



Immigrants in their host society: Barriers to immigrants' assimilation

*PSE Summer School – Migration Economics
June 2019*

Biagio Speciale | biagio.speciale@univ-paris1.fr

Paris School of Economics, Université Paris 1 Panthéon-Sorbonne

Outline of the talk

- Barriers to immigrants' assimilation
 - Language barriers
 - Language skills and labor market outcomes
 - Language training and labor market outcomes
 - Undocumented status
 - Undocumented status and labor market outcomes
 - Undocumented status and consumption behavior
 - Undocumented status and propensity to commit crime

Introduction

- Immigrants represent a disadvantaged group in European labor markets

**Table: Unemployment rates in EU countries and France.
2014/2015 data from OECD statistics.**

	Unemployment rates	
	foreign-born	native-born
EU countries	14.90%	9.80%
France	17.40%	9.50%

Language skills and labor market outcomes



- Lack of host country language skills as a main obstacle of immigrants' economic integration
- Several works study the relationship between language skills and labor market outcomes in different immigration countries (Chiswick, 1991; Chiswick and Miller, 1995; Dustmann and van Soest, 2001; Dustmann and Fabbri, 2003; Bleakley and Chin, 2004)
 - Positive and significant effect of host country language skills on immigrants' wages and probability of employment
 - Endogeneity of language skills due to self-reported assessments of language proficiency (measurement error) and omitted variables correlated with language skills and labor market outcomes

Language skills and labor market outcomes

(2)



- Methods to solve the endogeneity of language skills in previous literature :
 - IV estimation with parents' education level as instrumental variable (Dustmann and van Soest, 2001 REStat)
 - Information on whether the interview was done in the host country language only as IV (Dustmann and Fabbri, 2003 EJ)
 - Age at arrival interacted with a dummy for non-English-speaking country as instrumental variable (Bleakley and Chin, 2004 REStat)

Active Labor Market Policies (ALMPs)



- To improve immigrants' labor market integration, governments may use a wide range of **Active Labor Market Programs** (ALMPs)
 - **Language courses**
 - **Job search assistance** (e.g. counseling and monitoring of job search efforts)
 - **Training programs** (e.g. computer courses or courses providing specific occupational knowledge)
 - **Subsidized public and private sector employment** (e.g. wage subsidies for employers who hire disadvantaged workers; temporary job opportunities in community services.).



The Effect of Language Training on Immigrants' Economic Integration: Empirical Evidence from France

A. Lochmann, H. Rapoport and B. Speciale
CESifo Working Paper Series No. 6460

Main obstacles for immigrants' integration in France. Data source: ELIPA



	Freq.	Percentage
Lack of French language skills	1526	45.01
Lack of school support for immigrants' children	257	7.58
Lack of knowledge of usages, norms, cultural aspects	466	13.75
Lack of ethnic diversity in neighborhoods	189	5.58
Lack of policies favoring immigrants' integration	286	8.44
Lack of knowledge of school system, police, city hall	230	6.78
Lack of knowledge of French laws	331	9.76
None of these answers	105	3.10

Motivation

- Language barrier as one of the major obstacles for immigrants' integration into the host country's labor market (e.g. Borjas, 1994).
- Growing need to evaluate the channel of language skills and the effectiveness of integration plans on immigrants' integration.
 - We examine the impact of language training on the economic integration of immigrants in France.
 - The language training is offered by the French Ministry of Interior in the context of the *Contrat d'accueil et d'intégration*.

Contrat d'accueil et d'intégration

- 1st of January 2007: French Government introduced policies aimed at improving the economic and social integration of immigrants
- *Contrat d'accueil et d'intégration* (CAI) ("Welcome and Integration Contract"):
 - A civic training (e.g. on French institutions and values of the Republic)
 - An information session on life in France (about French society and access to public services)
 - A statement of professional competence (useful for employment access projects)
 - A language training (based on the results of an entry exam)

Contrat d'accueil et d'intégration (2)

- 1st of July 2016: French Government modified some aspects of these integration policies (*Contrat d'intégration républicaine*, CIR)
- **Language training:**
 - A total of 25% of all the individuals who signed the contract in 2010 were proposed the language training, between 60 and 400 hours with on average 264 hours (200 hours median).
 - The level of the CAI language training was basic: A1.1 level
 - By 2013 97% of the individuals who completed the training passed the DILF ("*Diplôme Initial de Langue Française*") French language exam.
 - 70% of the individuals who took the CAI training thought that this training was not enough to read and speak French – they would have wished to have additional hours.

Longitudinal Survey of the Integration of First-time arrivals (ELIPA)



- Collected by the *Département des statistiques, des études et de la documentation* (DSED) of the French Ministry of the Interior.
- Migrants from countries outside the EU and Switzerland, are at least 18 years old, were granted their residence permit at the end of 2009 and want to settle permanently in the country.
- Longitudinal survey (2010, 2011, 2013) on the integration of newly arrived immigrants in France.

Longitudinal Survey of the Integration of First-time arrivals (ELIPA) (2)



- Follows the path of the individuals, collecting a large set of information regarding socio-demographic characteristics, bureaucratic itinerary, employment, language skills, living conditions and social integration.
- Socio-demographic characteristics in 2010:

Variable	Mean	Std. Dev.	Min	Max
Educational level	8.46	6.19	0	17
Age	32.58	9.51	18	81
Years since arrival	3.82	4.75	0	56
Ile-de-France	0.59	0.49	0	1
Married	0.82	0.37	0	1
Male	0.47	0.49	0	1
Children	0.67	1.01	0	7
Household size	3.23	1.67	1	12

Descriptive statistics by migration reason



	Economic migration	Family migration	Refugees
<i>Gender</i>			
Men	0.80	0.41	0.58
Women	0.20	0.59	0.42
<i>Age in 2010</i>			
18 to 29	0.22	0.46	0.46
30 to 39	0.50	0.37	0.34
40 to 49	0.21	0.12	0.12
50 or more	0.07	0.04	0.08
<i>Age at arrival</i>			
0 to 17	0.03	0.06	0.13
18 to 29	0.65	0.57	0.48
30 to 39	0.24	0.27	0.24
40 to 49	0.06	0.07	0.10
50 or more	0.01	0.02	0.05
<i>Duration of stay in France in 2010</i>			
Less than 2 years	0.09	0.60	0.30
From 2 to 4	0.10	0.12	0.51
From 5 to 9	0.59	0.22	0.16
10 years or more	0.22	0.07	0.04
<i>Nationality</i>			
Maghreb	0.20	0.46	0.01
Sub-Saharan Africa	0.51	0.21	0.23
Other areas in Africa	0.05	0.06	0.14
Asia	0.12	0.16	0.44
Europe (excluding France) and CIS	0.05	0.05	0.14
America and Oceania	0.07	0.06	0.03

Labor market outcomes in 2013

Variable	Mean	Std. Dev.	Min	Max
Labor force participation	0.77	0.41	0	1
Being employed	0.59	0.49	0	1
Being empl. (cond. on LFP)	0.78	0.41	0	1
Income per HH member	1127.26	751.21	0	20000
Full time job	0.45	0.49	0	1
Full t. job (cond. on empl.)	0.77	0.41	0	1
Permanent contract	0.38	0.48	0	1
Perm. contr. (cond. on empl.)	0.65	0.47	0	7
Informal work	0.02	0.16	0	1
Informal work (cond. on empl.)	0.04	0.21	0	1

Methodology:

Regression Discontinuity design



- **Regression discontinuity design:**
 - One of the most advanced methods for causal inference in economics
 - Assignment of participants into the treated or control group depends on whether a certain variable exceeds a cutoff point
 - Comparison of observations lying closely on either side of the threshold

Methodology: Regression Discontinuity design (2)



- Empirical challenge to assess the effects of language training on immigrants' integration
 - Individuals who are assigned to the language training are likely to be different in terms of unobservables from those who are not assigned to the training
- Solution to the empirical challenge: Regression Discontinuity Design
 - We exploit a discontinuity related to the test score of a language entry exam, which is the main eligibility rule to the training

The Language Entry Test: Fuzzy RDD



- Each newly-arrived immigrant has to pass a test on the knowledge of the French language, written and spoken.
- If the result is insufficient, the person has to receive a linguistic training
 - The assignment to this training mainly depends on whether the entry test results are inferior or equal to a certain threshold (test score < 50, test score ranges from 0 to 100)
- The assignment to the language training also depends on socio-demographic characteristics of the individual: e.g. the country of origin (francophone vs. non francophone), employment status of the migrant at the time of the entry test, number of years already spent in France.

Empirical strategy

Second Stage

$$Y_i = \beta_0 + \beta_1 LC_i + \beta_2 X_i + \varepsilon_i$$

First Stage

$$LC_i = \alpha_0 + \alpha_1 T_i + \alpha_2 X_i + v_i$$

- Y_i is the outcome of interest of individual i (e.g. labor force participation, language skills, networks)
- $LC = 0$ if no language training was assigned, and LC = number of hours prescribed conditional on the fact that language training was assigned.
- T_i is the assignment variable. $T = 1$ if $Score < cutoff$ (50) and $T = 0$ if $Score \geq cutoff$ (50)
- X_i is a vector of control variables

Control variables

- **Education** level in years, **age**, age squared, a dummy variable equal to 1 if the immigrant is resident in the region “**Ile-de-France**”, a dummy variable equal to 1 if the immigrant is **married**, a dummy variable equal to 1 if the immigrant is **male**, the number of **children in the household**, the total number of **individuals in the household**, the number of **years that the migrant has already spent in France**, three dummy variables indicating the **reasons to migrate** (labor migration, refugee status, other), a dummy variable equal to one if the individual was **employed in 2010**, and **country of origin fixed effects**

Empirical strategy: two approaches

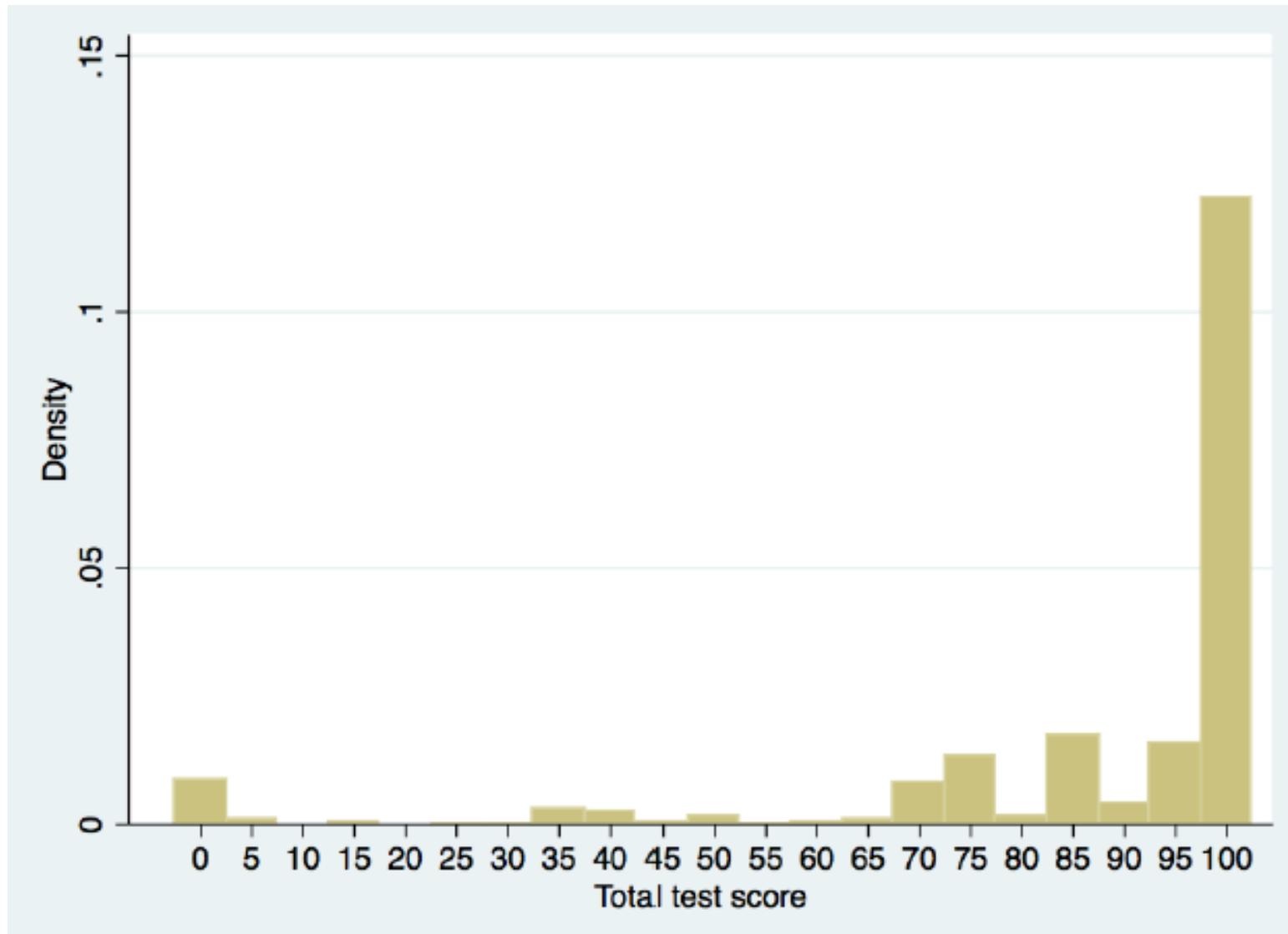
- Approach 1. Estimation sample: individuals around the cutoff (**local linear regressions**)
 - We use the optimal bandwidth selector suggested by Calonico, Cattaneo and Titiunik (2014), which is compatible with fuzzy RD designs.
 - The optimal bandwidth according to this estimator: individuals with test results between 35 and 60.
 - We also report the estimates using different bandwidths: trade-off between bias and estimate precision

Empirical strategy: two approaches (2)

