Immigration, Population Diversity and Innovation of Italian regions

Massimiliano Bratti[1], Chiara Conti[2]

3rd SEEK Conference - Engines for More and Better Jobs in Europe
Centre for European Economic Research (ZEW) – Mannheim, Germany
April 25-26, 2013

[1] **DEMM (University of Milan), CHILD (Turin), IZA (Bonn), Ld'A (Milan)**
[2] **DEMM (University of Milan)**

Motivation

Immigration has been recently at the centre of political and economical debate, given the size of the phenomenom and its recent fast growth

Potential impact of immigration has been studied on:

- Natives' wages (Borjas 2003;2005, Ottaviano and Peri 2012) and employment opportunities (Pischke and Velling 1997, Card 2001;2005)
- Firm productivity (Peri 2012)
- Trade creation (Gould 1994, Rauch and Trinidade 2002, Peri and Requena-Silvente 2010)
- Crime (Bell et al. 2010, Bianchi et al. 2012)

Motivation

Effect of immigration on innovation much less studied, though innovation is a key factor for economic growth

Recently, an increasing number of studies have focused on:

- Impact of immigrant's relative population size on host country's innovation (proxied by patents' applications)
- Effect of cultural diversity (complement of concentration index on the shares of 'ethnic groups') on productivity and innovation

→ mostly limited to **skilled** immigration in the US

Literature Review

- Effect of culturally diverse population (proxied by 'fractionalization index') on **productivity,** through the effect on natives' (i) wages and (ii) rental prices [Ottaviano and Peri (2006, US cities), Prarolo et al. (2009, EU regions)] or employment [Suedekum et al., 2009, German regions]

- Effect of **high skilled** immigrants and foreign graduate students on **US patents' applications** [Chelleraj et al. (2008), Hunt and Gauthier-Loiselle (2010), Kerr and Lincoln (2010)]

- Contribution of **foreign doctoral students** to **academic innovation**, measured by publications and citations [Stuen et al., 2012, American S&E Departments]

Literature Review

- Effect of the share of immigrants employed in **top skilled occupations** on **patenting and scientific publications** [Bosetti et al., 2012, EU countries]
- Cultural **diversity of R&D workforce** and **patenting** in German regions [Niebuhr, 2009]

→ **Agreed on beneficial effect of immigration on innovation and productivity**

- Ozgen C. (2012): effect of overall immigration and separate effects **by skill level** on innovativeness of EU regions → effects of high skilled and low skilled immigrants might work in **opposite directions**.

Our contribution

- Effect of immigration on Italian provinces' innovativeness (patents' applications)

  → country exposed to very fast and large wave of immigrations during the 2000s

  → very small geographical scale of analysis (NUTS-3)

- Focus on general impact of immigration

  → two different indicators: **share of migrants** and **population diversity**

- Separate the effects of **low-skilled** and **high-skilled** immigrants on innovation

Our contribution

The Italian context

- Immigrants tend to take **manual-intensive and routine-type occupations** (89.9% of foreign workers are blue collar workers)
- Most of foreign population: low schooling levels
- High educated immigrants often takes low skilled jobs
- *"Job Ethnicization":* for low-skilled jobs firms prefer to employ immigrants (foreigners are 9% of total workforce, but they are more than 80% of agricultural workers, 40% of workers in low skilled personal services, 18% of workers in construction sector, 33.5% of blue collars)

Conceptual framework

Why immigration may have effect on innovation?

1) “Mechanical” effects

Inflow of foreign population into a region produces changes in variables that are powerful predictors of innovation:

- Size of population → agglomeration and mkt size effects
- Avg skill level of population → human capital
- Age structure of population → younger individuals more creative

→ *Mediating ‘mechanical’ factors*

Conceptual framework

2) Other effects ("non-mechanical")

2.a Inflow of immigrants may affect firms' choices concerning technology adoption and investment in physical capital (long run)

