuring the effects of the regressors on the outcome, controlling for the average level of x_{it} when computing the partial effect. The π coefficients allow us to include time-invariant variables in the model, which cannot be used in a fixed-effects context.

The π coefficients measure the differences between within and between estimates. An additional advantage of the CRE approach is that it provides a formal way of choosing between the pure RE model and the augmented CRE model. If the u_i are indeed uncorrelated with x_{it}, the estimates of π will not be distinguishable from zero.

The β_w coefficients are then measuring the effects of the regressors on the outcome, controlling for the average level of x_{it} when computing the partial effect. The π coefficients allow us to include time-invariant variables in the model, which cannot be used in a fixed-effects context.

The π coefficients measure the differences between within and between estimates. An additional advantage of the CRE approach is that it provides a formal way of choosing between the pure RE model and the augmented CRE model. If the u_i are indeed uncorrelated with x_{it}, the estimates of π will not be distinguishable from zero.

As Schunck has shown, the CRE model is numerically equivalent to a so-called hybrid model from which both within and between estimates can be directly obtained:

$$y_{it} = \beta_0 + \beta_w(x_{it} - \bar{x}_i) + \beta_2 c_i + \beta_b \bar{x}_i + [\mu_i + \epsilon_{it}] \quad (2)$$

The within estimate, β_w, shows the effect of a variable which varies over time on the outcome for an individual, while the between estimate β_b can be interpreted as the long-run effect of that variable.

In our model, y_{it} is normalized monthly wage earnings for person i. The time-invariant regressors include membership in a cohort and gender. The time-varying regressors include occupational task categories, experience, marital status, preschool children, education, age, firm size, industry and region categories, as in the multinomial logit model.

Main findings

We estimate the CRE model first for all occupations, including occupational task category as a time varying control variable, yielding both within and between estimates. Estimation of the full model is based on 1,936,101 worker-year observations. We then estimate the model separately for each task category.

As the dependent variable is normalized relative income, coefficients can be interpreted as percentage deviations from the annual median income level. The coefficients for the time-invariant regressors are presented in the following table.

Main findings

We estimate the CRE model first for all occupations, including occupational task category as a time varying control variable, yielding both within and between estimates. Estimation of the full model is based on 1,936,101 worker-year observations. We then estimate the model separately for each task category.

As the dependent variable is normalized relative income, coefficients can be interpreted as percentage deviations from the annual median income level. The coefficients for the time-invariant regressors are presented in the following table.

Table: Determinants of wage earnings by occupational task category, CRE model, 2003–2013

	all occup	cogn non-rout	cogn rout	man non-rout	man rout
matched	0.014***	0.017***	-0.006	-0.002	-0.003
	[0.003]	[0.006]	[0.004]	[0.002]	[0.002]
Eur ref	0.037***	-0.033***	0.020***	0.043***	0.014***
	[0.003]	[0.008]	[0.006]	[0.003]	[0.003]
non-Eur ref	-0.019***	-0.084***	-0.023**	0.046***	-0.023***
	[0.004]	[0.010]	[0.008]	[0.003]	[0.004]
pre-90 ref	-0.040***	-0.094***	0.003	0.042***	-0.023***
	[0.004]	[0.009]	[0.008]	[0.004]	[0.004]
female	-0.261***	-0.337***	-0.160***	-0.148***	-0.147***
	[0.003]	[0.006]	[0.004]	[0.002]	[0.003]
worker-years	1,936,101	858,410	221,547	368,055	488,089
individuals	241,899	119,962	45,534	64,326	73,396
R^2_w	0.004	0.005	0.012	0.011	0.006
R^2_b	0.236	0.177	0.134	0.133	0.136

We first note that females earn lower wages than men, 26% overall and between 15%–37% over task categories.

European refugees earn slightly more than matched native-born workers over all occupations, but 5% less in the cognitive non-routine category, 2% more in the cognitive routine category, and over 4% more in the manual non-routine category. Non-European refugees earn less in all but the manual non-routine category.