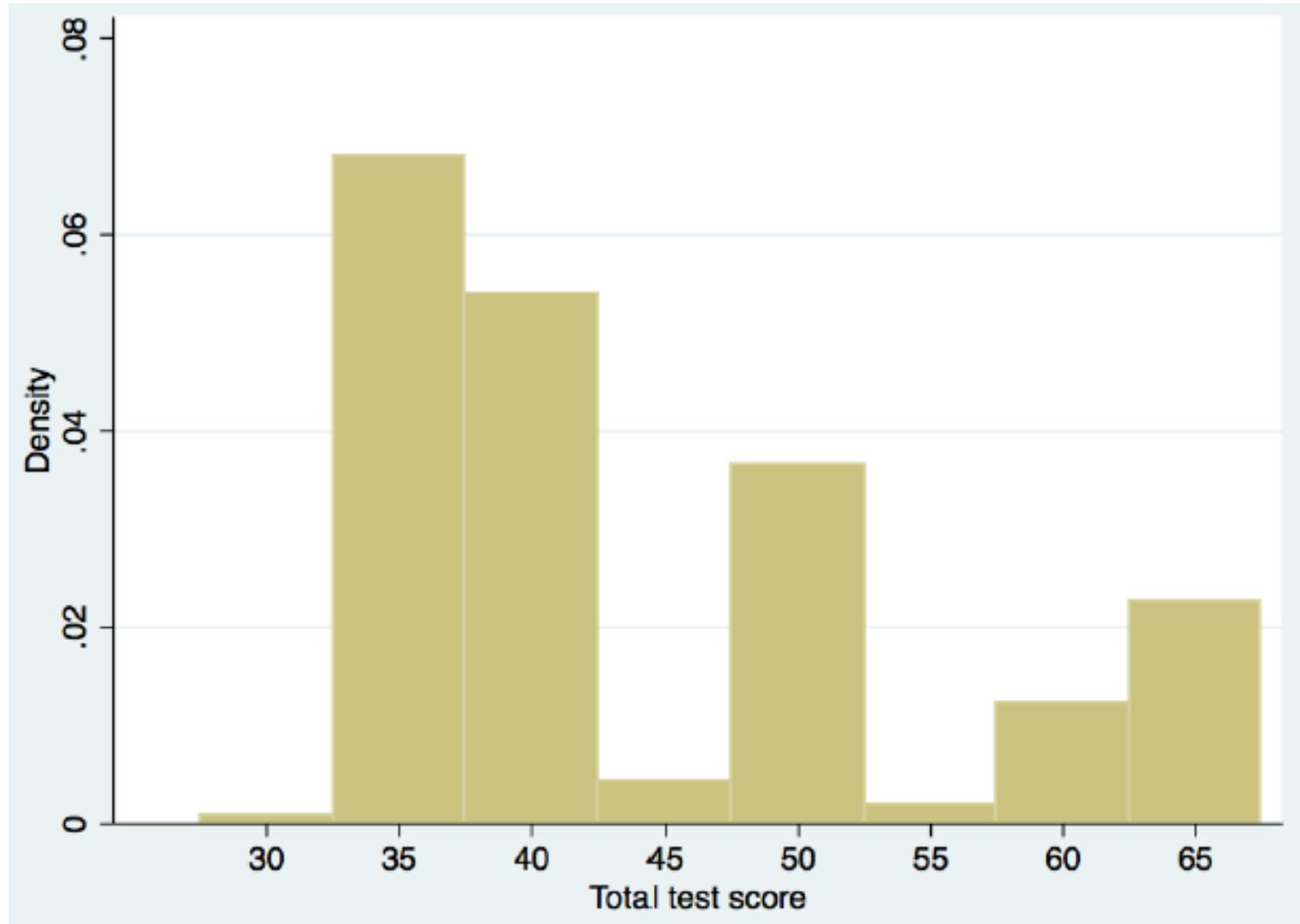


- Approach 2. **Parametric estimates** over the whole sample.
 - We control for the distance to the cutoff point (*Score-45*)
 - Using a polynomial function of the distance, we gradually include in the regressions the distance variable to the power of one and two.

Density of the score of the language entry test



Density of the score of the language entry test around the cutoff



Manipulation around the cutoff

- By the individuals:
 - Unlikely to manipulate the test result close to the cutoff. If individuals want to fail they would fail by a lot in order to make sure to get a high number of assigned hours of language classes.
- By the government:
 - If the government did not want to give training to certain individuals they would simply decide to do so, without manipulating the test result.
 - There is no upper limit to the number of people that can be admitted to the training - the training is provided for every person who needs it.

Test of balance of pre-treatment characteristics



Test of Balance of Pre-Treatment Characteristics, in 2010.

Panel A: Demographic and household characteristics

	Education level	Age	Permanence in France	Ile-de-France	Married	Male	Number of children	Number of people in HH	Employed in 2010	Taking other French classes
I(test score initial language exam < 50)	-0.772 [2.384]	0.787 [4.099]	-2.071 [1.866]	-0.397 [0.243]	-0.048 [0.125]	0.152 [0.211]	0.013 [0.379]	-0.897 [0.593]	-0.205 [0.202]	0.010 [0.136]
Polynomial of degree 2 Observations	2 4,969	2 5,299	2 5,298	2 5,307	2 4,231	2 5,307	2 5,307	2 5,307	2 5,303	2 5,304

Panel B: Types of migration and regions of origin

	Work migration	Refugee migration	Family migration	Europe	Asia	Maghreb	Sub-Sah. Africa	Africa (other)	America & Oceania
I(test score initial language exam < 50)	-0.080 [0.097]	0.120 [0.185]	0.035 [0.217]	0.113 [0.171]	-0.256 [0.271]	0.315 [0.452]	0.019 [0.219]	-0.098 [0.159]	-0.093 [0.118]
Polynomial of degree 2 Observations	2 5,307	2 5,307	2 5,307	2 5,307	2 5,307	2 5,307	2 5,307	2 5,307	2 5,307

Robust standard errors in brackets, clustered by country of origin and test scores of the initial language exam.

The columns include functions in the test scores of the initial language exam of degree 2, interacted with I(test score initial language exam < 50).

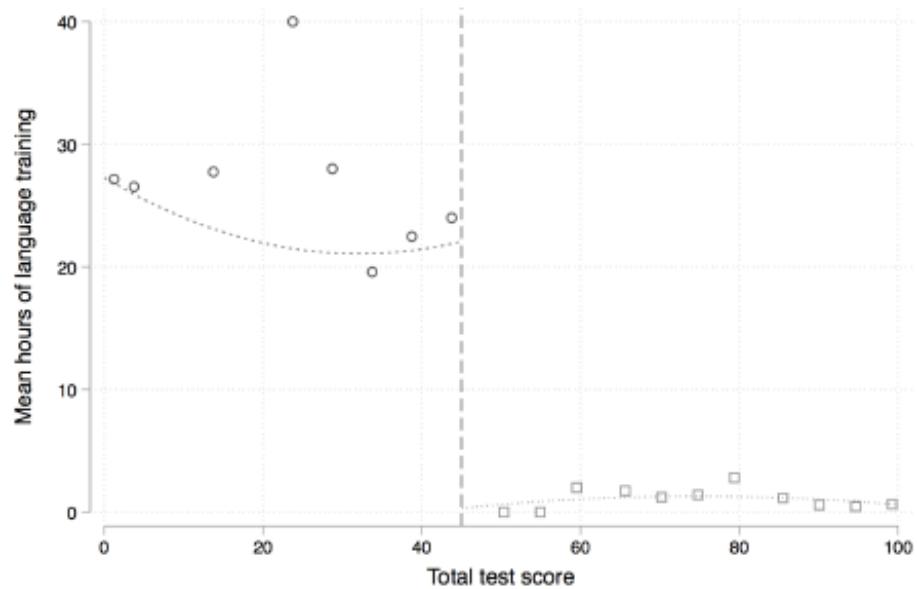
*** p<0.01, ** p<0.05, * p<0.1

Results:

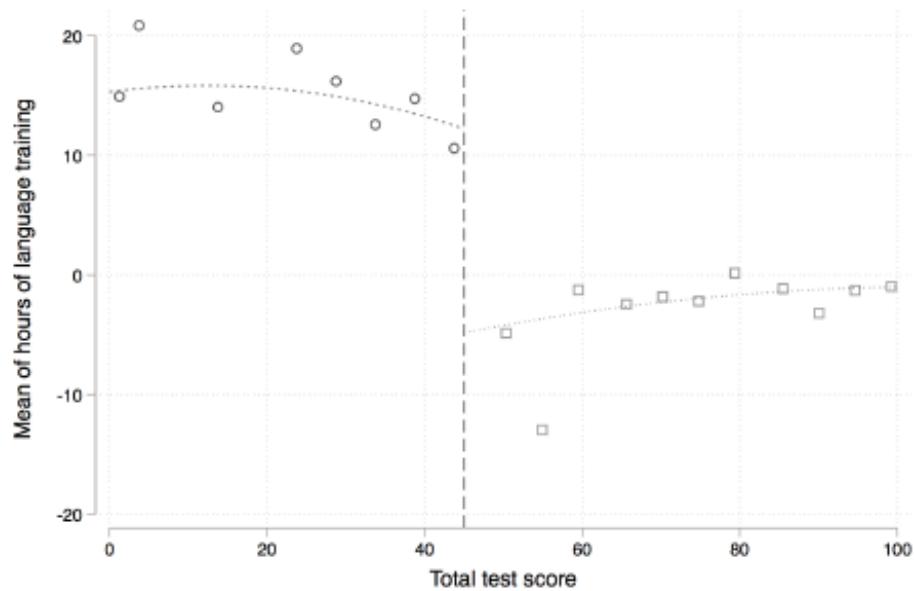
entry test and hours of language training



(a) Raw data



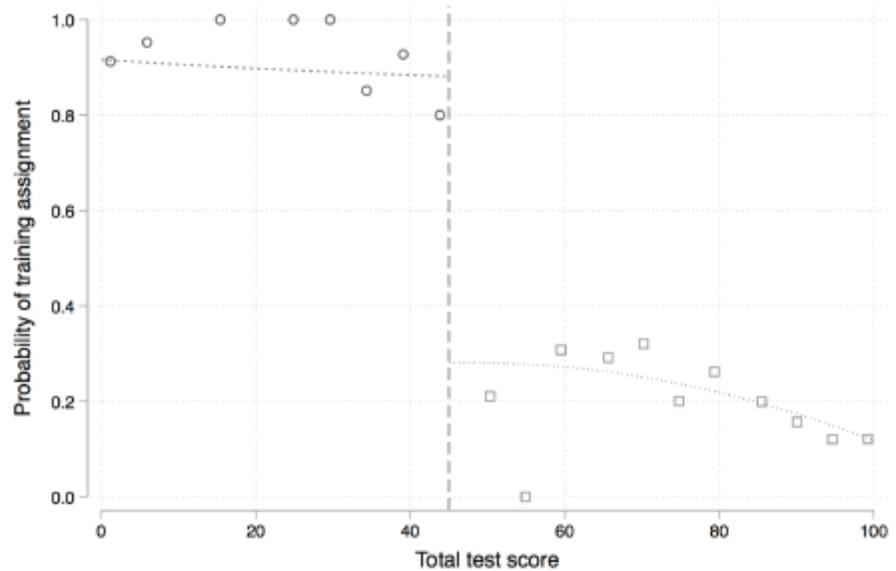
(b) Residuals from a regression on covariates



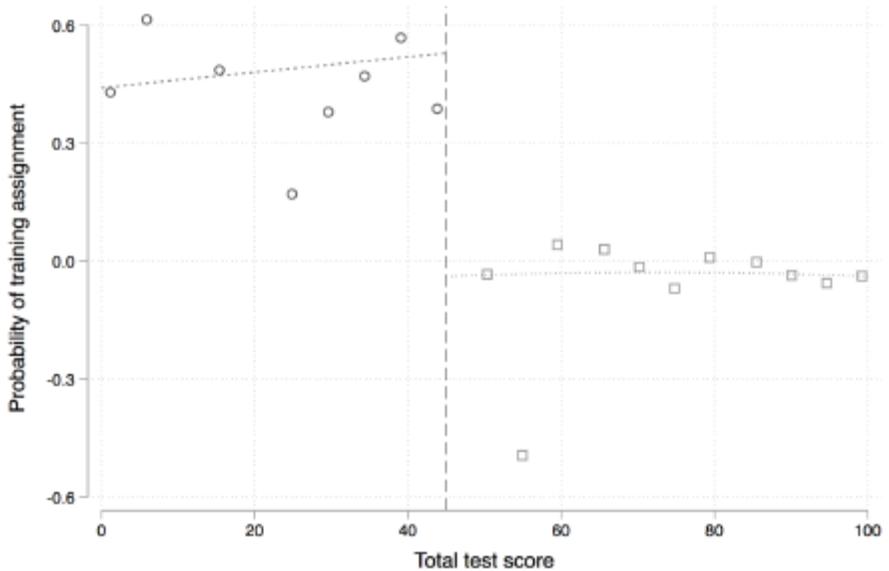
Results: entry test and probability of being assigned training



(c) Raw data



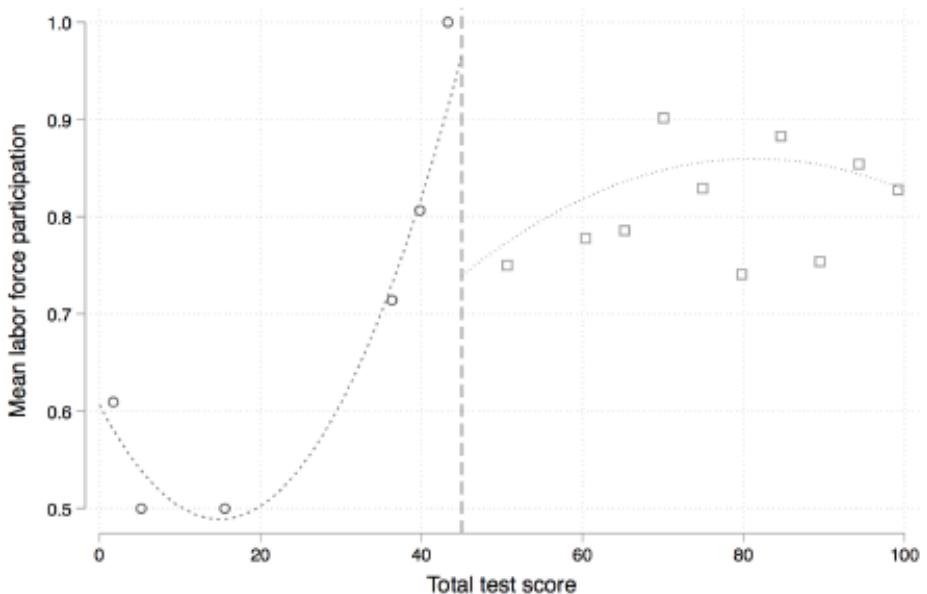
(d) Residuals from a regression on covariates



