Emigration and Entrepreneurial Drain

Online Appendix

Massimo Anelli (Bocconi, CESifo, IZA)
Gaetano Basso (Bank of Italy)
Giuseppe Ippedico (UC Davis)
Giovanni Peri (UC Davis and NBER)

Contents

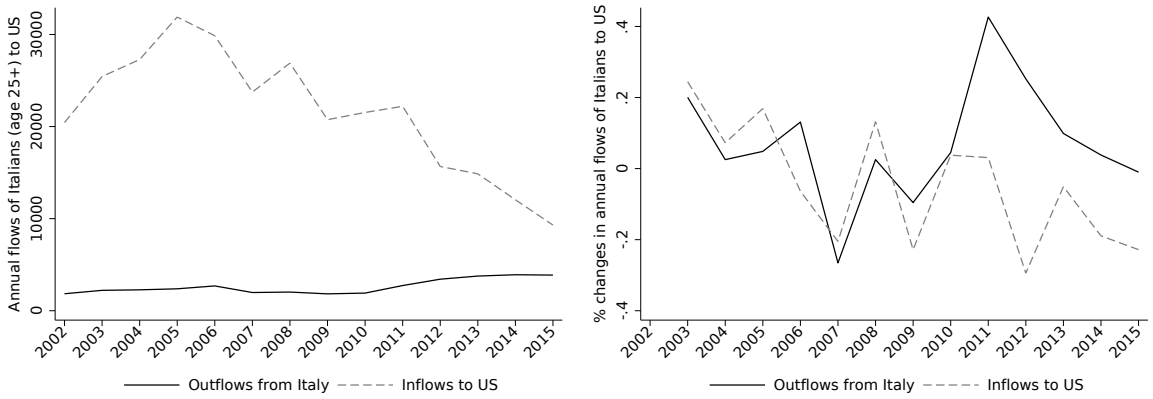
A.I Additional figures

In this section we present the additional figures discussed in the main text. Figure A1 shows that the outflows of Italians towards the US are underestimated if we compare Italian administrative and US Census Bureau American Community Survey (ACS) data (Ruggles et al. 2019; U.S. 2018). The year-to-year changes in the two datasets follow closely each other up to the most recent years (likely due to the inability of the ACS survey to capture recent immigration from Italy).

Figure A2 shows the pre-trends and the post-emigration wave change in the stock of firms in a reduced-form event-study graph. The estimates suggest that the largest decline in the number of firms occurs in the years after 2011, consistent with what we observe from the raw data in Figure 6.

Finally, Figure A3 shows how the correlation between the IV and the emigration rate, once we partial out the control variables (unemployment rate and value added per capita) and the 110 province FEs, is not driven by the presence of outliers.

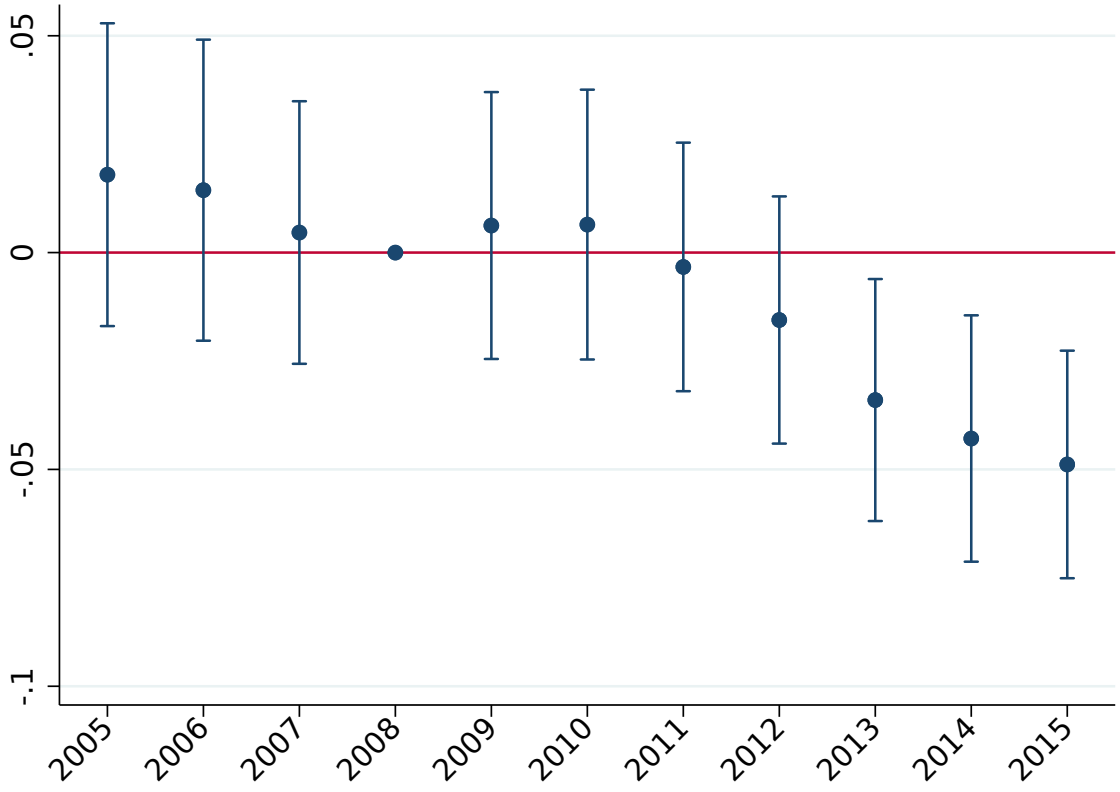
Figure A1: Recorded Emigration and Inflows of Italians to the US