→ *Mediating factors which requires economic agents to change their behavior*

2.b Immigration produces a more culturally diverse population

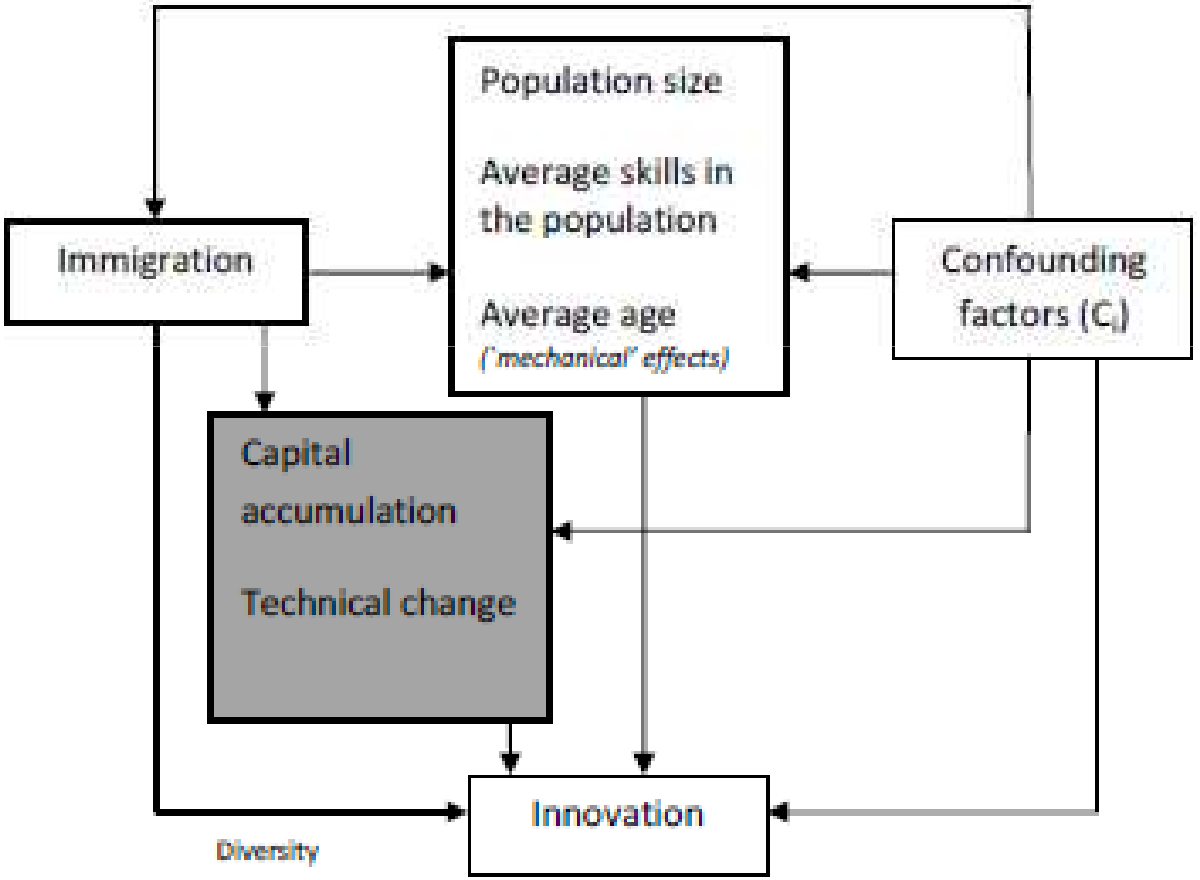
→ Skill complementarities and different backgrounds positively influence the production of new ideas (Hs)

→ Difficulties in communication, increased transaction costs, reduction of social capital can act as obstacle to innovation (Ls)

Mediating factors may be affected by *confounding factors* (if they depend on a 'third variable' which is also a determinant of immigration) or proxy them

Conceptual framework

Effect of immigrants on innovation



Conceptual framework

If immigration is endogenouos → mediating factors are endogenouos as well

Include mediating factors in the regression:

- need to be instrumented
- estimate the direct effect (diversity)

In Niebuhr (2009) and Ozgen (2012) population and measures of HC are included in the regression as **exogenous contemporaneouos variables**

Conceptual framework

If mediating factors are not included in the regression and they are also confounding factors:

- omitted variable bias
- even with consistent estimates → effects of immigration and mediating factors cannot be separately identified

Solution: include values of mediating factors in pre-estimation period (2001)

- control for very time-persistent confounding factors
- avoid to include variables affected by immigration during estimation period (endogenous mediating factors)

Data description

- Panel of 103 Italian provinces (NUTS-3) during the period 2002-2008 → *short-medium-run effects*
- Dependent variable: log of patents' application per 1000 inhab. (proxy for innovation)
- Number of **patent's applications to EPO by priority year** (closest to the actual timing of innovation): *EUROSTAT database-Regional Science&Technology statistics*
- Patents' applications assigned according to inventor's province of residence (fractional count)
- Share of immigrants and population diversity: *Demographic portal of ISTAT* (stock of legal immigrants from 195 countries of origin residing in each province)
- Distribution of immigrants by country of origin across provinces in 1995 (to build the instruments): **foreign residence permits** (Italian Ministry of Interior)