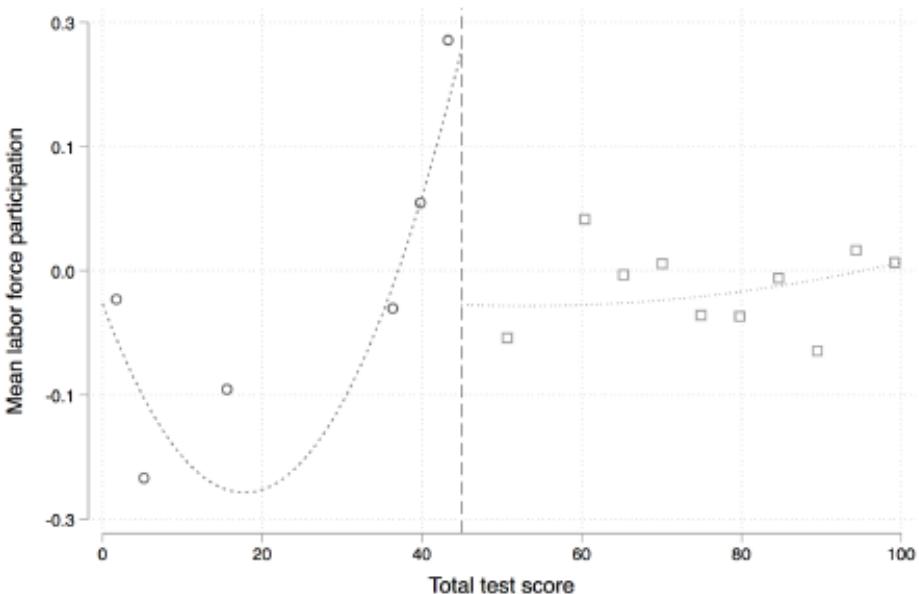
Results: language entry test and labor force participation. Entire sample



(a) Raw data



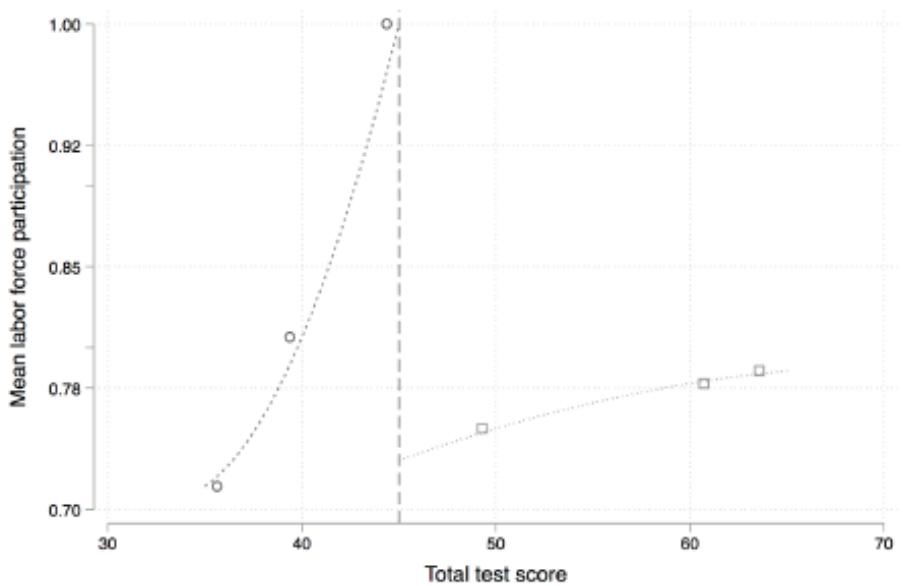
(b) Residuals from a regression on covariates



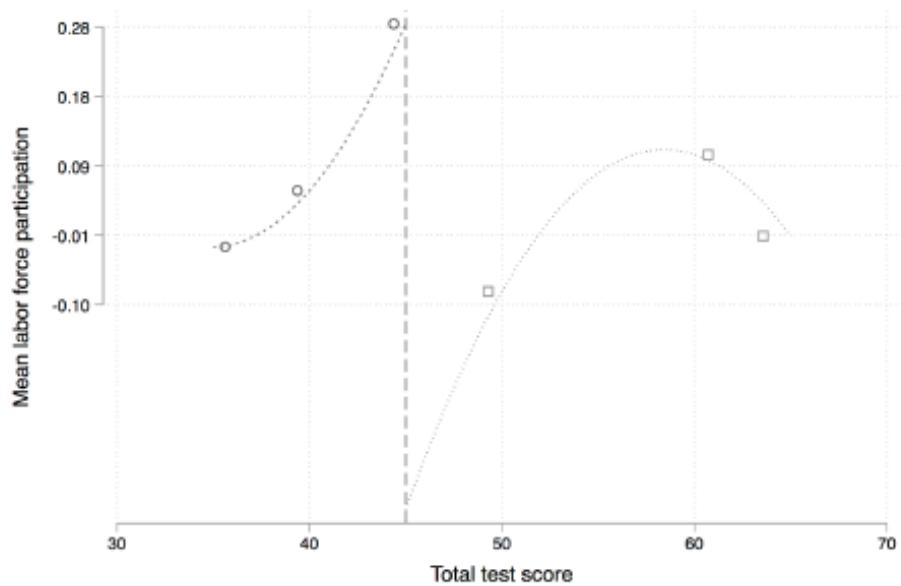
Results: language entry test and labor force participation. Bandwidth: test scores 30-65



(c) Raw data



(d) Residuals from a regression on covariates



Assigned French language training and labor force participation



Table 1: Assigned French Language Training and Labor Force Participation

	Local linear estimates		Parametric estimates		
	[35, 60]	[30, 65]			
Panel A: First stage regressions (dependent variable: number of assigned hours of French language training/100)					
I(test score initial language exam < 50)	3.294*** [0.936]	2.910*** [0.652]	2.576*** [0.466]	3.374*** [0.700]	3.602*** [1.072]
Panel B: Second stage regressions (dependent variable: labor force participation dummy)					
Hours of French language training (/100)	0.266* [0.141]	0.190* [0.106]	0.206** [0.098]	0.173* [0.103]	0.253** [0.106]
Polynomial of degree	1	1	2	3	4
Control variables	Yes	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	67	76	2,168	2,168	2,168

Robust standard errors in brackets, clustered by country of origin and test scores of the initial language exam. Columns 3-5 include functions in the test scores of the initial language exam of degree 1, 2, 3 and 4, interacted with I(test score initial language exam < 50). *** p<0.01, ** p<0.05, * p<0.1

- An increase in the language course duration by 100 hours leads to a 17 to 26 percentage point increase in the probability of participating in the labor force.

Assigned French language training and labor force participation (2)



Table 1a: Assigned French Language Training and Labor Force Participation. Heterogeneity effects by education level.

	Local linear estimates		Parametric estimates		
	[35, 60]	[30, 65]			
Panel A: Second stage regressions (dependent variable: labor force participation dummy)					
Hours of French language training (/100) (a_0)	0.095 [0.072]	0.097 [0.065]	0.162* [0.088]	0.148 [0.100]	0.176** [0.085]
Hours of French classes (/100) X education level (a_1)	0.024* [0.013]	0.024 [0.015]	0.011* [0.006]	0.010 [0.006]	0.012** [0.006]
$H_0: a_0 + 12 \cdot a_1 = 0$ (p-value)	0.0238	0.0631	0.0363	0.0918	0.0147
Polynomial of degree	1	1	2	3	4
Control variables	Yes	Yes	Yes	Yes	Yes
Country of origin fixed effects	Yes	Yes	Yes	Yes	Yes
Observations	67	76	2,168	2,168	2,168

Robust standard errors in brackets, clustered by country of origin and test scores of the initial language exam. The median years of education are 12. Columns 3-5 include functions in the test scores of the initial language exam of degree 1, 2, 3 and 4, interacted with I(test score initial language exam < 50). *** p<0.01, ** p<0.05, * p<0.1.

- The language training is less effective for immigrants with no schooling (about 21.5% of the estimation sample)

Additional results

- No statistically significant differential effect of the language training on labor force participation:
 - by type of migration (family migrants, refugees and labor migrants)
 - by gender
 - by age
- No statistically significant differential effect of the language training on employment, permanent contract, informal work and income per household member

Robustness checks

- Robustness checks to take into account possible manipulation by the language class provider (government)
 - Reclassification of certain individuals' test score from 50 to 45, choosing characteristics that may induce positive selection out of training (education level, permanence in France, distance of French to the native language and the fact of being employed at the time of the test)
 - IV that considers this reclassification and the probability of misclassification

Robustness checks (2)

Table 8: Positive Selection out of Language Training: Labor Force Participation.

Positive Selection Out Of Language Training: Labor Force Participation.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Local linear estimates				Parametric estimates			
	[35, 60]	[35, 60]	[30, 65]	[30, 65]				
Panel A: By education level.								
Second stage regressions (dependent variable: labor force participation dummy)								
Hours of French classes (/100)	0.168** [0.068]	0.215*** [0.069]	0.153** [0.067]	0.144*** [0.043]	0.040 [0.063]	0.143** [0.071]	0.160** [0.071]	0.188** [0.095]
Panel B: By permanence in France.								
Second stage regressions (dependent variable: labor force participation dummy)								
Hours of French classes (/100)	0.198** [0.080]	0.245** [0.121]	0.209** [0.086]	0.171* [0.094]	0.063 [0.063]	0.153** [0.072]	0.219** [0.086]	0.213** [0.105]
Panel C: By language distance (Francophone vs. Non-Francophone).								
Second stage regressions (dependent variable: labor force participation dummy)								
Hours of French classes (/100)	0.174** [0.077]	0.264** [0.115]	0.151* [0.078]	0.200*** [0.077]	0.039 [0.064]	0.143** [0.073]	0.157** [0.079]	0.199** [0.101]
Panel D: By employment in 2010.								
Second stage regressions (dependent variable: labor force participation dummy)								
Hours of French classes (/100)	0.181** [0.073]	0.233** [0.101]	0.194** [0.080]	0.186** [0.082]	0.060 [0.062]	0.150** [0.072]	0.207** [0.081]	0.208** [0.102]
Polynomial of degree	1	1	1	1	1	1	2	2
Control variables	No	Yes	No	Yes	No	Yes	No	Yes
Country of origin fixed effects	No	Yes	No	Yes	No	Yes	No	Yes
Observations	67	67	76	76	2.168	2.168	2.168	2.168

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.

The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50).

The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Île-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects.

*** p<0.01, ** p<0.05, * p<0.1

Robustness checks (3)

- Highly motivated individuals may want to participate in free language classes even though their level of French is high enough to achieve the passing grade in the test
- The high density of "zero" test scores and high value of the residuals in the labor force participation equation for immigrants with "zero" test score may signal the individuals' manipulation of the assignment variable
 - Estimates excluding all the individuals with a "zero" test score, implicitly considering a worst-case scenario where all immigrants with a "zero" test score cheated their entry test to be assigned to the language classes

Robustness checks (4)

Assigned French Language Training and Labor Force Participation: Excluding Test Result = 0.

	(1)	(2) Parametric estimates	(3)	(4)
Second stage regressions (dependent variable: labor force participation dummy)				
Hours of French classes (/100)	0.085 [0.059]	0.178** [0.070]	0.160** [0.072]	0.181* [0.097]
Polynomial of degree	1	1	2	2
Control variables	No	Yes	No	Yes
Country of origin fixed effects	No	Yes	No	Yes
Observations	2,093	2,093	2,093	2,093

Robust standard errors in brackets, clustered by country of origin times test score of the initial language exam.

The columns include functions in the test scores of the initial language exam (distance from the cutoff value) of degree 1 and 2, interacted with I(test score initial language exam < 50). The set of control variable includes: education level in years, age, age squared, a dummy variable equal to 1 if the immigrant is resident in the region "Ile-de-France" (the region around Paris), a dummy variable equal to 1 if the immigrant is married, a dummy variable equal to 1 if the immigrant is male, the number of children in the household, the total number of individuals in the household, the number of years that the migrant has already spent in France, dummy variables indicating the reason to migrate (labor migration, refugee, other. Family migration is the reference category), a dummy variable equal to one if the individual was employed in 2010 and country of origin fixed effects.

*** p<0.01, ** p<0.05, * p<0.1

Language training: possible mechanisms



- **Objective improvement in language skills:** In wave three, thus in 2013, the individuals were asked to take a written and oral language test. We exploit this information to see whether the language training objectively increased the language skills. We observe no significant improvement in objective measures of language skills at the median value of language distance
- **Signaling effect:** We observe a significant impact on labor force participation, but not necessarily on employment. This hints against a signaling effect.
- **Networks:** We find little evidence for networks (French, international, or origin country) playing a significant role.

Language training: possible mechanisms (2)



- **Behavioral effect:** In principle, positive effects if immigrants are encouraged to participate in the labor force because the government implements an integration plan. Instead suggestive evidence of negative effects, probably because immigrants are demotivated after they realize the integration plan is not effective in increasing the probability of employment
- **Information effect:** Positive effects of the language classes on information related to practical aspects of life and employment in France (interactions with classmates and teachers during classes)

Language training: possible mechanisms (3)



Table 7: Mechanisms. Information effect and behavioral effect.

	Information effect				Behavioral effect				Interested in politics of home country
	Job search through: "Pôle Emploi"	Found job through: "Pôle Emploi"	Passed driver's licence in France	Applied for recognition of the latest academic diploma	Feels "at home" in France	Interested in French politics	Interested in European politics		
Hours of French classes (/100)	0.140*** [0.027]	0.255* [0.146]	0.195*** [0.074]	0.640** [0.273]	-0.311** [0.157]	-0.395** [0.457]	-0.217 [0.169]	-0.379* [0.216]	
Polynomial of degree	2	2	2	2	2	2	2	2	2
Observations	363	1,344	797	665	2,157	2,147	2,141	2,148	

Robust standard errors in brackets, clustered by country of origin and test scores of the initial language exam. All columns include functions in the test scores of the initial language exam of degree 1 and 2, interacted with I(test score initial language exam < 50). *** p<0.01, ** p<0.05, * p<0.1.