(a) Annual emigration of Italians to US (b) % changes in annual flows of Italians to US

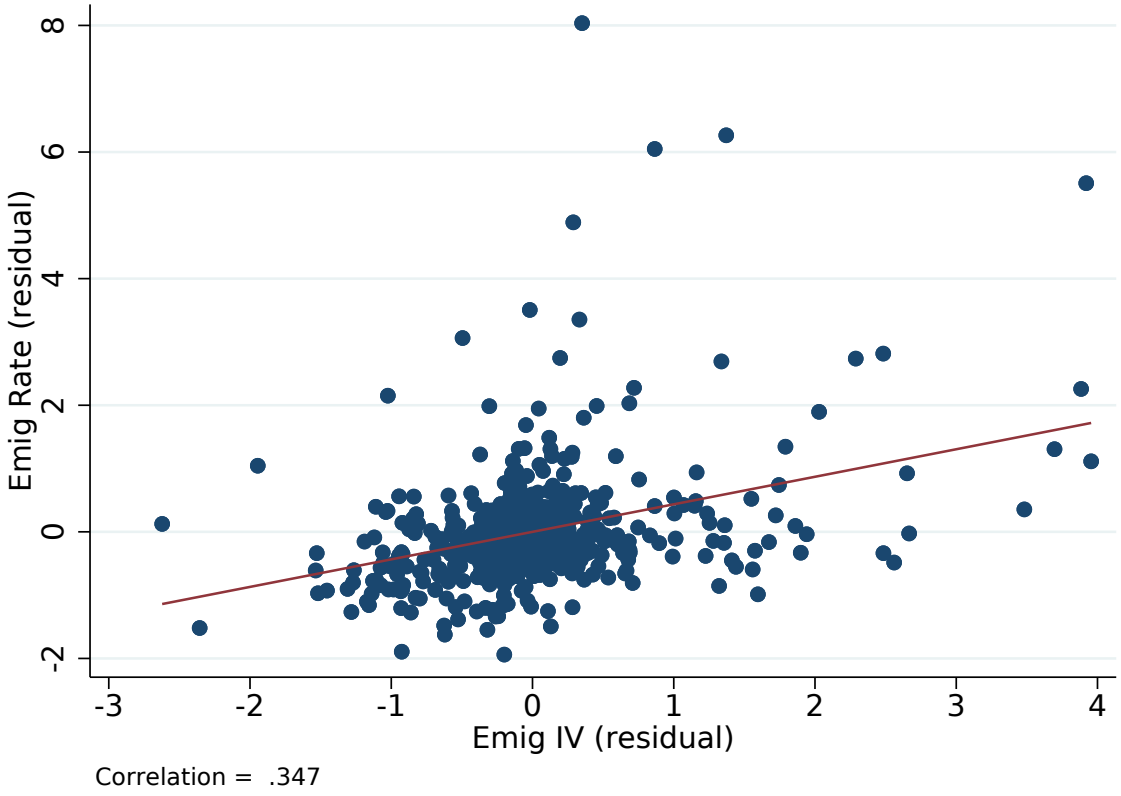
Notes: In Figure (a), the black solid line shows the annual outflows of Italians to the United States recorded in the AIRE-Istat data, while the grey dashed line shows the corresponding annual inflows of Italians to the US according to the American Community Survey (ACS) data. Figure (b) shows the percentage changes in the annual flows from the two data sources.