Data description

Time-varying control variables

- Shares of VA accounted for by agricultural, services and construction (industrial structure – very time persistent): *ISTAT-Sistemi Indicatori Territoriali* (share of manufacturing is omitted – multicollinearity)
- R&D expenditure (%GDP); only available at NUTS-2 level → assign to each province the value of the corresponding region

Time-invariant control variables (2001 values):

- Resident population (ISTAT)
- Share of working-age population (ISTAT)
- College share (n.of graduates from *Census 2001*)

Data description

- Patents: imperfect indicator of overall innovative activity (Griliches 1990)
  - → not all innovation are patented (partial indicator)
  - → not all patented innovation have the same level of quality
  - → propensity to patent changes across areas, sectors and time
- no generally recognized methods for measuring the value of patents
- applications to EPO: homogeneous, high quality patents (applying is difficult, time consuming and expensive (Moreno et al. 2005))
- controlling for the industrial structure: account for different propensity to patent across sectors
- Still the best available, most reliable measure of innovation output, commonly used in empirical research

Empirical strategy

Linear specification:

$$\ln PATN_{ijt} = \alpha_0 + \delta_t + \delta_j + \alpha_1 MIGsh_{it-1} + \alpha_2 \mathbf{X}_{it-1} + \alpha_3 X_{jt-1} + \alpha_4 \mathbf{D}_{i2001} + \varepsilon_{ijt}$$

i, j, t: province, region and time subscripts

***MIGsh*$_{it-1}$:** lagged share of immigrants (as a whole) on population

Alternative measure: *Ethnolinguistic Fractionalization Index* (ELF)

$$POPdiv_{it} = 1 - \sum_{g=1}^{Git} \left(\frac{P_{git}}{P_{it}} \right)^2$$

P_{git} is the population of ethnic group *g* in province *i* and year *t*

→ Probability that 2 randomly drawn individuals in the population will not belong to the same ethnic group (higher values → more diversity)

Endogeneity and identification: 2SLS

- Instruments for $MIGsh_{it-1}$: *shift-share* procedure (Card 2001)
- Intensively employed in empirical literature on immigration (Peri, 2012, Hunt and Gauthier-Loiselle, 2010, Bosetti et al., 2012)
- Based on the idea that immigrants tend to settle where individuals of the same nationality already live (*immigrants' enclaves*)
- Impute the yearly stock of immigrants by nationality in Italy as a whole (M_{gt}) to provinces according to the distribution of nationalities across provinces in 1995 (θ_{i1995})
- Predicted stock of immigrants by nationality:

$$\hat{M}_{git} = \theta_{i1995} M_{gt}$$

- Aggregate across nationalities at province level and divide by the predicted province's population (according to 1995 distribution)
- → Predicted share of immigrants in province *i* at time *t*

Endogeneity and identification: 2SLS

- Same procedure to compute the 'predicted population diversity'

$$\hat{POPdiv}_{it} = 1 - \sum_{g=1}^{Gi1995} \left(\frac{\hat{P}_{git}}{\hat{P}_{it}} \right)^2$$

- Instruments are based on 2 components:

  1) The total stock of individuals by nationality in Italy → uncorrelated with single provinces' supply and demand shocks impatting on local innovation

  2) The distribution of immigrants and of the total population in 1995 → should be uncorrelated with unobserved factors affecting patenting more than 7 years later, conditional on the observables included in the regression (industrial structure and proxies of agglomeration economies)

Table 1: OLS estimates of the effect of immigrants on patents' applications

	(1)	(2)	(3)	*benchmark* (4)	(5)
share of immigrants	0.364***	0.107***	0.093***	-0.017	
	(0.024)	(0.020)	(0.021)	(0.019)	
population diversity (ELF)					-0.933
					(1.020)
RD expenditures (% GDP)[a]			1.071***	1.034***	1.033***
			(0.397)	(0.391)	(0.390)
share VA agriculture			-0.119***	-0.032**	-0.032**
			(0.014)	(0.015)	(0.015)
share VA services			-0.021***	-0.064***	-0.064***
			(0.004)	(0.006)	(0.006)
share VA construction			-0.126***	-0.021	-0.021
			(0.028)	(0.025)	(0.025)
log pop 2001				0.277***	0.277***
				(0.051)	(0.051)
active age pop share 2001				0.056**	0.056**
				(0.024)	(0.024)
% of graduates on pop 18-64				0.191***	0.191***
				(0.019)	(0.019)
Year fixed effects	No	Yes	Yes	Yes	Yes
Region (NUTS-2) fixed effects	No	Yes	Yes	Yes	Yes
N. observations	607	607	607	607	607
R-squared	.46	.76	.80	.85	.85

*** significant at 1%; ** significant at 5%; * significant at 10%.

Note. The dependent variable is log patents' applications per 1000 inhabitants at the province (NUTS-3) level for Italy, 2003-2008. When not differently specified all independent variables are lagged one year. Standard errors are clustered at the *region* × *year* level because of the inclusion of an 'aggregated' variable (Moulton 1990) and robust to heteroskedasticity. Diversity of immigrants is measured using the ELF index (Mauro 1995).
[a] only available at the NUTS-2 level.