- Suggestive evidence of negative behavioral effects and positive effects on information related to practical aspects of life and job search strategies in France

Concluding remarks

- We examine the impact of language training on the economic integration of immigrants in France
- We use a local randomized experiment in the form of a Regression Discontinuity Design to estimate the impact of language classes on immigrants' integration
- Main findings:
 - Assigned hours of language training significantly increase labor force participation. The effect is weaker for immigrants with no schooling
 - The mechanisms hint towards an information effect as main driver towards increased labor force participation



Integrating Immigrants: The Impact of Restructuring Active Labor Market Programs

M. Sarvimäki and K. Hämäläinen

Journal of Labor Economics 2016

The treatment and the discontinuity

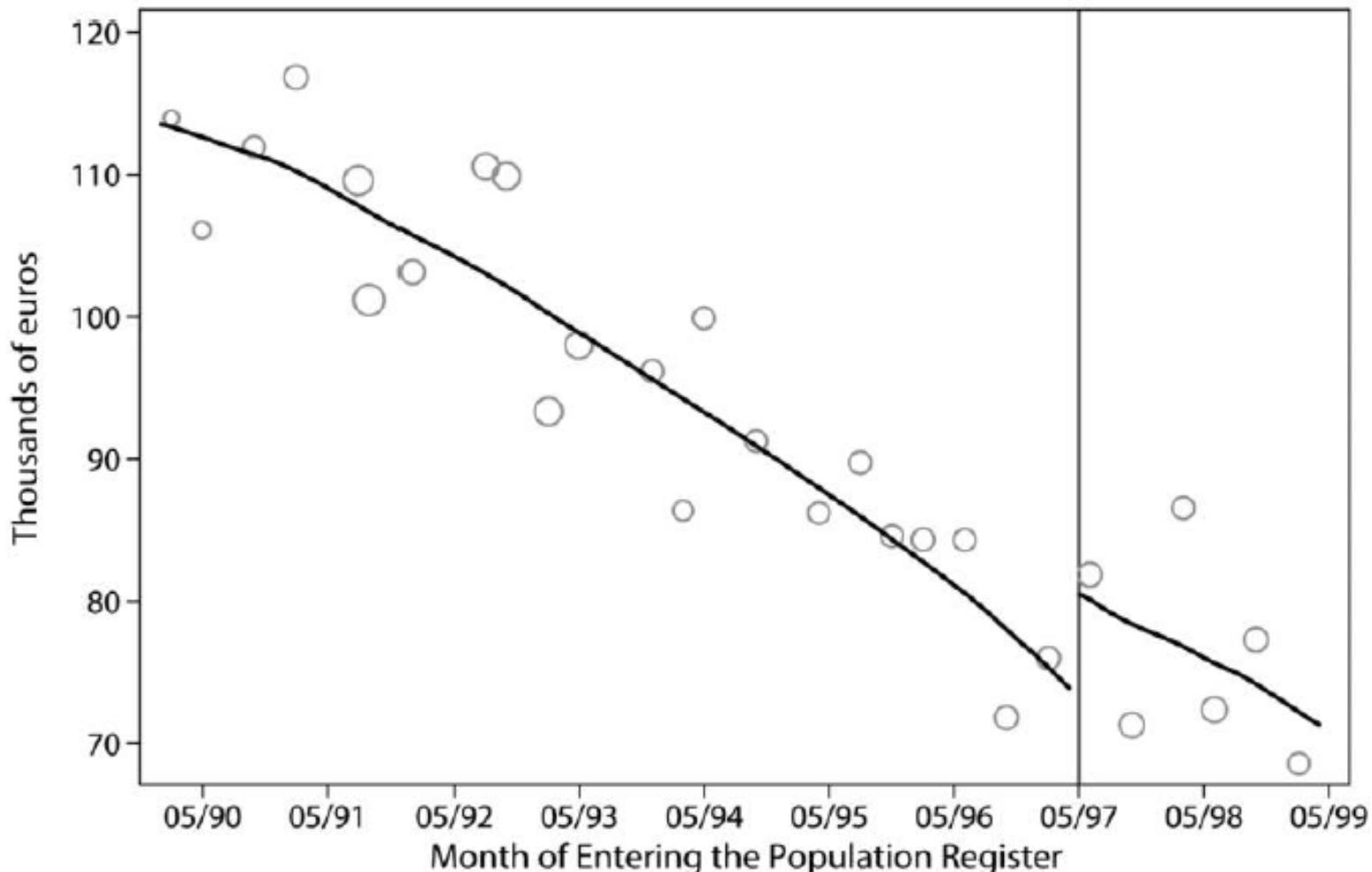


- Individualized integration plans for non-working immigrants who had lived in Finland for less than 3 years
- Integration plans include language courses, training in civic and working life skills, vocational training, subsidized job placements, rehabilitation.
- **Discontinuity:** Participation was mandatory for unemployed immigrants who had entered the population register after May 1, 1997

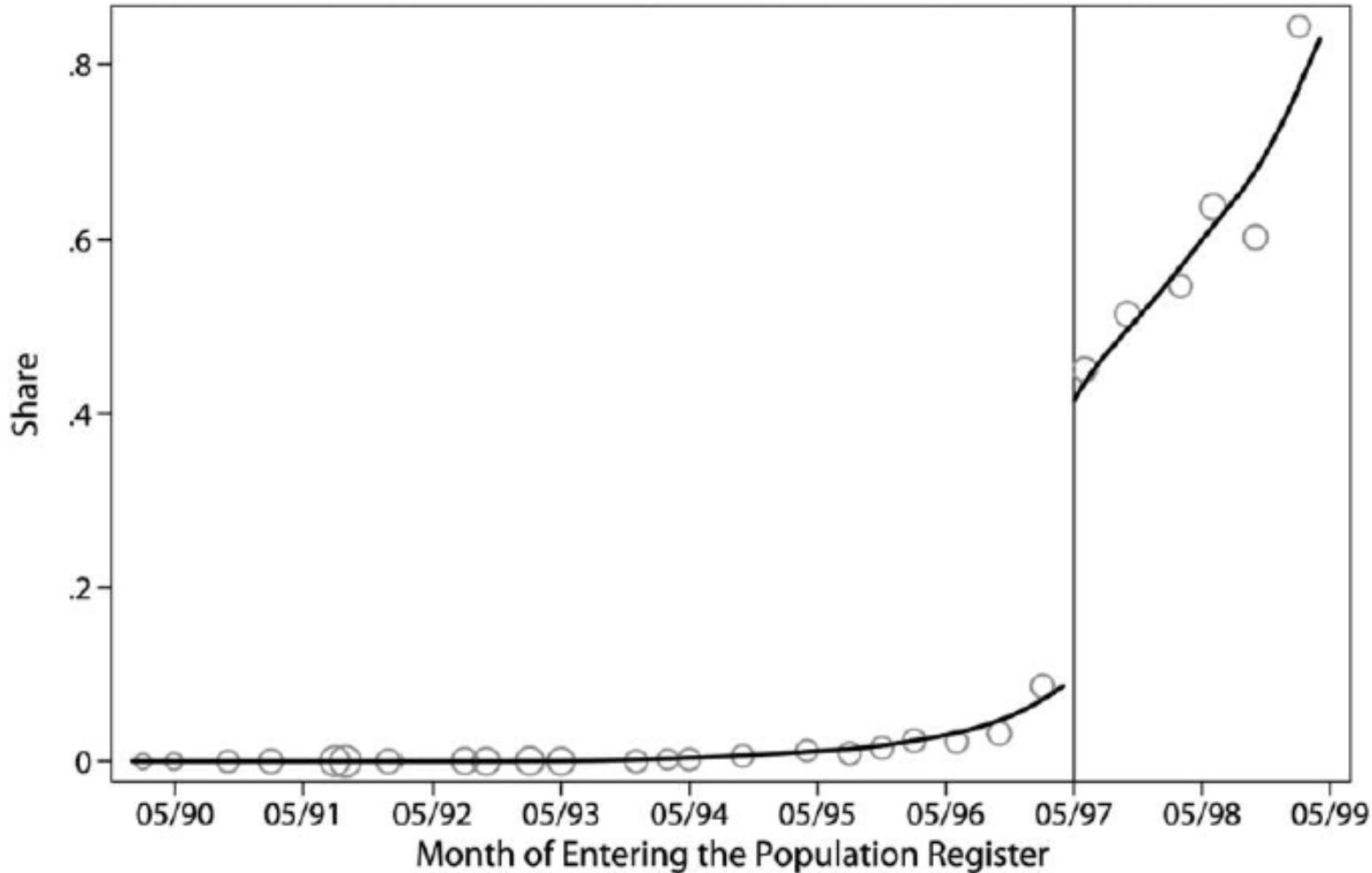
Outcomes of interest

- Cumulative earnings over the period 2000-2009
- Cumulative social benefits over the period 2000-2009
- An indicator for working at all during the 2000s
- Time between arrival and starting in the first job
- Total days in employment during the 2000s
- Occupational quality (measured as the average annual earnings among natives in the same occupation)

Earnings by month of arrival



Integration plans by month of arrival



Integration plans, earnings and benefits

Table 3
Impact of the Integration Plans on Earnings and Benefits

	Earnings		Benefits	
	(1)	(2)	(3)	(4)
Reduced form	7,286 (4,094)	7,238 (3,091)	-2,785 (1,758)	-2,684 (1,281)
First-stage	.35 (.02)	.35 (.02)	.35 (.02)	.35 (.02)
Local average treatment effect (LATE)	20,916 (11,891)	20,702 (9,107)	-8,017 (5,103)	-7,698 (3,681)
Compliers' expected outcomes in the absence of the treatment	44,445 (9,962)	44,420 (8,900)	61,249 (4,314)	60,810 (3,049)
LATE relative to the baseline	.47	.47	-.13	-.13
Additional covariates	No	Yes	No	Yes
Bandwidth (months)	42	42	40	40
First-stage <i>F</i> -statistic for the excluded instrument	322.0	390.1	318.1	384.5
Observations	16,615	16,615	16,173	16,173

NOTE.—The table shows the local linear estimates for the jump at the May 1, 1997, cutoff using the edge (triangle) kernel and the optimal bandwidth selection algorithm of Imbens and Kalyanaraman (2012). Standard errors (in parentheses) are clustered by month of arrival. Additional covariates are age, age squared, sex, marital status, has children, mother tongue Estonian, type of residence municipality (urban, semi-urban, rural), log regional unemployment rate at the year of arrival, region of residence (20 categories), region of birth (10 categories), legal status (refugee, Ingrian Finn, family member, other/unknown). All time-variant characteristics are measured at the first full calendar year the person is resident in Finland. Observations with zero earnings or social benefits are included in the regressions. Social benefits are measured at the household level using an equivalence scale, which assigns a value of 1 to the first household member, .7 to other adults, and .5 to each child.

Integration plans, employment and occupation



Table 4
Impact on Employment and Occupations

	Any Employment		Days to First Job		Days Employed		Occupational Quality	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reduced form	−.02 (.02)	−.01 (.01)	39 (76)	45 (60)	15 (48)	8 (37)	696 (332)	582 (346)
First-stage	.33 (.02)	.32 (.02)	.34 (.02)	.34 (.02)	.35 (.02)	.35 (.02)	.36 (.02)	.36 (.02)
Local average treatment effect (LATE)	−.07 (.05)	−.04 (.04)	115 (224)	133 (178)	42 (137)	24 (105)	1,860 (892)	1,537 (902)
Compliers' expected outcomes in the absence of the treatment	.78 (.05)	.76 (.04)	1,430 (202)	1,400 (169)	1,061 (108)	1,078 (91)	23,033 (824)	23,381 (861)
LATE relative to the baseline	−.09	−.05	.08	.10	.04	.02	.08	.07
Additional covariates	No	Yes	No	Yes	No	Yes	No	Yes
First-stage F-statistic for the excluded instrument	260.2	251.8	356.0	445.3	323.6	392.4	476.4	589.4
Bandwidth (months)	17	17	28	28	42	42	54	54
No. of observations	9,486	9,486	10,851	10,851	16,615	16,615	14,021	14,021

NOTE.—The table shows local linear estimates for the jump at the May 1, 1997, cutoff using the edge (triangle) kernel and the optimal bandwidth selection algorithm of Imbens and Kalyanaraman (2012). Standard errors (in parentheses) are clustered by month of arrival. Any employment is an indicator of the immigrant working at least 2 weeks in a nonsubsidized job during the 2000s. Cumulative employment is the number of days in nonsubsidized employment during the 2000s. Occupational quality is measured as the average native earnings in the occupations that the immigrant held between 2004 and 2009. For additional covariates, see the note to table 3.

Outline of the talk

- Barriers to immigrants' assimilation
 - Language barriers
 - Language skills and labor market outcomes
 - Language training and labor market outcomes
 - Undocumented status
 - Undocumented status and labor market outcomes
 - Undocumented status and consumption behavior
 - Undocumented status and propensity to commit crime



Illegal Migration and Consumption Behavior of Immigrant Households

Dustmann, C., Fasani, F., and B. Speciale,
Journal of the European Economic
Association 2017, 15(3)

Introduction

- **Research question:** What are the effects of immigrants' legal status on their consumption behavior?
- **Importance of the research question:**
 - Large size of the undocumented immigrant population
 - Adoption of immigration policies that affect immigrants' legal status is at the core of the public policy debate in several OECD countries

Stocks of undocumented immigrants

Country	Estimated stock	Year
US	12-13 million	2008
Austria	100 thousand	2003
Italy	650 thousand	2006
Germany	500 thousand	2004
Greece	250 thousand	2007
Spain	570 thousand	2008
UK	430 thousand	2001

Source: PEW Hispanic Center (US); Clandestino project (Undocumented migrants: counting the uncountable) (Europe)

Related literature on immigrants' legal status and labor market outcomes



- Legal status and labor market outcomes (Borjas and Tienda, 1993; Kossoudji and Cobb-Clark, 2002; Kaushal, 2006; Amuedo-Dorantes et al., 2007)
- After legalization, newly legalized immigrants have higher wages than before legalization
- The employment effect remains empirically ambiguous

Introduction (2)

- **Why is the analysis of this research question challenging?**
 - Selection into illegal status that depends on unobservable characteristics
 - Importance of identifying the mechanism driving the relationship of interest
- **Methodology:**
 - Rainfall shocks in the origin country at the time of emigration as IV for immigrants' legal status
 - Exposure to amnesty quotas that grant legal residence status to undocumented immigrants as IV
 - Variations in deportation risk as an alternative source of variation in uncertainty within the illegal immigrant population

Background information on immigration in Italy



- Background information on immigration in Italy:
 - Italy is a new immigration country
 - Foreign-born **documented** population: 3mln in 2006 ($\approx 5\%$ of total population)
 - Estimated **undocumented** immigrant population: 650 th in 2006 ($\approx 15\%$ of foreign-born population)
 - Large **shadow economy**: 26% of official GDP (8% in the US) (Schneider, 2007)
 - Relevant **probability of deportation**: 4% in 2005 (≈ 0 in US)
 - Channels for documented migration: 5 amnesties from 1986 to 2002; Quota system