The between R_b^2 values are much higher than the within R_w^2s. The occupational task category, which only enters the 'all occup' model, has considerable explanatory power, as the table below shows. The within effect captures the impact of workers changing task categories.

We first note that females earn lower wages than men, 26% overall and between 15%–37% over task categories.

European refugees earn slightly more than matched native-born workers over all occupations, but 5% less in the cognitive non-routine category, 2% more in the cognitive routine category, and over 4% more in the manual non-routine category. Non-European refugees earn less in all but the manual non-routine category.

The between R_b^2 values are much higher than the within R_w^2s. The occupational task category, which only enters the 'all occup' model, has considerable explanatory power, as the table below shows. The within effect captures the impact of workers changing task categories.

We first note that females earn lower wages than men, 26% overall and between 15%–37% over task categories.

European refugees earn slightly more than matched native-born workers over all occupations, but 5% less in the cognitive non-routine category, 2% more in the cognitive routine category, and over 4% more in the manual non-routine category. Non-European refugees earn less in all but the manual non-routine category.

The between R_b^2 values are much higher than the within R_w^2s. The occupational task category, which only enters the 'all occup' model, has considerable explanatory power, as the table below shows. The within effect captures the impact of workers changing task categories.

Table: Within and between coefficients for occupational task categories, CRE model

	within	between
cogn non-routine	0.042***	0.348***
	[0.002]	[0.004]
cogn routine	-0.008***	0.016***
	[0.003]	[0.004]
manual non-routine	-0.022***	0.034***
	[0.003]	[0.005]

The short-run effect of switching from manual routine tasks to cognitive non-routine tasks is only 4.2%, but the between estimates show that the long-term effect is almost 35%. The other two groups have much smaller positive long-run effects of switching from manual routine tasks, with negative short-run effects.

Table: Within and between coefficients for occupational task categories, CRE model

	within	between
cogn non-routine	0.042***	0.348***
	[0.002]	[0.004]
cogn routine	-0.008***	0.016***
	[0.003]	[0.004]
manual non-routine	-0.022***	0.034***
	[0.003]	[0.005]

The short-run effect of switching from manual routine tasks to cognitive non-routine tasks is only 4.2%, but the between estimates show that the long-term effect is almost 35%. The other two groups have much smaller positive long-run effects of switching from manual routine tasks, with negative short-run effects.

A number of other interesting findings appear in the full CRE model results, available in the paper. The effects of an additional year of experience is highest (over 11%) for cognitive non-routine tasks, but significant for all task categories.

Wage earnings are almost 27% higher for cognitive non-routine tasks in larger metro areas relative to other municipalities or rural locations. The long-term effects of higher education are also largest for those performing cognitive non-routine tasks.

In the following figure, the task category indicators are interacted with year effects to see how these impacts change over time. In the cognitive non-routine category, differences between cohorts are persistent.

A number of other interesting findings appear in the full CRE model results, available in the paper. The effects of an additional year of experience is highest (over 11%) for cognitive non-routine tasks, but significant for all task categories.

Wage earnings are almost 27% higher for cognitive non-routine tasks in larger metro areas relative to other municipalities or rural locations. The long-term effects of higher education are also largest for those performing cognitive non-routine tasks.

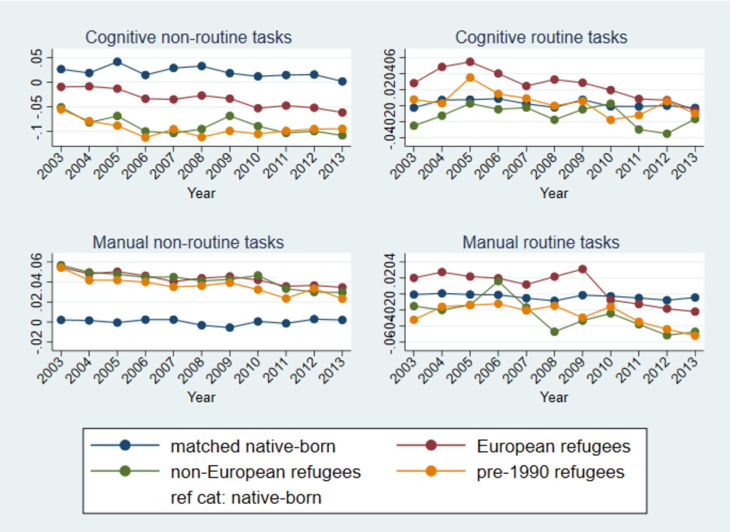
In the following figure, the task category indicators are interacted with year effects to see how these impacts change over time. In the cognitive non-routine category, differences between cohorts are persistent.