Table 2: 2SLS estimates of the effect of immigrants on patents' applications

	(1)	(2)
Second stage		
share of immigrants	-0.064**	
	(0.031)	
population diversity (ELF)		-3.457**
		(1.693)
RD expenditures (% GDP)[a]	0.944**	0.942**
	(0.378)	(0.378)
share VA agriculture	-0.025*	-0.025*
	(0.014)	(0.014)
share VA services	-0.069***	-0.069***
	(0.007)	(0.007)
share VA constructions	-0.022	-0.022
	(0.025)	(0.025)
log pop 2001	0.316***	0.317***
	(0.050)	(0.050)
active age pop share 2001	0.055**	0.055**
	(0.024)	(0.024)
% of graduates on pop 18-64	0.199***	0.199***
	(0.019)	(0.019)
First stage		
predicted share of immigrants	0.375***	
	(0.028)	
predicted population diversity		0.374***
		(0.029)
F-test excluded instrument	181.76	170.56
N. observations	607	607
R-squared	.37	.37

*** significant at 1%; ** significant at 5%; * significant at 10%.

Note. The dependent variable is log patents' applications per 1000 inhabitants at the province (NUTS-3) level for Italy, 2003-2008. When not differently specified all independent variables are lagged one year. All models include year and region (NUTS-2) fixed effects. Standard errors are clustered at the *region* × *year* level because of the inclusion of an 'aggregated' variable (Moulton 1990) and are robust to heteroskedasticity. Diversity of immigrants is measured using the ELF index (Mauro 1995).

[a] only available at the NUTS-2 level.

Analysis by immigrants' skill level

- **2SLS: negative effect of overall immigration on patents' applications** (in contrast with existing literature focused on HS)
- Likely to be the result of the features of Italian immigration

→ **split immigrants in HS and LS**

- *Dataset Docquier-Marfoukof*: number of emigrants to Italy from 195 countries, divided in low, medium and high skilled (1991 and 2001)
- Compute for each country the share of medium-high skilled emigrants on total emigrants to Italy
- A) High skilled: all immigrants from a country for which the above share is larger than the median value (0.56)
- B) Total n. of immigrants from a country is splitted according to the share of medium-high skilled emigrants on total emigrants to Italy
- Compute the share of low skilled and high skilled immigrants in each province and the corresponding instrument (*shift-share* procedure)

Analysis by immigrants' skill level

Results:

- the coefficient of the share of low skilled is negative and significant (1% with procedure A, 5% with procedure B)
- an increase of the share of low skilled of 1 p.p. induces a reduction in patents' application per 1000 inhabitants in a range btw 0.09% (procedure A) and 0.19% (procedure B)
- the coefficient of the share of high skilled is positive but not significant → not precisely estimated in our sample

→ Consistent with the empirical literature on HS immigration

→ In line with those of Ozgen et al. (2012) and Suedekum et al. (2009): effects of HS and LS immigrants works in opposite direction

Table 3: OLS and IV estimates by skill level (1)

		2SLS		
	OLS	1st stage: high skilled	1st stage: low skilled	2nd stage
share of immigrants: high skilled[b]	-0.079			0.025
	(0.055)			(0.102)
share of immigrants: low skilled	-0.001			-0.094***
	(0.023)			(0.036)
RD expenditures (% GDP)[a]	0.998**	-0.355	-0.908*	0.988**
	(0.384)	(0.314)	(0.546)	(0.391)
share VA agriculture	-0.027*	0.061***	0.037*	-0.031**
	(0.016)	(0.010)	(0.022)	(0.015)
share VA services	-0.061***	0.013***	-0.081***	-.073***
	(0.007)	(0.005)	(0.011)	(0.008)
share VA construction	-0.015	0.044**	-0.093*	-0.031
	(0.026)	(0.018)	(0.048)	(0.026)
log population (2001)	0.281***	0.099**	0.193**	0.313***
	(0.052)	(0.039)	(0.095)	(0.051)
15-65 population share (2001)	0.058**	0.038***	0.142***	0.052**
	(0.025)	(0.014)	(0.047)	(0.023)
% of graduates on pop 18-64 (2001)	0.194***	0.022	-0.045**	0.195***
	(0.020)	(0.013)	(0.020)	(0.020)
predicted share of immigrants: high skilled[c]		0.337***	-0.213***	
		(0.032)	(0.049)	
predicted share of immigrants: low skilled		-0.027	0.528***	
		(0.020)	(0.028)	
Year fixed effects	yes	yes	yes	yes
Region (NUTS-2) fixed effects	yes	yes	yes	yes
F-test (1st stage)		57.78	176.88	
N. obs.	607	607	607	607
R^2	0.846	0.409	0.467	0.364