Background information on immigration in Italy (2)



- Deportations of undocumented migrants (Average values: 2004-2006)

	Undocumented migrants stock	Deportations	% apprehended	% deported
Italy	595500	24624	15.4	4.1
US	12000000	256023	2	0.00003

Sources: ISMU, Italian Ministry of Interior, Pew Hispanic Center and US Department for Homeland Security

- High risk of deportation → high levels of uncertainty for undocumented immigrants in Italy

Econometric model

- Estimated equation:

$$C_{ict} = \alpha + \beta Illegal_{ict} + \gamma X_{ict} + f_c + d_t + \varepsilon_{ict}$$

where i=individual migrant, c=country of origin and t=year of interview

Selection into illegal status

- Main source of bias for OLS estimates: **Selection into illegal status**
 - Unobservable characteristics that affect both the legal status and the consumption behavior (e.g., risk aversion)
 - Solution of the identification issue: Instrumental Variable estimation
- **Identification strategy:**
 - Rainfall shocks as IV
 - Exposure to amnesty quotas as IV

Rainfall shocks as IV

- **Rainfall shocks in the origin country at the time of emigration** as exogenous source of variation in current legal status
 - Almost all countries of origin of immigrants residing in Italy are low and middle-income countries, and highly dependent on the agricultural sector
 - Legal migration is a time-consuming option (individuals need to wait for the visa) → unexpected temporary income shocks allow those who would like to emigrate to cover the cost of an immediate illegal migration (rather than applying for the lengthy legal pathway)

Countries of origin

Table A1—Countries of origin, undocumented shares and agricultural sector: First 20 nationalities

Country	Sample		Agriculture sector (1995-2007)		
	Observations	Share of total sample	Share of undocumented immigrants	Employment share	GDP share
MOROCCO	1576	0.12	0.10	44.88	15.83
ALBANIA	1346	0.10	0.09	66.17	30.66
ROMANIA	874	0.06	0.22	38.18	15.52
SENEGAL	787	0.06	0.20	-	18.76
EGYPT	682	0.05	0.20	30.06	16.54
ECUADOR	615	0.04	0.16	7.81	11.12
PERU	593	0.04	0.18	4.05	8.30
UKRAINE	579	0.04	0.29	21.47	14.12
PAKISTAN	547	0.04	0.10	45.40	24.99
INDIA	526	0.04	0.04	66.70	23.94
TUNISIA	477	0.03	0.08	-	12.24
CHINA	447	0.03	0.08	46.64	15.60
MOLDOVA	327	0.02	0.33	46.36	26.66
COTE D'IVOIRE	307	0.02	0.05	-	23.45
PHILIPPINES	238	0.02	0.09	38.85	16.84
SRI LANKA	235	0.02	0.08	34.74	20.29
BRAZIL	222	0.02	0.15	22.80	5.98
NIGERIA	219	0.02	0.15	-	29.63
BANGLADESH	209	0.02	0.08	59.00	24.07
GHANA	203	0.01	0.08	55.00	36.63
Other 79 nationalities	2663	0.19	0.09	-	-
Total	13672	1.00	0.13		
Sample weighted avg (first 20 countries)				40.7	19.1
Italy				5.5	2.8
OECD countries				4.2	1.9

Note. This table reports the nationality composition of our sample. For each country of origin, we report, in the first two columns, the number of observations and the percentage they represent in our sample. In the last two columns, we report World Bank data on shares of employment and GDP in agriculture. In the last three rows, we report shares of employment and GDP in agriculture for our sample (after computing the average across countries weighted by the number of observations from each country), Italy, and all OECD countries.

Rainfall shocks as IV (2)

- Advantages of using rainfall shocks as IV
 - Exogenous to individual decisions/Non correlated with individual risk aversion
 - Transitory income shocks

Rainfall shocks as IV (3)

- The IV approach allows to identify the Local Average Treatment Effect (**LATE**) parameter
- Group of **compliers**: those who respond to positive rainfall shocks by switching from the legal to the illegal option

Rainfall shocks as IV (4)

- Monthly **rainfall data from NASA**, from 1979 onwards (Global Precipitation Climatology Project). Rainfall estimates for each point at which latitude and longitude degree lines cross, at 2.5-degree intervals
- **Instrumental variable construction:**
 - Logarithm of rainfall in the country of origin → Conditional on origin country dummies, deviation of rainfall in origin countries from its long term mean at the time of emigration (weather-induced transitory income shock)

Rainfall shocks as IV (5)

- **Instrumental variable construction:**
 - Possible lag between decision to migrate and actual migration → Average of rainfalls at the time of emigration (T) and the year before ($T-1$)

Rainfall shocks as IV (6)

- Related literature that uses weather shocks as IV: Munshi (QJE, 2003), Miguel et al. (JPE, 2004), Miguel (RES, 2005), Maccini and Yang (AER, 2008), Bruckner and Ciccone (Econometrica, 2011)
- Our contribution to this literature:
 - Use of **rainfall shocks** in the origin country at the time of emigration to predict **legal status of migrants**

First stage regressions: Rainfall shocks as IV



	Alternative definitions of Illegal residence status						Demographics characteristics				
	Illegal						Illegal2	Illegal3	In (Age)	Female	Higher Education
	1	2	3	4	5	6	7	8	9	10	11
In (Rainfall (T))	0.032*** [0.005]	0.030*** [0.005]	0.031*** [0.005]								
In (Rainfall (T-1))		0.018*** [0.005]	0.017*** [0.005]								
In (Rainfall (T-2))			0.006 [0.004]								
In (Rainfall (T,T-1))				0.046*** [0.007]	0.046*** [0.007]	0.045*** [0.007]	0.049*** [0.008]	0.039*** [0.007]	0.002 [0.003]	-0.004 [0.005]	0.007 [0.006]
In (Current Rainfall)				-0.002 [0.006]	-0.002 [0.006]	-0.003 [0.007]	0.002 [0.006]	0.005 [0.008]			
Baseline controls	X	X	X	X	X	X	X	X	X	X	X
Individual controls				X	X	X	X	X			
HH controls					X	X					
In(income)						X					
Observations	13,672	13,672	13,672	13,672	13,672	13,672	13,672	13,672	13,672	13,672	13,672
R-squared	0.211	0.214	0.214	0.234	0.240	0.254	0.237	0.205	0.133	0.182	0.078
IV: F-stat	34.46	20.93	15.18	40.15	39.49	35.55	38.51	29.32	0.31	0.45	1.40
IV: p-value F-stat	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.58	0.50	0.24

Note. This table reports LPM estimates of the "Illegal residence status" dummy (columns 1-8) and of individual characteristics (log age, gender, education; columns 9-11) on rainfall and other controls. The dummy *Illegal residence status* equals one if the respondent lacks legal status (alternative definitions of illegal residence status are used in columns 7 and 8). The variables *In (Rainfall)* are the logarithm of rainfall in the country of origin (normalized by the average within-country standard deviation) in the year of emigration (T), one year before emigration (T-1), two years before emigration (T-2), and averaged over the year of migration and the year before (T, T-1). The variable *In (Current Rainfall)* is the logarithm of rainfall in the country of origin (normalized by the average within-country standard deviation) in the year of the interview. *Baseline controls* include: origin country dummies, year dummies (2004-2007) and number of years of residence in Italy of the respondent. *Individual controls* include: gender, age, age squared, dummies for education level (none, primary, secondary, tertiary) and dummies for Italian province of residence. *HH controls* include: number of members in the household living in Italy, a dummy for spouse living abroad, number of children living in Italy and abroad, dummy for home ownership. *In (income)* is log monthly income. For each column, the table reports the F-statistic (*IV: F-stat*) and the p-value (*IV: p-value F-stat*) from a joint significance test of the excluded instrument(s). Sample: immigrants with 1–10 years of residence in Italy. ISMU survey years: 2004–2007. Standard errors: robust and clustered by country of origin (103 clusters); *significant at 10%; **significant at 5%; ***significant at 1%.

First stage regressions: Rainfall shocks as IV (2)



- A one standard deviation increase in the level of precipitations in the year of emigration (the year earlier) leads to a 3 (2) percentage points increase in the probability of being an illegal resident in Italy
- Sizeable effects: weather conditions are an important factor in explaining illegal migration flows.

Rainfall shocks at the time of migration and immigrants' predetermined characteristics



- No significant relationship between weather shocks and predetermined personal characteristics at emigration (age, gender and emigration)

First stage regressions: Rainfall shocks as IV. Agriculture and financial development



Illegal residence status	Large countries excluded:											
	First 10	First 20	First 30	8	9	10	1	2	3	4	5	6
Log Rainfall (T,T-1)	0.046*** [0.007]	0.044*** [0.008]	0.041*** [0.008]	0.044*** [0.009]	0.040*** [0.009]	0.053*** [0.007]	0.049*** [0.008]	0.047*** [0.007]	0.048*** [0.008]	0.044*** [0.008]		
Log Rainfall (T,T-1): * dummy (agriculture GDP share above 75th pct)		0.038*** [0.012]	0.043*** [0.010]									
* dummy (agriculture GDP share below 25th pct)		-0.016 [0.020]	-0.008 [0.019]									
* dummy (agriculture empl share above 75th pct)				0.040** [0.016]	0.044*** [0.016]							
* dummy (agriculture empl share below 25th pct)				-0.003 [0.045]	0.014 [0.044]							
* dummy (banking sector - above 75th pct)						-0.012** [0.005]	-0.010** [0.005]					
* dummy (banking sector - below 25th pct)						0.018 [0.011]	0.017 [0.011]					
Current log rainfall	-0.002 [0.006]	-0.002 [0.006]	-0.001 [0.006]	-0.005 [0.007]	-0.002 [0.007]	-0.003 [0.007]	-0.002 [0.007]	-0.002 [0.006]	-0.003 [0.007]	0.003 [0.005]		
Baseline controls	X	X	X	X	X	X	X	X	X	X		
Individual controls	X	X	X	X	X	X	X	X	X	X		
Log GDP per capita (T)			X		X		X					
Observations	13,672	13,613	13,602	11,251	11,249	13,459	13,454	12,120	11,290	9,534		
R-squared	0.234	0.235	0.236	0.243	0.245	0.240	0.241	0.242	0.245	0.230		
IV: F-stat	40.15	48.29	54.76	24.20	20.80	20.36	13.00	39.60	37.36	27.20		
IV: p-value F-stat	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00		

Note. This table reports LPM estimates of the "illegal residence status" dummy on rainfall and other controls. The dummy *illegal residence status* equals one if the respondent lacks legal status. In columns 2-5, *Log Rainfall (T, T-1)* is interacted with dummies that identify countries whose GDP (employment) share in agriculture is, respectively, above the 75th percentile or below the 25th percentile of the distribution in the countries we have in our sample. In columns 6-7, *Log Rainfall (T, T-1)* is interacted with dummies that identify countries whose banking sector (as a share of GDP) is, respectively, above the 75th percentile or below the 25th percentile of the distribution in the countries we have in our sample. In columns 3, 5 and 7, we control for the log of real GDP per capita in the country of origin at the time of emigration (*Log GDP per capita (T)*). In columns 8-10, we exclude immigrants from, respectively, the first 10, 20 and 30 largest countries in our sample. The countries of origin ranking according to their total land area is: 1) China; 2) Brazil; 3) India; 4) Argentina; 5) Kazakhstan; 6) Sudan; 7) Algeria; 8) Congo, the Democratic Republic; 9) Mexico; 10) Saudi Arabia; 11) Indonesia; 12) Libya; 13) Iran; 14) Peru; 15) Chad; 16) Niger; 17) Angola; 18) Mali; 19) Colombia; 20) Ethiopia; 21) Bolivia; 22) Mauritania; 23) Egypt; 24) Tanzania; 25) Nigeria; 26) Venezuela; 27) Namibia; 28) Pakistan; 29) Mozambique; 30) Turkey. *Current Log Rainfall*, *Baseline controls* and *Individual controls* are defined as in the note to Table 3. For each column, the table reports the F-statistic (IV: F-stat) and the p-value (IV: p-value F-stat) from a joint significance test of the excluded instrument(s). Sample: immigrants with 1–10 years of residence in Italy. ISMU survey years: 2004–2007. Standard errors: robust and clustered by country of origin (103 clusters); *significant at 10%; **significant at 5%; ***significant at 1%.

Exposure to amnesty quotas as IV

- In the late 1990s, Italy adopted a quota system that was meant to regulate the entry of migrant workers to the Italian labor market.
- Since its introduction, however, the system has been widely used to grant legal status to undocumented immigrants who reside in Italy
- Total number of residence permits offered through the quota system since arrival in Italy as IV

First stage regressions: quotas as IV



	1	2	3	4
Panel A: main quota IV				
In (Total Quotas)	-0.309*** [0.039]	-0.301*** [0.039]	-0.306*** [0.039]	-0.294*** [0.038]
IV: F-stat	61.32	59.72	61.69	58.59
IV: p-value F-stat	0.00	0.00	0.00	0.00
Panel B: alternative quota IVs				
In (Total Quotas) - "adaptive" expect.	-0.118*** [0.042]	-0.114*** [0.041]	-0.116*** [0.041]	-0.111*** [0.039]
IV: F-stat	7.876	7.543	7.754	8.012
IV: p-value F-stat	0.01	0.01	0.01	0.01
In (Country-Reserved Quotas)	-0.324*** [0.049]	-0.314*** [0.048]	-0.318*** [0.048]	-0.308*** [0.047]
In (Non Country-Reserved Quotas)	-0.149*** [0.027]	-0.146*** [0.026]	-0.149*** [0.026]	-0.140*** [0.026]
IV: F-stat	49.60	48.07	50.67	46.68
IV: p-value F-stat	0.00	0.00	0.00	0.00
Baseline controls	X	X	X	X
Individual controls		X	X	X
HH controls			X	X
In (income)				X
Observations	13,672	13,672	13,672	13,672

Note. This table reports LPM estimates of the “Illegal residence status” dummy on the quota system instruments. The dummy *Illegal residence status* equals one if the respondent lacks legal status. In Panel A, we use our main quota instrument, *In (Total Quotas)*, i.e. the residuals of log of total quotas since arrival in Italy after estimating equation (2) (see Section 2.3). In panel B, we use two alternative quota instruments. First, we use the instrument *In (Total Quotas) - “adaptive” expect.* (as defined in Section 2.3) and we then jointly use the residuals of the total number of residence permits offered through the quota system to citizens of Albania, Morocco and Tunisia (*In (Country-Reserved Quotas)*) and of the total number of residence permits offered to immigrants of all other nationalities (*In (non-Country-Reserved Quotas)*). *Baseline*, *individual* and *HH controls* and *In (income)* are defined as in the note to Table 3. For each regression, the table reports the F-statistic (IV: F-stat) and the p-value (IV: p-value F-stat) from a joint significance test of the excluded instrument(s). Sample: immigrants with 1–10 years of residence in Italy. ISMU survey years: 2004–2007. Standard errors: robust and clustered by years since arrival in Italy and year of interview (40 clusters); *significant at 10%; **significant at 5%; ***significant at 1%.