A number of other interesting findings appear in the full CRE model results, available in the paper. The effects of an additional year of experience is highest (over 11%) for cognitive non-routine tasks, but significant for all task categories.

Wage earnings are almost 27% higher for cognitive non-routine tasks in larger metro areas relative to other municipalities or rural locations. The long-term effects of higher education are also largest for those performing cognitive non-routine tasks.

In the following figure, the task category indicators are interacted with year effects to see how these impacts change over time. In the cognitive non-routine category, differences between cohorts are persistent.

Figure: Marginal effect of cohort on normalized wage earnings in occupational category k



We perform a Blinder–Oaxaca wage decomposition (Blinder, *J.of Human Resources*, 1973; Oaxaca *International Economic Review*, 1973)to examine whether wage differences can be explained with different characteristics of native and refugee workers, or whether unexplained differences exist which could suggest wage discrimination, or differentials in unobserved ability.

We apply the so-called twofold decomposition, which is defined as (Jann, *Stata Journal*, 2008):

$$R = \underbrace{[E(X_A)E(X_B)]'\beta^*}_{\text{explained part}} + \underbrace{E(X_A)'(\beta_A - \beta^*) + E(X_B)'(\beta^* - \beta_B)}_{\text{unexplained part}} \qquad (3)$$

where R is the difference in wage earnings between the groups and β^* has been estimated for a reference group: in our case for the matched natives. In our model we have $\beta_A = \beta^*$, so the second term disappears. Thus, the first term shows that differences in characteristics (endowments) explain wage differences, while differences in coefficients imply unexplained wage differences.

We perform the Blinder–Oaxaca decomposition for each cohort over 2003–2013 using the CRE model outlined above, using matched natives as the reference group and the respective refugee group as a comparison group.

The decomposition is available for each refugee immigrant group. We focus here on the group of European refugees. Results for the other groups are given in the full paper.

We perform the Blinder–Oaxaca decomposition for each cohort over 2003–2013 using the CRE model outlined above, using matched natives as the reference group and the respective refugee group as a comparison group.

The decomposition is available for each refugee immigrant group. We focus here on the group of European refugees. Results for the other groups are given in the full paper.

Table: Twofold Blinder–Oaxaca wage decomposition for European refugees, 2003–2013

	(1) all occup	(2) cogn non-rout	(3) cogn rout	(4) man non-rout	(5) man rout
matched natives	1.284***	1.547***	1.014***	0.877***	1.117***
	[0.002]	[0.004]	[0.003]	[0.002]	[0.002]
European refugees	1.029***	1.229***	0.958***	0.860***	1.055***
	[0.002]	[0.005]	[0.004]	[0.002]	[0.002]
difference	0.254***	0.318***	0.056***	0.017***	0.062***
	[0.003]	[0.006]	[0.005]	[0.003]	[0.003]
explained	0.291***	0.249***	0.088***	0.065***	0.092***
	[0.004]	[0.008]	[0.005]	[0.003]	[0.003]
unexplained	-0.037***	0.069***	-0.032***	-0.048***	-0.030***
	[0.005]	[0.009]	[0.007]	[0.004]	[0.004]
N natives	705,148	346,149	88,237	106,611	167,151
N Eur refugees	229,243	46,676	21,168	63,497	97,902
Worker-years	934,391	392,825	109,405	167,108	265,053

Notes: Standard errors in brackets. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Reference group matched natives. Wage earnings relative to median wage earnings in respective year.