Figure A2: Event study: effect of the instrument interacted with years fixed effects on stock of firms, 2005-2015



Notes: The graph plots the coefficients γ_τ of the interaction between the instrument and year fixed effects from the regression: $y_{l,t} = \alpha + \beta Pull_l + \sum_{\tau \neq 2008} \gamma_\tau Pull_l * \mathbb{I}(\tau = t) + \xi X_{l,2005} + \phi_p + \lambda_t + \psi_{p,t} + \varepsilon_{l,t}$, where the outcome $y_{l,t}$ is the stock of firms in LLM l in year t as a fraction of LLM population 25-64 years old (average 2005-2008), $Pull_l$ is the predicted emigration rate based on the shares of pre-2000 emigrants to different countries relative to the LLM population in 2000 interacted with GDP growth of each country between 2008-2015, $X_{l,2005}$ include unemployment rate and value added per capita in 100,000 euros in 2005, ϕ_p are province fixed effects, λ_t are year FEs and $\psi_{p,t}$ are province-by-year FEs. Standard errors are clustered at the province-by-year level and bars show confidence intervals at the 5-percent level.

Figure A3: First stage scatter plot correlation



Notes: The scatter plot shows the correlation between the Pull IV and the Emigration Rate after partialling out the control variables (unemployment rate and value added per capita) as well as the 110 province FEs. Both variables are normalized to have mean zero and unit variance.

A.II Accounting for under-registration in AIRE-Istat emigration data

In this section we validate the 2.6 adjustment factor used in the empirical analysis. To circumvent the issue that not all Italian emigrants report their change of residence by registering in the AIRE, we compare yearly outflows of Italians recorded by Istat-AIRE to the yearly inflows of Italians to three among the top-5 destination countries of Italian emigrants, namely the UK, Switzerland and the United States. For the UK, we obtained administrative data from the UK Social Security Registry based on "National Insurance number allocations to adult overseas nationals entering the UK" (NINo 2018), which include all individuals applying to work in the UK or to claim any benefit or tax credit. For Switzerland, we use Federal Statistical Office (BFS 2018) administrative data based on the migration registry (PETRA-STATPOP), which include only permanent residents ("ständige Wohnbevölkerung"). For the US, we use weighted survey data from the American Community Survey (ACS) (Ruggles et al. 2019; U.S. 2018), using information on the year of arrival and country of birth.

Table A1 compares the emigration flows registered in the Italian AIRE-Istat data to the immigration flows registered by each foreign source respectively. The variable *Factor* shows the ratio between the immigration and emigration flow in each year. The data shows that emigration flows are systematically under-reported in the AIRE-Istat data, in almost every year and for all the three countries considered. In the last two columns, we construct a weighted average of the correction factors (weighted by the emigration flows). If we include the US (penultimate column), the average correction factor for the period 2009-15 is about 3.48. However, as the ACS data are survey-based and thus less reliable than the administrative sources from UK and Switzerland, in the last column we only consider the two latter countries, for which the average correction

factor ranges between 2.62 and 3.23 over the period 2009-15 and is 2.87 on average. Based on these results, we use the minimum value of the average correction factor, 2.6, to adjust upwards the emigration flows between 2009-15 throughout our empirical analysis.

Table A1: Correction factor based on destination country data

	United Kingdom			Swizterland			United States			Emig-weighted	Emig-weighted
Year	Emig	Immig	Factor	Emig	Immig	Factor	Emig	Immig	Factor	Avg Factor	Avg Factor - No US
2002	2400	7717	3.22	4587	5961	1.30	1846	20439	11.07	3.86	1.96
2003	2740	8122	2.96	6021	5820	0.97	2216	25435	11.48	3.59	1.59
2004	3097	8180	2.64	5068	5859	1.16	2272	27282	12.01	3.96	1.72
2005	4003	10361	2.59	4911	5622	1.14	2382	31892	13.39	4.24	1.79
2006	4561	11048	2.42	5271	5689	1.08	2694	29865	11.09	3.72	1.70
2007	5033	15735	3.13	3647	8540	2.34	1979	23746	12.00	4.51	2.80
2008	5474	16460	3.01	4165	10025	2.41	2029	26887	13.25	4.57	2.75
2009	**4981**	**16876**	**3.39**	**4097**	**8668**	**2.12**	**1835**	**20749**	**11.31**	**4.24**	**2.81**
2010	**5167**	**18461**	**3.57**	**4522**	**10226**	**2.26**	**1918**	**21532**	**11.23**	**4.33**	**2.96**
2011	**5317**	**24882**	**4.68**	**5669**	**10651**	**1.88**	**2736**	**22200**	**8.11**	**4.21**	**3.23**
2012	**7293**	**26599**	**3.65**	**8238**	**14098**	**1.71**	**3427**	**15668**	**4.57**	**2.97**	**2.62**
2013	**12756**	**44120**	**3.46**	**9663**	**17662**	**1.83**	**3766**	**14870**	**3.95**	**2.93**	**2.76**
2014	**13332**	**51210**	**3.84**	**10151**	**19006**	**1.87**	**3910**	**12055**	**3.08**	**3.00**	**2.99**
2015	**17248**	**58653**	**3.40**	**11227**	**18894**	**1.68**	**3871**	**9306**	**2.40**	**2.69**	**2.72**
								Average	**2009-15**	**3.48**	**2.87**