*** significant at 1%; ** significant at 5%; * significant at 10%.

Note. The dependent variable is log patents' applications per 1000 inhabitants at the province (NUTS-3) level for Italy, 2003-2008. When not differently specified all independent variables are lagged one year. All models include year and region (NUTS-2) fixed effects. Standard errors are clustered at the *region* × *year* level because of the inclusion of an 'aggregated' variable (Moulton 1990) and are robust to heteroskedasticity.

[a] only available at the NUTS-2 level.

[b] the skill level is assigned to immigrants according to their country of origin: all immigrants from a country whose share of high-medium skilled emigrants on total emigrants to Italy in 2001 (Docquier-Marfoukof database) is larger than 0.56 (median value) are considered high skilled.

[c] for the construction of the instruments the year of reference is 1991 (Docquier-Marfoukof database).

Table 4: OLS and IV estimates by skill level (2)

		2SLS		
	OLS	1st stage: high skilled	1st stage: low skilled	2nd stage
share of immigrants: high skilled[b]	0.001			0.113
	(0.083)			(0.154)
share of immigrants: low skilled	-0.029			-0.186**
	(0.053)			(0.091)
RD expenditures (% GDP)[a]	1.041***	-0.617*	-0.688	1.010**
	(0.389)	(0.356)	(0.433)	(0.394)
share VA agriculture	-0.033**	0.062***	0.027	-0.033**
	(0.016)	(0.013)	(0.017)	(0.015)
share VA services	-0.064***	-0.017***	-0.049***	-.073***
	(0.007)	(0.006)	(0.008)	(0.007)
share VA construction	0.002	-0.008	-0.053	-0.027
	(0.026)	(0.024)	(0.034)	(0.026)
log population (2001)	0.275***	0.159***	0.105	0.304***
	(0.052)	(0.041)	(0.066)	(0.053)
15-65 population share (2001)	0.056**	0.069***	0.104***	0.053**
	(0.024)	(0.016)	(0.032)	(0.024)
% of graduates on pop 18-64 (2001)	0.190***	-0.016	-0.020	0.190***
	(0.020)	(0.014)	(0.014)	(0.020)
predicted share of immigrants: high skilled[c]		0.163***	-0.401***	
		(0.055)	(0.063)	
predicted share of immigrants: low skilled		0.145***	0.742***	
		(0.044)	(0.051)	
Year fixed effects	yes	yes	yes	yes
Region (NUTS-2) fixed effects	yes	yes	yes	yes
F-test (1st stage)		70.63	165.81	
N. obs.	607	607	607	607
R^2	0.846	0.429	0.462	0.368

*** significant at 1%; ** significant at 5%; * significant at 10%.

Note. The dependent variable is log patents' applications per 1000 inhabitants at the province (NUTS-3) level for Italy, 2003-2008. When not differently specified all independent variables are lagged one year. All models include year and region (NUTS-2) fixed effects. Standard errors are clustered at the *region* × *year* level because of the inclusion of an 'aggregated' variable (Moulton 1990) and are robust to heteroskedasticity.

[a] only available at the NUTS-2 level.

[b] for each province, the total number of immigrants from a given country is splitted by skill level according to the shares of high-medium skilled and low skilled emigrants on total emigrants from that country to Italy in 2001 (Docquier-Marfoukof database).

[c] for the construction of the instruments the year of reference is 1991 (Docquier-Marfoukof database).

Conclusions

Looking for a causal relationship between the presence of immigrants and innovation (proxied by patents applications) in Italian provinces during the period 2002-2008, we found that:

- The overall stock of immigrants has a significant negative effect on Italian provinces' patenting activities
- This is mostly driven by (i) a negative effect of low educated immigrants and (ii) the prevalence in Italy of unskilled immigration
- Expected positive effects of complementary skills and different background may not emerge: (i) more than compensated by the negative effect of low skilled immigrants, (ii) under-utilization of human capital of high skilled immigrants
- Importance of immigration policies and correct functioning of labor mkt (negative effect of LS can intensify in the long run)