Over all occupations, there are relatively small unexplained wage differences between refugees and matched natives, with -3.7% for European refugees. An analysis of the contribution of explanatory variables to the explained difference shows that it is mainly due to differences in accumulated work experience of refugees and natives.

Larger unexplained differences are found for cognitive non-routine task categories, where the unexplained difference is 6.9%. This result might be indicative of wage discrimination in the labor market. On the other hand, for the other three task categories, the unexplained differences are generally negative and significant, implying that refugees actually earn higher wages than predicted by the model.

Over all occupations, there are relatively small unexplained wage differences between refugees and matched natives, with -3.7% for European refugees. An analysis of the contribution of explanatory variables to the explained difference shows that it is mainly due to differences in accumulated work experience of refugees and natives.

Larger unexplained differences are found for cognitive non-routine task categories, where the unexplained difference is 6.9%. This result might be indicative of wage discrimination in the labor market. On the other hand, for the other three task categories, the unexplained differences are generally negative and significant, implying that refugees actually earn higher wages than predicted by the model.

Conclusions

- We find that the observed wage gap between refugees and natives is mainly explained by two factors. The first is occupational sorting into different work tasks. The marginal probability of refugee immigrants to work in higher paid cognitive non-routine occupational jobs is significantly lower.
- Mobility across occupational categories is limited for both native-born workers and refugee workers, but is lower for refugee workers.
- Native-born workers have, on average, more accumulated work experience. Holding other factors equal—age, gender, family status, education, place of residence, company size, industry, and job task—refugees have less work experience, which explains a large part of the wage disparity. However, a significant part of the wage gap remains unexplained by the model.

Conclusions

- We find that the observed wage gap between refugees and natives is mainly explained by two factors. The first is occupational sorting into different work tasks. The marginal probability of refugee immigrants to work in higher paid cognitive non-routine occupational jobs is significantly lower.
- Mobility across occupational categories is limited for both native-born workers and refugee workers, but is lower for refugee workers.
- Native-born workers have, on average, more accumulated work experience. Holding other factors equal—age, gender, family status, education, place of residence, company size, industry, and job task—refugees have less work experience, which explains a large part of the wage disparity. However, a significant part of the wage gap remains unexplained by the model.

Conclusions

- We find that the observed wage gap between refugees and natives is mainly explained by two factors. The first is occupational sorting into different work tasks. The marginal probability of refugee immigrants to work in higher paid cognitive non-routine occupational jobs is significantly lower.
- Mobility across occupational categories is limited for both native-born workers and refugee workers, but is lower for refugee workers.
- Native-born workers have, on average, more accumulated work experience. Holding other factors equal—age, gender, family status, education, place of residence, company size, industry, and job task—refugees have less work experience, which explains a large part of the wage disparity. However, a significant part of the wage gap remains unexplained by the model.

Our findings have important policy implications with respect to both income inequality and economic efficiency. Occupational sorting is accompanied by increasing wage differentials for high-skilled and low-skilled workers while occupational mobility is limited. This may counteract the long-run process of narrowing wage gaps due to reduced differences in work experience.

As many companies face difficulties in recruiting competent and qualified personnel, refugee workers may have unexploited skill potentials that could be used to reduce the shortage of skilled labor in Sweden and other developed economies facing the demographic challenges of an increasing ratio of pensioners to workers.

Our findings have important policy implications with respect to both income inequality and economic efficiency. Occupational sorting is accompanied by increasing wage differentials for high-skilled and low-skilled workers while occupational mobility is limited. This may counteract the long-run process of narrowing wage gaps due to reduced differences in work experience.

As many companies face difficulties in recruiting competent and qualified personnel, refugee workers may have unexploited skill potentials that could be used to reduce the shortage of skilled labor in Sweden and other developed economies facing the demographic challenges of an increasing ratio of pensioners to workers.