A.III Instrument validity: Additional checks

In this Section we report four figures that corroborate our identification strategy and inference. First, as reported in the main text, we cluster the standard errors at the province level: independence across labour markets is needed under an identifying assumption based on the exogeneity of the emigration networks and our identification is effectively within provinces (Adão, Kolesár, and Morales 2019; Goldsmith-Pinkham, Sorkin, and Swift 2020). To validate this choice, we follow similar exercises proposed by Adão, Kolesár, and Morales (2019) and implemented in Fouka, Mazumder, and Tabellini (2020), and perform two placebo exercises. We replace the shifters and the shares, respectively, with random numbers extracted from $N(0, 5)$. The two exercises confirm that the clustered standard errors are valid and, if anything, too conservative: in the case of shifters, only 0.4 percent wrongly reject the null hypothesis of $\beta = 0$ at the 10 percent level (Figure A4); in the case of shares randomization, none of the 500 replications is significantly different from zero (Figure A5).

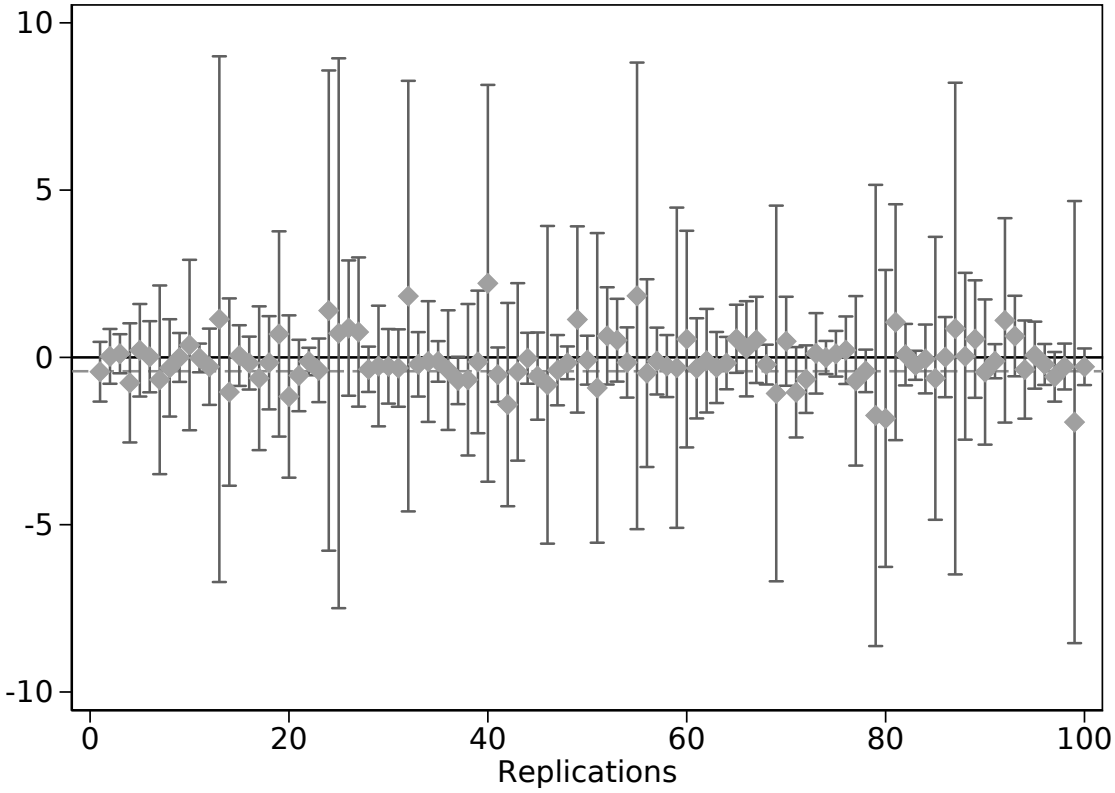
Then, we perform two additional exercises to validate the identification strategy. Differently from assigning a random shift as above, which effectively does not allow to identify any effect, here we first hold the emigration networks (the shares) fixed and we permute the GDP changes (the shifts); then, we do the opposite (hold the shares fixed and permute the shifters).[1] Considering that the country-of-emigration shares drive most of the identifying variation (Table 4), the random permutation of the shifts still allow us to identify our results (Figure A6). To the contrary, randomly permuting the shares while keeping the right shifts does not allow to identify any effect, consistent with our identifying assumption (Figure A7).