Second stage regressions: Illegal status and consumption



	IV estimates												Alternative quota IVs		
	OLS estimates			In (Rainfall (T,T-1))			In (Total Quotas)			In (Rainfall (T,T-1)) and In (Total Quotas)		In (Total Quotas) - "adaptive" expect.		In (Country-Reserved Quotas) - and In (Non Country-Reserved Quotas)	
	1	2	3	4	5	6	7	8	9	10	11	12			
Illegal residence status	-0.303*** [0.026]	-0.287*** [0.025]	-0.578*** [0.200]	-0.571*** [0.185]	-0.510*** [0.060]	-0.533*** [0.058]	-0.519*** [0.067]	-0.537*** [0.064]	-0.557*** [0.106]	-0.591*** [0.100]	-0.523*** [0.067]	-0.542*** [0.064]			
Baseline controls	X	X	X	X	X	X	X	X	X	X	X	X			
Individual controls	X	X	X	X	X	X	X	X	X	X	X	X			
HH controls		X		X		X		X		X		X			
Observations	13,672	13,672	13,672	13,672	13,672	13,672	13,672	13,672	13,672	13,672	13,672	13,672			
IV: F-stat	-	-	39.12	38.52	59.72	61.69	54.33	55.71	7.543	7.754	48.07	50.67			
IV: p-value F-stat	-	-	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00			
Durbin-Wu-Hausman test: p-value	-	-	0.13	0.09	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00			
Hansen overidentification test: p-value	-	-	-	-	-	-	0.71	0.82	-	-	0.44	0.54			

Note. This table reports OLS (columns 1-2) and IV estimates (columns 3-12) of log monthly consumption on illegal residence status and other controls. The dummy *Illegal residence status* equals one if the respondent lacks legal status. In columns 3-12 we treat the variable *Illegal residence status* as endogenous and instrument it with alternative sets of instrumental variables: a) *In (Rainfall (T, T-1))* (columns 3-4); b) *In (Total Quotas)* (columns 5-6); c) *In (Rainfall (T, T-1))* and *In (Total Quotas)* (columns 7-8); d) *In (Total Quotas)- "adaptive" expect.* (columns 9-10); e) *In (Country-Reserved Quotas)* and *In (non-Country-Reserved Quotas)* (columns 11-12). Odd columns condition on baseline and individual controls, while even columns condition also on household controls. *Baseline*, *individual* and *HH controls* are defined as in the note to Table 3. For each column, the table reports the F-statistic (IV: F-stat) and the p-value (IV: p-value F-stat) from a joint significance test of the excluded instrument(s) and the p-value from a regression-based version of the Durbin-Wu-Hausman test for endogeneity (Durbin-Wu-Hausman test: p-value). When illegal residence status is instrumented with more than one instrument (columns 7-8 and 11-12) the p-value from the Hansen overidentification test is reported. Sample: immigrants with 1–10 years of residence in Italy. ISMU survey years: 2004–2007. Standard errors: robust and clustered by country of origin (103 clusters); *significant at 10%; **significant at 5%; ***significant at 1%.

Illegal status and consumption

- **Upward bias in the OLS estimates**
 - Less risk averse individuals tend to select into illegal migration
 - Less risk averse individuals tend to consume more/save less for precautionary reasons

Illegal status and consumption

Conditioning on income



	IV estimates						
	OLS	In (Rainfall (T,T-1))		In (Total Quotas)		In (Rainfall (T,T-1)) and In (Total Quotas)	
	1	2	3	4	5	6	7
Illegal residence status	-0.132*** [0.025]	-0.296** [0.132]	-0.394** [0.161]	-0.182* [0.107]	-0.323** [0.141]	-0.196** [0.097]	-0.338*** [0.128]
In (income)	0.773*** [0.014]	0.757*** [0.016]	0.487*** [0.177]	0.768*** [0.016]	0.458** [0.179]	0.766*** [0.016]	0.446** [0.182]
Variables instrumented:							
Illegal residence status	-	X	X	X	X	X	X
In (income)	-	-	X	-	X	-	X
All controls	X	X	X	X	X	X	X
Observations	13,672	13,672	13,672	13,672	13,672	13,672	13,672

Note. This table reports OLS (columns 1) and IV estimates (columns 2-7) of log monthly consumption on illegal residence status, log monthly income and other controls. The dummy *Illegal residence status* equals one if the respondent lacks legal status. In columns 2-7, we treat the variable *Illegal residence status* as endogenous and instrument it with: a) *In (Rainfall (T, T-1))* (columns 2-3); b) *In (Total Quotas)* (columns 4-5); c) *In (Rainfall (T, T-1))* and *In (Total Quotas)* (columns 6-7). In columns 3, 5 and 7 we instrument *In (income)* with the unemployment rate in the province of residence. All controls include *baseline*, *individual* and *HH controls* as defined in the note to Table 3. Sample: immigrants with 1–10 years of residence in Italy. ISMU survey years: 2004–2007. Standard errors: robust and clustered by country of origin (103 clusters); *significant at 10%; **significant at 5%; ***significant at 1%.

Illegal status and consumption (2)



- **Without conditioning on household income:**
 - **IV estimates:** Illegal immigrants consume about 40% less than legal immigrants
- **Keeping income constant between the two groups:**
 - **IV estimates:** illegal immigrants consume about 25%-30% less than legal immigrants

Illegal status and consumption (3)

- Comparing the IV estimates with and without household income: about 25 percent of the reduction in consumption is driven by illegal immigrants having lower incomes than legal immigrants

Illegal status and household consumption: robustness checks



	Alternative measures of consumption			Alternative definitions of Illegal residence status		Subsamples	
	"Modified" OECD equivalence scale	OECD equivalence scale	Unweighted normalization	Illegal2	Illegal3	Excluding home owners	Only one- member HHs
	1	2	3	4	5	6	7
Panel A: OLS estimates							
Illegal residence status	-0.303*** [0.026]	-0.262*** [0.027]	-0.203*** [0.029]	-0.316*** [0.027]	-0.238*** [0.025]	-0.285*** [0.027]	-0.275*** [0.025]
Panel B: IV estimates							
IV: $\ln(\text{Rainfall}(T, T-1))$							
Illegal residence status	-0.578*** [0.200]	-0.633*** [0.201]	-0.712*** [0.208]	-0.546*** [0.193]	-0.680*** [0.240]	-0.571*** [0.180]	-0.534*** [0.131]
IV: $\ln(\text{Total Quotas})$							
Illegal residence status	-0.510*** [0.060]	-0.531*** [0.062]	-0.570*** [0.073]	-0.511*** [0.065]	-0.568*** [0.065]	-0.505*** [0.065]	-0.543*** [0.084]
IV: $\ln(\text{Rainfall}(T, T-1))$ and $\ln(\text{Total Quotas})$							
Illegal residence status	-0.519*** [0.067]	-0.544*** [0.068]	-0.587*** [0.076]	-0.516*** [0.072]	-0.580*** [0.072]	-0.514*** [0.068]	-0.541*** [0.077]
Baseline controls	X	X	X	X	X	X	X
Individual controls	X	X	X	X	X	X	X
Observations	13,672	13,672	13,672	13,672	13,672	11,879	6,069

Note. This table reports OLS (Panel A) and IV estimates (Panel B) of log monthly consumption on illegal residence status and other controls. The dummy *Illegal residence status* equals one if the respondent lacks legal status. Each row in Panel B reports IV estimates obtained using a different instrument (set of instruments): a) $\ln(\text{Rainfall}(T, T-1))$; b) $\ln(\text{Total Quotas})$; c) $\ln(\text{Rainfall}(T, T-1))$ and $\ln(\text{Total Quotas})$. We perform the following robustness checks: a) we use three alternative definitions of individual consumption, as defined in Section 3.1 (columns 1-3); b) we use two alternative definitions of illegal status, as defined in online-Appendix Table A.2 (columns 4-5); c) we drop from the main sample all respondents who report to be living in their own house (columns 6); d) we restrict the estimating sample to households with only one member (column 7). *Baseline* and *individual controls* are defined as in the note to Table 3. Sample: immigrants with 1–10 years of residence in Italy. ISMU survey years: 2004–2007. Standard errors: robust and clustered by country of origin (103 clusters); *significant at 10%; **significant at 5%; ***significant at 1%.

Alternative definitions of undocumented immigrants



Table A2—Alternative definitions of undocumented immigrants

Current residence status	obs.	%	Undocumented immigrants		
			Illegal	Illegal2	Illegal3
Italian citizenship	320	2.34			
Permanent residence permit	2,022	14.79			
Temporary residence permit	8,852	64.75			
No residence permit	1,697	12.41	X	X	X
Applicant for legalization	110	0.8	X		X
Renewing temporary residence permit	671	4.91			X
Total observations	13672	100	1,807	1,697	2,478
Undocumented share			0.13	0.12	0.18

Note. ISMU data. Sample: immigrants with 1–10 years of residence in Italy. ISMU survey years: 2004–2007.

Probability of deportation and household consumption



	Undocumented immigrants				Documented immigrants			
	1	2	3	4	5	6	7	8
Prob. of deportation	-0.157*** [0.036]	-0.115*** [0.041]	-0.117*** [0.039]	-0.116*** [0.032]	-0.012 [0.015]	-0.012 [0.014]	-0.012 [0.014]	-0.010 [0.006]
Province and year dummies	X	X	X	X	X	X	X	X
Individual controls		X	X	X		X	X	X
HH controls			X	X			X	X
ln (income)				X				X
Observations	1,299	1,299	1,299	1,299	8,413	8,413	8,413	8,413
R-squared	0.100	0.317	0.325	0.569	0.141	0.169	0.190	0.588

Probability of deportation (%)	Mean	Std Dev	Within Std Dev	Min-Max
	2.59	1.68	1.15	0.04 - 6.76

Note. This table reports OLS regression estimates of log of monthly consumption on the probability of deportation in the province of residence and other controls. *Province and year dummies* include: dummies for Italian province of residence and year dummies (2004-2006). *Individual controls* include: gender, age, age squared, dummies for education level (none, primary, secondary, tertiary), origin country dummies, number of years of residence in Italy of the respondent. *HH controls* include: number of members in the household living in Italy, a dummy for spouse living abroad, number of children living in Italy and abroad, dummy for home ownership. Sample: immigrants with 1–10 years of residence in Italy. ISMU survey years: 2004–2006. Standard errors (in brackets): robust and clustered by Italian province of residence (11 provinces) and year of the interview (3 years); *significant at 10%, **significant at 5%, ***significant at 1%.

Probability of deportation and household consumption (2)



- Within-province variation in deportation probabilities: alternative source of variation in uncertainty within the illegal immigrant population
- Measure of risk of deportation = # of migrants deported / estimated stock of undocumented immigrants
- A one (within-province) standard deviation increase in the probability of deportation reduces consumption by about 13% for undocumented immigrants. Statistically insignificant effect for legal immigrants

Concluding remarks

- Analysis of the effects of legal status on immigrants' consumption behavior
- Severe bias of OLS estimates due to **selection into legal status**
- Solutions for the identification issue:
 - Rainfall shocks in the origin country at the time of emigration as exogenous source of variation in current legal status
 - Exposure to amnesty quotas as IV

Concluding remarks (2)

- **Main findings:**
 - Undocumented immigrants consume about 40% less than documented immigrants
 - About 25% of this difference is due to lower income
 - The remaining difference is due to higher uncertainty, which leads to precautionary behavior

Concluding remarks (3)

- Findings from first stage regressions: **Weather shocks** can affect the **dynamics of the undocumented immigrant population**
- **Regularization programs** in immigrants' host countries may have **large distributional consequences** in favor of host countries through an increase in consumption