Finally, Tables A2, A3 and A4 show the pre-trends as discussed in Section II.B of

[1] In practice, we randomly reshuffle the observed shifts 500 different times, run our main specification, retrieve and plot the estimate of the emigration effect and its confidence interval (for visualization clarity, the graph reports only the first 100 replications).

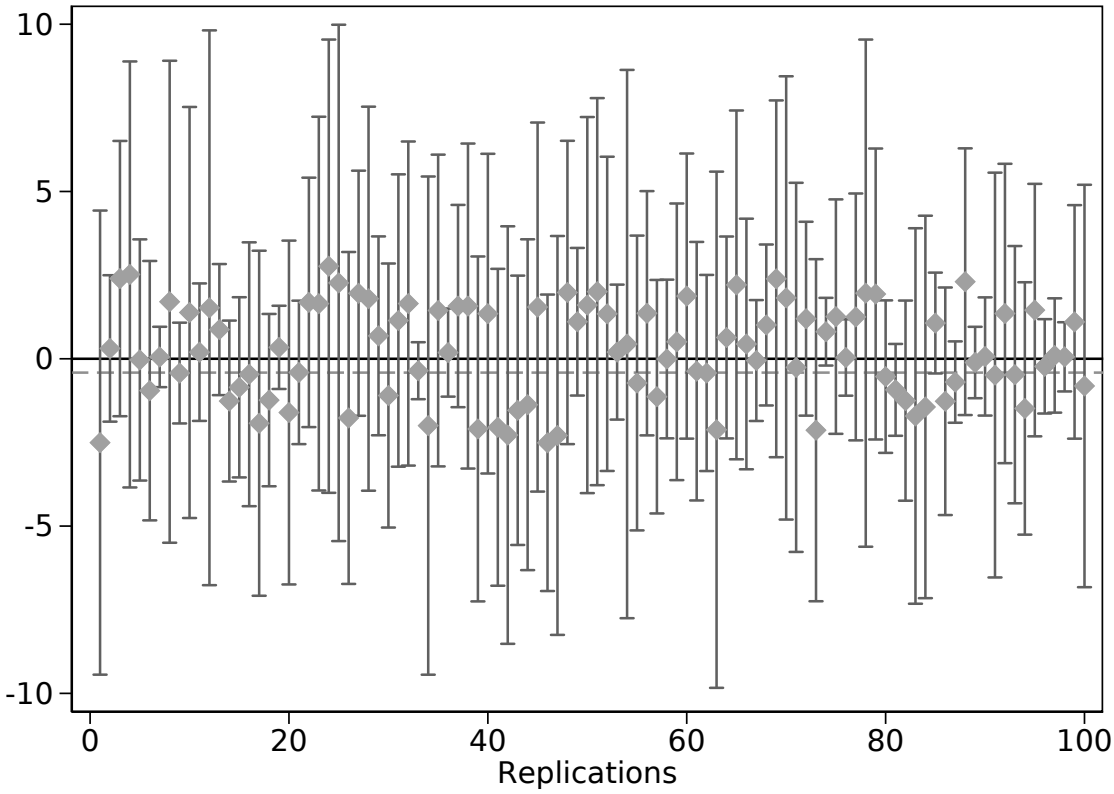
the main text.

Figure A4: Randomization of the Shifter components



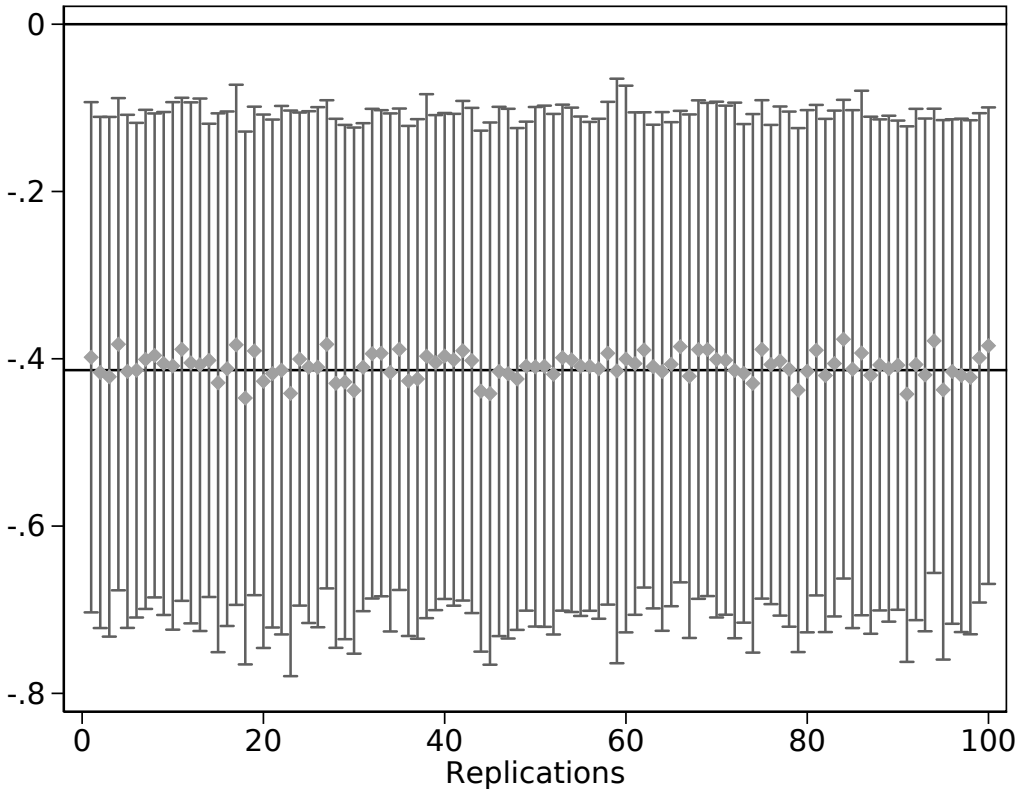
Notes: The graph reports the emigration effect estimates obtained in 100 different random draws of the GDP shifters using our baseline specification (the exercise is based on 500 replications, but for visualization clarity only 100 are reported in the graph). The estimated coefficients are significant 0.4 percent of the times at 10 percent level, and never significant at 5 percent level.

Figure A5: Randomization of the Share Components



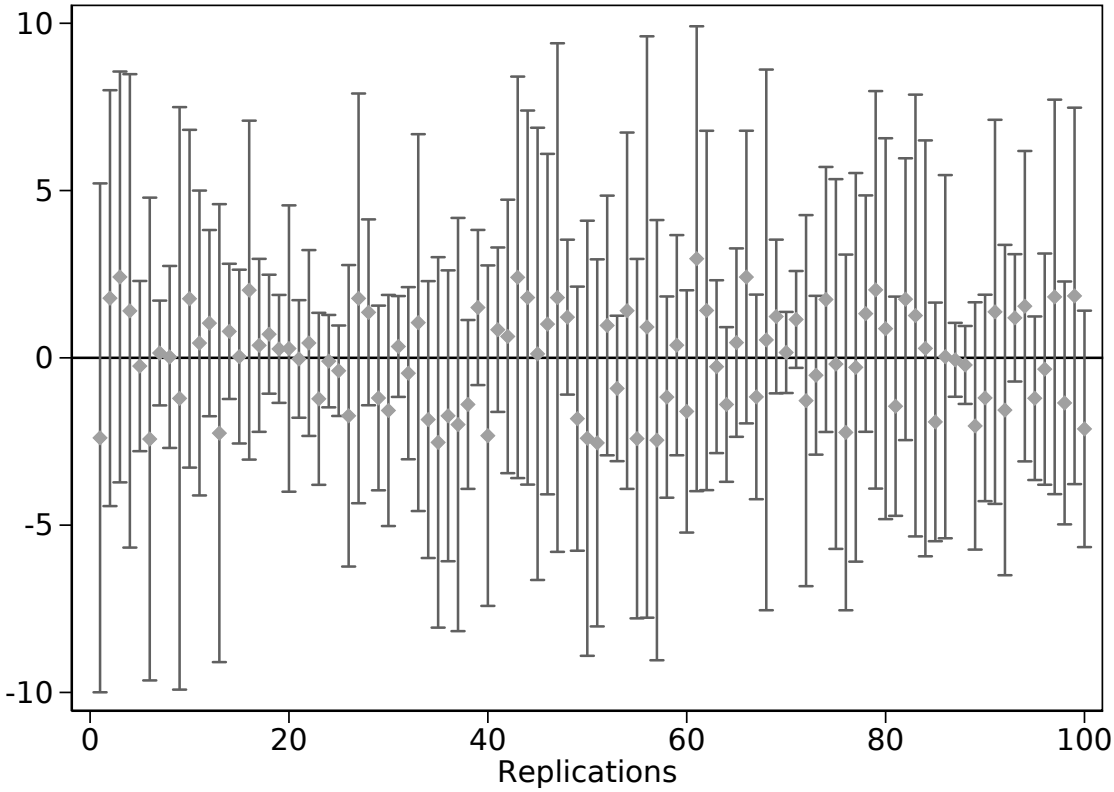
Notes: The graph reports the emigration effect estimates obtained in 100 different random draws of the emigration network shares using our baseline specification (the exercise is based on 500 replications, but for visualization clarity only 100 are reported in the graph). The estimated coefficients are never significant at 10 percent level.