Consumption sub-categories

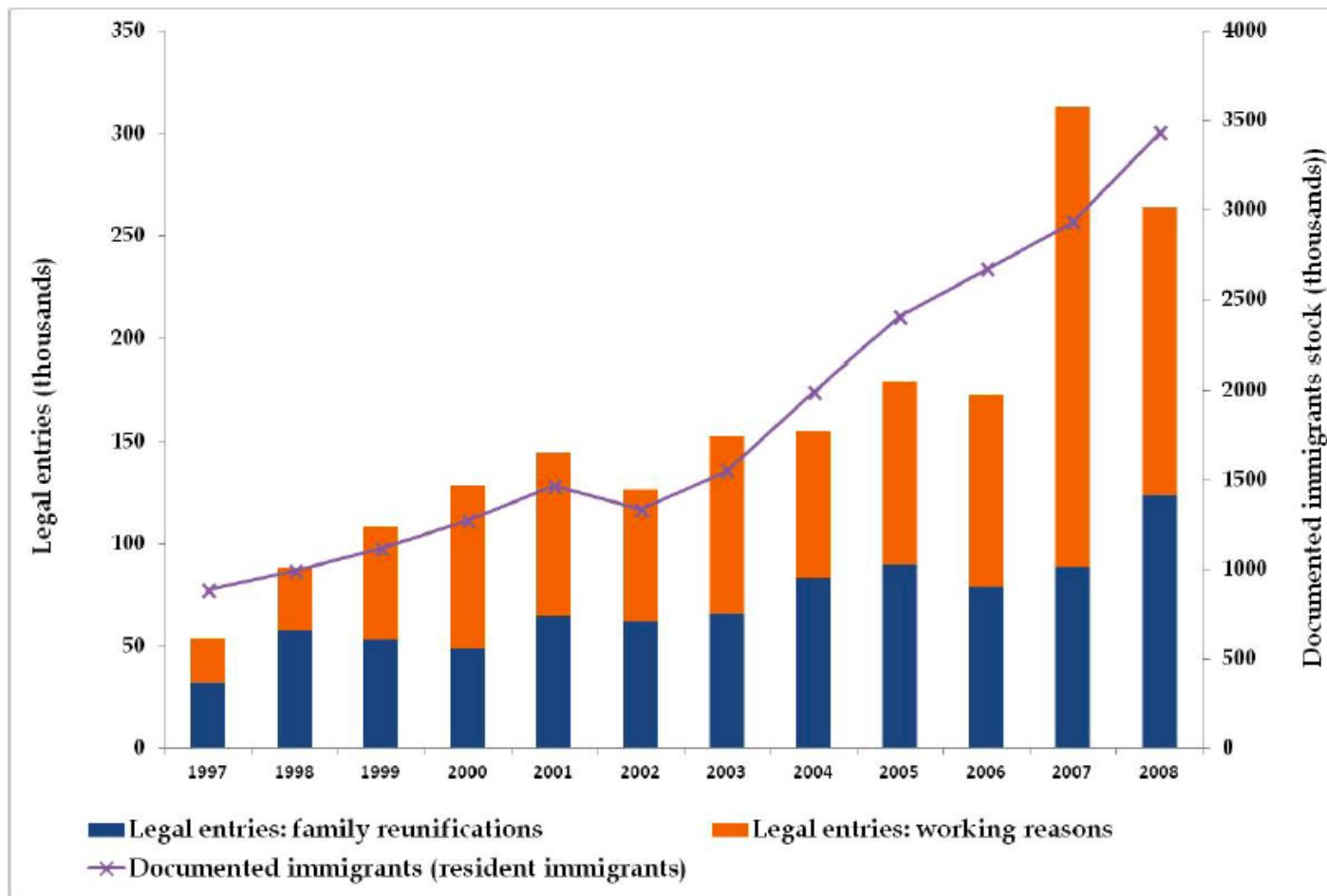
	Food & Clothes			Housing			Other		
	1	2	3	4	5	6	7	8	9
Panel A: OLS estimates									
Illegal residence status	-0.292*** [0.049]	-0.274*** [0.044]	-0.130** [0.053]	-0.821*** [0.085]	-0.748*** [0.079]	-0.471*** [0.072]	-0.308*** [0.045]	-0.314*** [0.044]	-0.126*** [0.036]
Panel B: IV estimates									
IV: $\ln(\text{Rainfall}(T, T-1))$									
Illegal residence status	-0.721*** [0.239]	-0.718*** [0.233]	-0.470** [0.191]	-1.610*** [0.507]	-1.660*** [0.467]	-1.183*** [0.450]	-0.612 [0.387]	-0.567 [0.380]	-0.228 [0.282]
IV: $\ln(\text{Total Quotas})$									
Illegal residence status	-0.554*** [0.098]	-0.573*** [0.099]	-0.252** [0.124]	-1.227*** [0.333]	-1.375*** [0.315]	-0.756*** [0.291]	-0.777*** [0.150]	-0.736*** [0.145]	-0.313*** [0.119]
IV: $\ln(\text{Rainfall}(T, T-1))$ and $\ln(\text{Total Quotas})$									
	-0.574*** [0.098]	-0.590*** [0.100]	-0.278** [0.117]	-1.273*** [0.318]	-1.409*** [0.302]	-0.808*** [0.272]	-0.757*** [0.149]	-0.715*** [0.144]	-0.303*** [0.110]
Baseline controls	X	X	X	X	X	X	X	X	X
Individual controls	X	X	X	X	X	X	X	X	X
HH controls		X	X		X	X		X	X
$\ln(\text{income})$			X			X			X
Observations	13,672	13,672	13,672	13,672	13,672	13,672	13,672	13,672	13,672

Note. This table reports OLS (Panel A) and IV estimates (Panel B) of log monthly expenditure for three different categories of consumption (food and clothes, columns 1-3; housing, columns 4-6; other, column 7-9) on illegal residence status and other controls. The dummy *Illegal residence status* equals one if the respondent lacks legal status. Each row in Panel B reports IV estimates obtained using a different instrument (set of instruments): a) $\ln(\text{Rainfall}(T, T-1))$; b) $\ln(\text{Total Quotas})$; c) $\ln(\text{Rainfall}(T, T-1))$ and $\ln(\text{Total Quotas})$. All these instruments are defined as in the notes to Table 3 and Table 4. *Baseline, individual* and *HH controls* and *$\ln(\text{income})$* are defined as in the note to Table 3. Sample: immigrants with 1–10 years of residence in Italy. ISMU survey years: 2004–2007. Standard errors: robust and clustered by country of origin (103 clusters); *significant at 10%; **significant at 5%; ***significant at 1%.

Legal entries and documented migrants stock



Figure 1—Legal entries and documented migrants stock in Italy (years: 1997–2008)



Source. Authors's elaboration from Italian Ministry of Internal Affairs data.

Descriptive statistics

Variable		Documented immigrants		Undocumented immigrants		significant difference (5%)
		Mean	Std. Dev.	Mean	Std. Dev.	
Respondent:	Age	33.32	7.64	31.60	8.50	*
	Female	0.42	0.49	0.37	0.48	*
	Education: none	0.09	0.28	0.11	0.31	*
	primary/compulsory	0.37	0.48	0.35	0.48	
	secondary	0.40	0.49	0.42	0.49	
	tertiary	0.14	0.35	0.12	0.33	*
	Residence in Italy (years)	5.81	2.44	2.68	1.78	*
	Unemployed	0.04	0.20	0.08	0.27	*
	Single	0.33	0.47	0.55	0.50	*
	Children (dummy)	0.58	0.49	0.45	0.50	*
Household:	# household members in Italy	2.43	1.47	1.45	0.95	*
	spouse abroad (if married)	0.25	0.44	0.59	0.49	*
	children abroad (if children>0)	0.38	0.49	0.84	0.37	*
	# children abroad (if any)	1.82	0.98	1.92	1.02	*
	children in Italy (if children>0)	0.68	0.47	0.23	0.42	*
	# children in Italy (if any)	1.76	0.89	1.40	0.64	*
	living in own house in Italy	0.15	0.36	0.01	0.11	*
Area of origin:	Subsaharan Africa	0.17	0.38	0.16	0.36	
	East Asia (and Pacific)	0.06	0.23	0.03	0.18	*
	Eastern Europe & Central Asia	0.27	0.44	0.36	0.48	*
	Latin America	0.15	0.35	0.19	0.39	*
	Middle East & North Africa	0.23	0.42	0.20	0.40	*
	South Asia	0.12	0.33	0.06	0.24	*
Observations		11865		1807		
Total observations				13672		
Undocumented immigrants share				0.13		

Note. * denotes a difference between documented and undocumented immigrants that is significant at least at 5% level. Sample: immigrants with 1–10 years of residence in Italy. ISMU survey years: 2004–2007.

Descriptive statistics (2)

		Documented immigrants		Undocumented immigrants		significant difference (5%)
		Mean	Std. Dev.	Mean	Std. Dev.	
Monthly values (euros)						
Consumption:	Total	581.6	267.2	424.6	243.8	*
	Food & Clothes	220.0	116.9	177.5	113.1	*
	Housing	244.4	156.3	151.5	137.1	*
	Other	117.3	95.5	95.6	82.5	*
Total Income		815.5	376.9	710.4	339.2	*
Share of total income						
Consumption:	Total	0.74	0.22	0.65	0.27	*
Share of total consumption						
	Food & Clothes	0.39	0.14	0.44	0.20	*
	Housing	0.40	0.17	0.31	0.21	*
	Other	0.21	0.13	0.25	0.21	*
Observations		11,865		1,807		

Note. * denotes a difference between documented and undocumented immigrants that is significant at least at 5% level. Sample: immigrants with 1–10 years of residence in Italy. ISMU survey years: 2004–2007.



Clicking on Heaven's Door: The Effect of Immigrant Legalization on Crime

Paolo Pinotti

Forthcoming, American Economic Review

Introduction



- Estimate of the effects of **immigrant legalization** on the **crime rate of immigrants in Italy**
- **Background information on acquiring legal status in Italy:**
 - A possibility to acquire legal status is **work-related residence permits sponsored by the immigrant's employer**.
 - A large part of immigrants who enter Italy irregularly, start working unofficially for an employer, and hope that the employer will subsequently sponsor them for a residence permit.

Methodology

- **Regression discontinuity design:**
 - Fixed quotas of residence permits are available each year
 - Applications must be submitted electronically on specific “**Click Days**” and are processed on a **first-come, first-served basis** until the available quotas are exhausted
 - Several thousand applicants are denied legal status every year simply because their employers applied a few minutes (or seconds) after the cutoff time

Methodology (2)

- **Regression discontinuity design:**
 - For most groups of applicants, this cutoff occurred fewer than 30 minutes after the start of the Click Day.
 - The exact timing of the cutoff for each group was unknown *ex ante*: it depended on the timing of all applications as well as on how many applications were rejected for being inaccurate, false, or incomplete.

Data

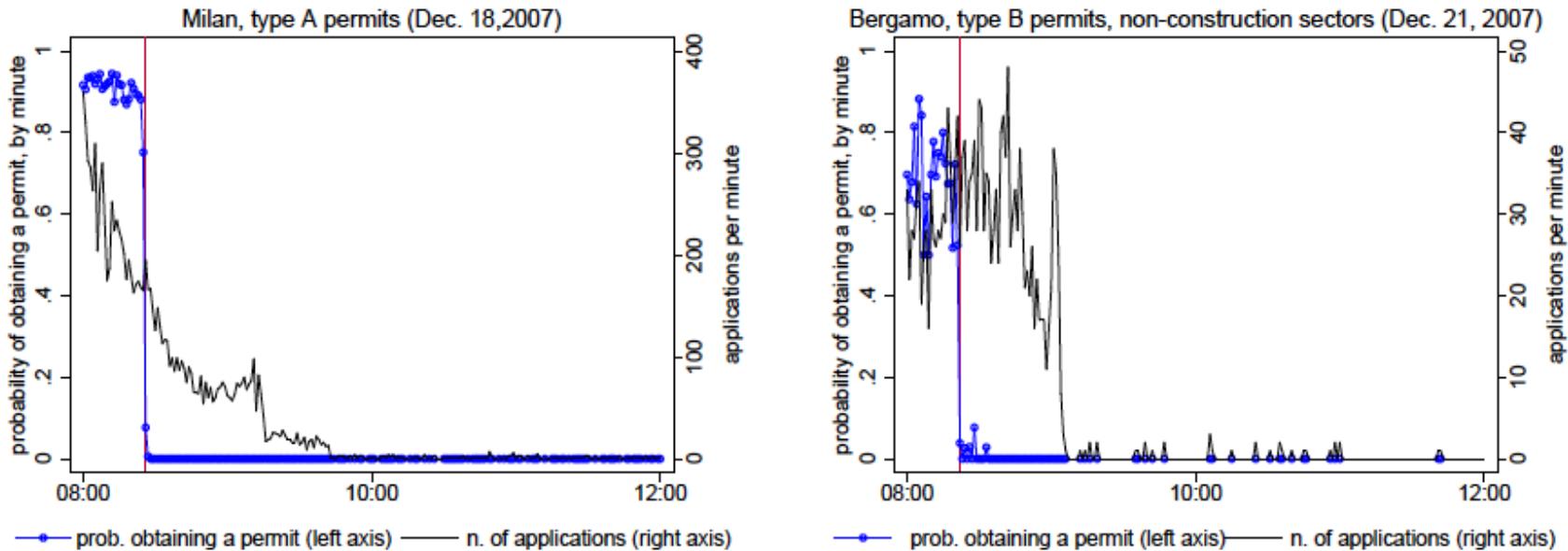
- Data on the universe of **applicants** sponsored on the **Click Days in December 2007** for receiving a permit in 2008. The total number of applicants reached 610 thousand.
- Data on applications are matched at the individual level with the **criminal records** of all offenders reported by the police for having committed (at least) one serious crime (**robbery, theft, drug trafficking, smuggling, extortion, kidnapping, murder, and rape**) in 2007 or 2008 (i.e., the year before and the year after the 2007 Click Days)



Types of residence permits

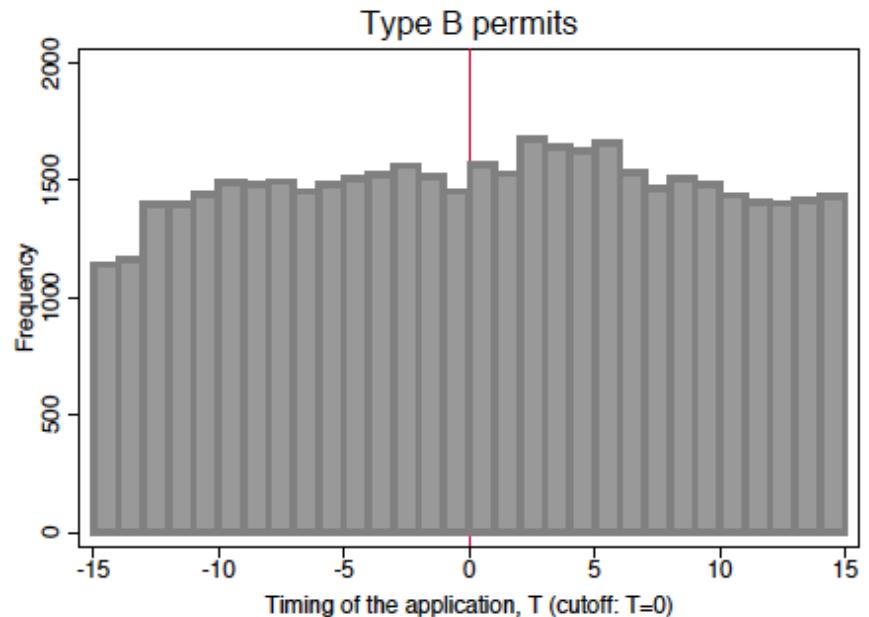
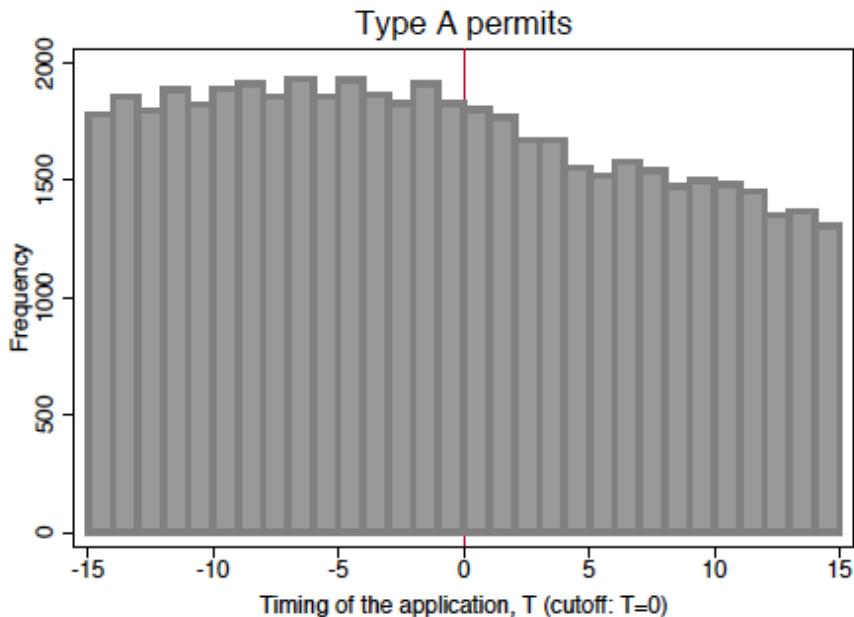
- Immigrants can be sponsored for two main types of permits:
 - **Type-A permits** for domestic and care workers employed by individuals and families
 - **Type-B permits** for firm employees, further distinguished into B1 and B2 permits for construction or non-construction workers, respectively.