Figure A6: Permutation of the Shifter components



Notes: The graph reports the emigration effect estimates obtained in 100 different random permutations of the GDP shifters using our baseline specification (the exercise is based on 500 replications, but for visualization clarity only 100 are reported in the graph). The average estimated coefficient is -0.43 and the estimated coefficients from all replications are significant at 5 percent level.

Figure A7: Permutation of the share components



Notes: The graph reports the emigration effect estimates obtained in 100 different random permutations of the emigration network shares using our baseline specification (the exercise is based on 500 replications, but for visualization clarity only 100 are reported in the graph). The average estimated coefficient is -0.989 and the estimated coefficients are significant 0.2% of the times at 5 percent level, and never at the 1 percent level.

Table A2: Instrument validity: effect of the instrument on pre-shock change in stock and flows of Young-owned firms (2005-08)

VARIABLES	(1) Young Firms Δ Stock 2005-08	(2) Young Firms $\sum$ Births 2005-08	(3) Young Firms $\sum$ Deaths 2005-08
Pull IV	0.015	-0.002	-0.017
	(0.049)	(0.048)	(0.028)
Observations	686	686	686
R-squared	0.222	0.530	0.340
Avg. Outcome	-0.135	2.783	2.918
Mean Pull IV	0.046	0.046	0.046
S.d. Pull IV	0.049	0.049	0.049
Controls	X	X	X
Province FE	X	X	X

Notes: OLS estimates. The sample is composed of 686 local labor markets (LLMs). The dependent variable is the change in stock (1), cumulative entry (2) and exit (3) of firms owned and managed by under 45 ("Young firms") between 2005-2008 as a fraction of LLM population 25-64 years old (average 2005-2008) times 100. The independent variable is the predicted emigration rate based on the shares of pre-2000 emigrants to different countries relative to the LLM population in 2000 interacted with GDP growth of each country between 2008-2015, $Pull_l = \sum_c NTWK_{l,c} * G_c$, and normalized to have mean zero and unit variance. Controls include unemployment rate and value added per capita in 100,000 euros in 2005 at the LLM level as well as 110 province FEs. Standard errors are clustered at the province level (110 clusters).

Table A3: Instrument validity check: effect of the instrument on pre-shock change in LLM employment (2005-08)

VARIABLES	(1) Δ Employees 2005-08	(2) Δ Emp/Pop 2005-08	(3) Δ Avg. Size 2005-08	(4) Δ Wage Bill 2005-08
Pull IV	-0.007	-0.001	-0.008	-0.009
	(0.010)	(0.010)	(0.009)	(0.011)
Observations	686	686	686	686
R-squared	0.282	0.260	0.177	0.194
Avg. Outcome 2005	16709.0	0.3	5.5	348.6
Mean Pull IV	0.046	0.046	0.046	0.046
S.d. Pull IV	0.049	0.049	0.049	0.049
Controls	X	X	X	X
Province FE	X	X	X	X

Notes: OLS estimates. The sample is composed of 686 local labor markets (LLMs). The dependent variable is the change in LLM employment (1), employment to population ratio (2), average firm size (3) and total wage bill in 100,000 euros (4) between 2005-2008, as a fraction of each outcome in 2005. The independent variable is the predicted emigration rate based on the shares of pre-2000 emigrants to different countries relative to the LLM population in 2000 interacted with GDP growth of each country between 2008-2015, $Pull_l = \sum_c NTWK_{l,c} * G_c$, and normalized to have mean zero and unit variance. Controls include unemployment rate and value added per capita in 100,000 euros in 2005 at the LLM level as well as 110 province FEs. Standard errors are clustered at the province level (110 clusters).