Timing of applications and probability of obtaining a residence permit



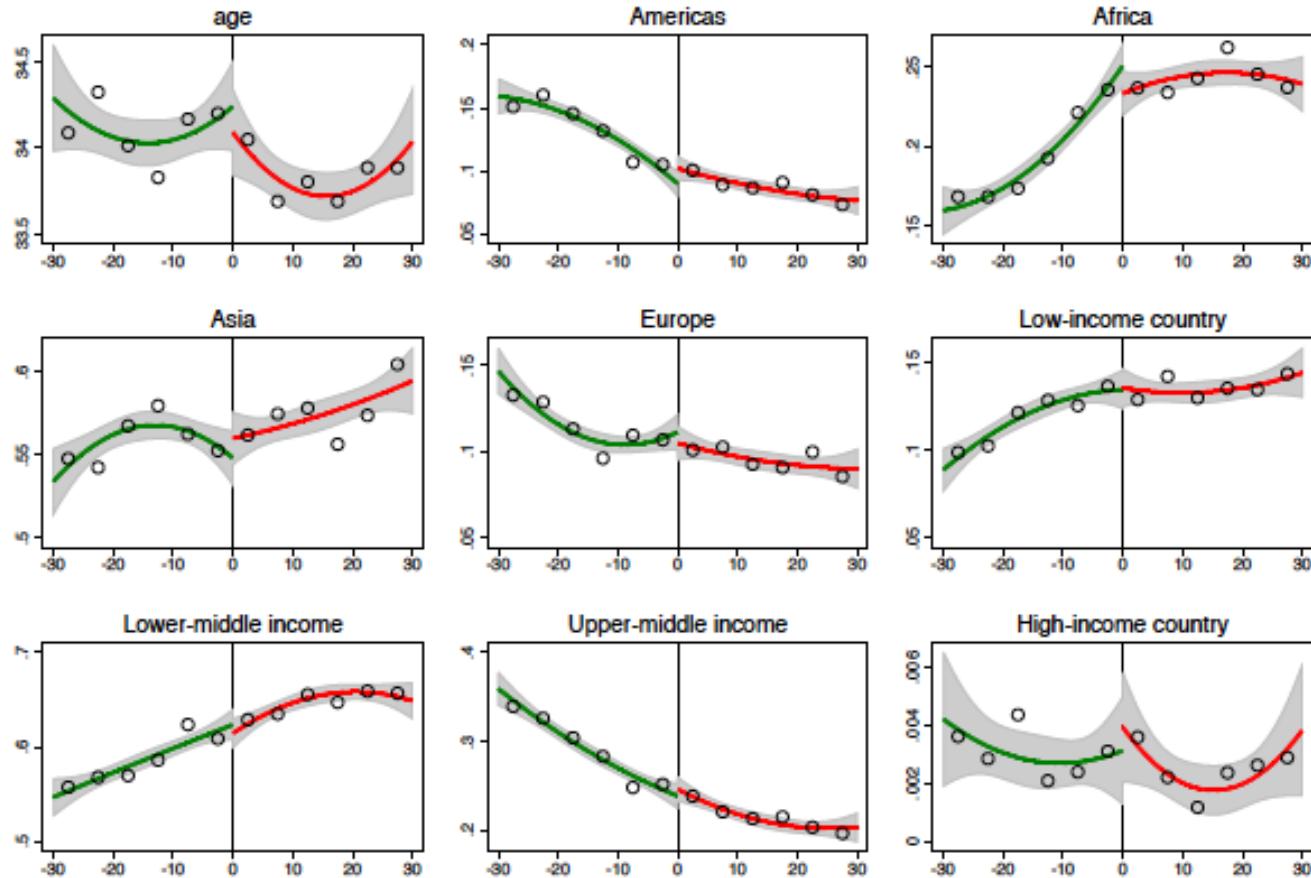
Note: These graphs show the total number of applications received (black line, right axis) and the fraction of those that were eventually accepted (blue line, left axis) at each minute in time between 8:00 and 12:00 for the case of two lotteries: type-A applicants in Milan and type-B applicants (non-construction, B1) in Bergamo. The vertical lines in each graph show the timing of the structural breakpoint in the probability of acceptance, according to the Andrews (1993) test.

Frequency of applications near the cutoff



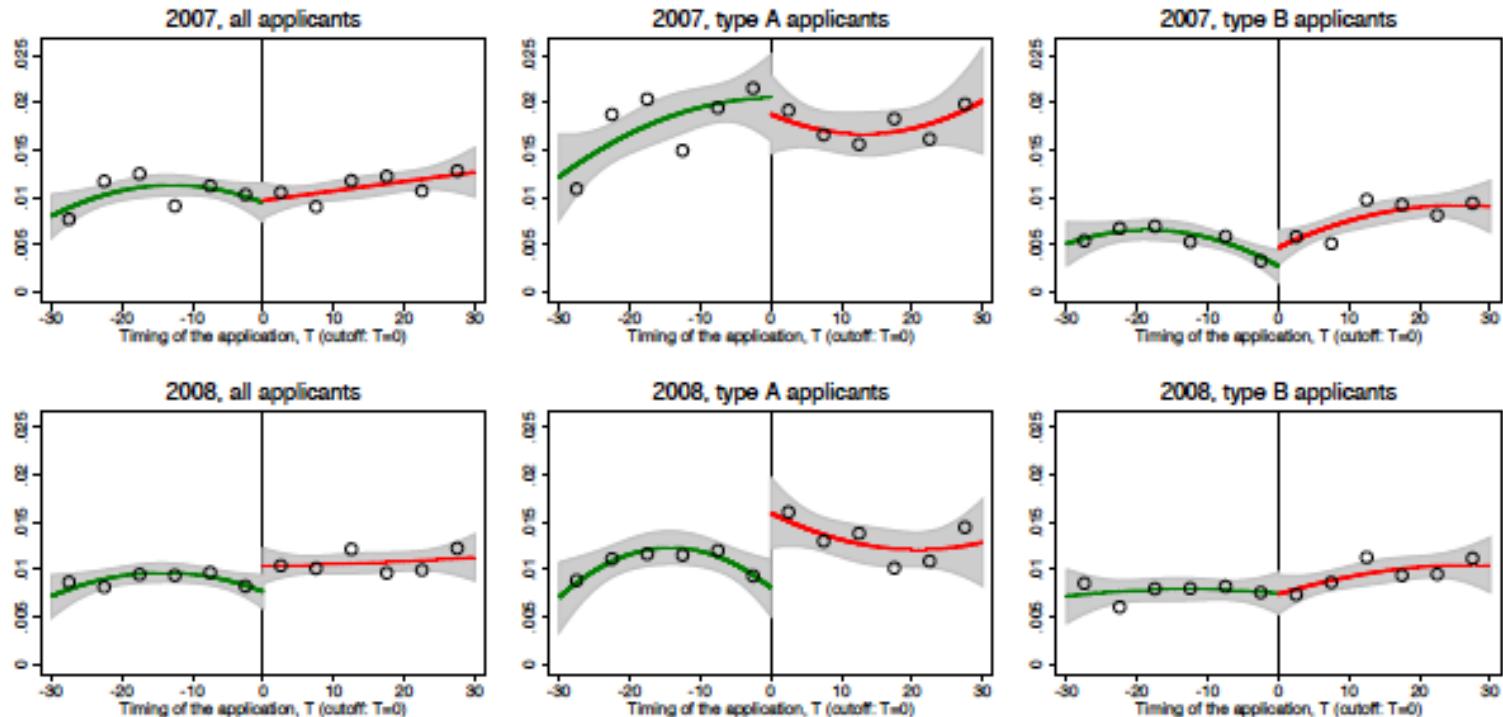
Note: The graph shows the density of applications received by type-A and type-B applicants at each minute in time before and after the cutoff.

Individual characteristics and timing of the application



Note: These graphs show the average individual characteristics of type-A applicants, conditional on the timing of application. The scatterplots are averages within five-minute bins, and the solid lines and shaded areas are the predicted values and associated confidence intervals, respectively, based on a quadratic polynomial regression.

Reduced-form plots



Note: The graphs show the average number of crimes per applicant committed before and after Click Days – conditional on the timing of application. The scatterplots are averages within five-minute bins, and the solid line and shaded areas are the predicted outcomes and associated confidence intervals, respectively, based on quadratic polynomial regressions. The graphs in the top and bottom row show the crime rates in the year before and after Click Days (2007 and 2008, respectively). The left, middle, and right column show the results for all applicants, the subsample of type A applicants, and the subsample of type B applicants, respectively.

Legal status and crime: Global polynomial regressions



	(1)	(2)	(3)	(4)	(5)	(6)
	Baseline estimates			Lottery FE, clustered s.e.		
	all	type A	type B	all	type A	type B
Year 2008:						
Reduced form	-0.003*	-0.008***	0.000	-0.003	-0.008**	0.000
	(0.002)	(0.003)	(0.002)	(0.002)	(0.003)	(0.002)
2SLS effect of legal status	-0.006*	-0.013***	0.000	-0.006	-0.013**	0.001
	(0.003)	(0.005)	(0.005)	(0.004)	(0.005)	(0.005)
First stage coeff. of Z	0.459***	0.610***	0.372***	0.463***	0.621***	0.374***
	(0.007)	(0.010)	(0.009)	(0.029)	(0.032)	(0.041)
first stage F-stat. for Z	4304.08	3492.26	1645.23	248.73	369.39	83.15
Year 2007:						
Reduced form	-0.000	0.002	-0.002	-0.000	0.002	-0.002
	(0.002)	(0.004)	(0.001)	(0.002)	(0.004)	(0.002)
$E(C Z = 0)$	[0.011]	[0.018]	[0.008]	[0.011]	[0.018]	[0.008]
Observations	110,337	40,451	69,886	110,337	40,451	69,886

Note: This table shows parametric estimates of the effect of legal status on the crime rate of Click Day applicants. The top panel reports reduced form and 2SLS regressions of a dummy $C = 1$ for applicants that committed at least one serious offense in 2008 on a dummy $L = 1$ for applicants obtaining legal status on Click Days. The first stage instrument is a dummy $Z = 1$ for having applied before the cutoff time, the first stage coefficient of Z and the F-statistic for the excluded instrument (adjusted for heteroskedastic and clustered standard errors) are also reported. The bottom panel of the table reports the reduced form regression of a dummy for committing at least one serious offense in the year before Click Days (2007) on the instrument Z and (in square brackets) the baseline crime rate in such year, $E(C|Z = 0)$. All regressions control for a quadratic polynomial in the time elapsed since the cutoff (by the millisecond) and its interaction with Z , and the specifications in columns (4)-(6) also include lottery-fixed effects and a quadratic polynomial in age. Robust standard errors are reported in parentheses, and they are clustered by lottery in columns (4)-(6). *, **, and *** denote statistical significance at the 90%, 95%, and 99% confidence levels, respectively.

Legal status and crime: Local linear regressions



	(1)	(2)	(3)	(4)	(5)	(6)
	Bandwidth: IK2012			Bandwidth: CCT2014		
	all	type A	type B	all	type A	type B
Year 2008:						
Reduced form	-0.004** (0.002)	-0.008** (0.003)	0.001 (0.002)	-0.005** (0.002)	-0.009** (0.003)	0.002 (0.002)
2SLS effect of legal status	-0.008** (0.004)	-0.013** (0.005)	-0.003 (0.006)	-0.010** (0.005)	-0.014** (0.006)	-0.005 (0.007)
size of bandwidth	9:26	10:52	9:55	6:37	8:17	7:57
obs. inside bandwidth	41,743	17,378	27,801	29,805	13,454	22,532
Year 2007:						
Reduced form	-0.001 (0.002)	-0.003 (0.003)	0.002 (0.002)	-0.001 (0.002)	-0.004 (0.004)	0.001 (0.002)
size of bandwidth	10:20	25:43	11:15	7:27	9:51	7:31
obs. inside bandwidth	45,472	36,362	31,160	33,310	15,905	21,371
Observations	110,337	40,451	69,886	110,337	40,451	69,886

Note: This table shows non-parametric estimates of the effect of legal status on the crime rate of Click Day applicants. The top panel reports reduced form and 2SLS regressions of a dummy $C = 1$ for applicants that committed at least one serious offense in 2008 on a dummy $L = 1$ for applicants obtaining legal status on Click Days. The first stage instrument is a dummy $Z = 1$ for having applied before the cutoff time. The bottom panel of the table reports the reduced form regression of a dummy for committing at least one serious offense in the year before Click Days (2007) on the instrument Z . All coefficients are estimated using a kernel local linear regression in a symmetric bandwidth around the cutoff. The table reports the bandwidth selection rule, IK2012 or CCT2014, the size of the bandwidth (expressed in [minutes:seconds]), and the number of observations included in the bandwidth. Robust standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 90%, 95%, and 99% confidence levels, respectively.

Summary of results

- Main result:
 - Legalization reduces the crime rate of legalized immigrants by 0.6 percentage points on average, on a baseline crime rate of 1.1 percent.
 - The result is driven by a reduction in the number of economically-motivated crimes committed by applicants sponsored by other immigrants as domestic workers (e.g., housekeepers, elder care givers, and so on) – some of them being fictitious job offers
 - The effect is zero for firm-sponsored employees