Table A4: Instrument validity check: effect of the instrument on pre-shock change in LLM skills (2005-08)

VARIABLES	(1) Δ Blue Coll 2005-08	(2) Δ White Coll 2005-08	(3) Δ Managers 2005-08
Pull IV	0.002	-0.011	0.293
	(0.011)	(0.012)	(0.286)
Observations	686	686	584
R-squared	0.323	0.137	0.135
Avg. Outcome 2005	8950.1	6737.4	191.7
Mean Pull IV	0.046	0.046	0.039
S.d. Pull IV	0.049	0.049	0.040
Controls	X	X	X
Province FE	X	X	X

Notes: OLS estimates. The sample is composed of 686 local labor markets (LLMs). The dependent variable is the change in LLM employment of blue collar workers (1), white collars (2) and managers (3) between 2005-2008, as a fraction of each outcome in 2005. The independent variable is the predicted emigration rate based on the shares of pre-2000 emigrants to different countries relative to the LLM population in 2000 interacted with GDP growth of each country between 2008-2015, $Pull_l = \sum_c NTWK_{l,c} * G_c$, and normalized to have mean zero and unit variance. Controls include unemployment rate and value added per capita in 100,000 euros in 2005 at the LLM level as well as 110 province FEs. Standard errors are clustered at the province level (110 clusters).

A.IV IV diagnostics for the consulate-based IV

The following Tables, A5 and A6, replicate the main tests proposed by Goldsmith-Pinkham, Sorkin, and Swift (2020) for the IV based on destination regions (Italian consulates abroad) rather than on the countries. Similarly to what shown in Table 4, Table A5 shows that the cross-sectional components of the pull emigration instrumental variable is driven by networks of Italian emigrants towards German and Swiss regions. Interestingly, the estimated coefficients of Stuttgart/Friburg and Dortmund/Koln, that alone make up about 40 percent of the IV variation, are close to each other (-0.498 and -0.355) and close to the main estimate at the consulate (-0.432) and country level (-0.433). Table A6 shows the correlations between the share of emigrants towards the most relevant regions and the main labor market characteristics: we fail to find statistically significant correlations with observable LLM characteristics, similarly to what shown in the country level analysis in the main text.

Table A5: Pull IV diagnostics (destination regions IV)

Panel A: Negative and positive weights

	Sum	Mean	Share
$\hat{\alpha}_c \leq 0$	-0.007	0	0.007
$\hat{\alpha}_c > 0$	1.007	0.008	0.993

Panel B: Correlations

	$\hat{\alpha}_c$	G_c	$\hat{\beta}_c$	$\hat{F}_c$	$Var(NTWK_c)$
$\hat{\alpha}_c$	1.0000				
G_c	-0.0735	1.0000			
$\hat{\beta}_c$	0.0072	0.0694	1.0000		
$\hat{F}_c$	0.0150	0.0296	0.0050	1.0000	
$Var(NTWK_c)$	0.7561	-0.1236	0.0043	-0.0041	1.0000

Panel C: Top 5 destination regions

	$\hat{\alpha}_c$	G_c	$\hat{\beta}_c$	$\hat{F}_c$	95% C.I.
Stuttgart/Friburg	0.250	1.075	-0.491	7.33	(-3.20, -0.10)
Zurich	0.105	1.01	-0.145	12.63	(-0.30, 0.00)
Dortmund/Koln	0.090	1.075	-0.343	2.64	$(-\infty,\infty)$
Lugano	0.076	1.010	-0.009	6.45	(-0.40, 0.10)
France	0.075	1.007	-0.364	3.56	$(-\infty, 1.10)$

Notes: The table reports the Pull IV diagnostics as suggested by Goldsmith-Pinkham, Sorkin, and Swift (2020). Panel A reports the sum, the mean and the share of negative and positive Rotemberg weights $\hat{\alpha}_c$. Panel B reports correlations between the weights ($\hat{\alpha}_c$), the 2008-2015 destination region/country GDP growth (G_c), the just-identified coefficients ($\hat{\beta}_c$), the first stage F-statistics for the just-identified instruments ($\hat{F}_c$) and the variance in the emigrant networks across destination region/country ($Var(NTWK_c)$). Panel C reports the top five destination regions according to the Rotemberg weights. The 95% CI are the weak instrument robust confidence intervals obtained with the Chernozhukov and Hansen (2008) method with a range from -10 to 10 ($(-\infty, \infty)$ indicates that the CI is undefined). The coefficients $\hat{\beta}_c$ are based on the regression of Table 7, column (1), where the outcome is the change 2008-2015 in the stock of firms per capita, and control variables include LLM value added per capita and unemployment rate in 2005 as well as 110 province FEs. We computed the Rotemberg decomposition using Goldsmith-Pinkham, Sorkin, and Swift (2020)'s Stata package. Sources for the regional GDP and population data: Eurostat (2019) and BFS (2017).

Table A6: Relationship between destination regions' emigration networks and pre-period LLMs characteristics

VARIABLES	(1) Share to Stuttgart/Friburg	(2) Share to Zurich	(3) Share to Dortmund/Koln	(4) Share to Lugano	(5) Share to France
Δ Stock	-0.002	-0.001	0.001	-0.001	0.002
	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)
$\sum$ Births	0.139	-0.039	0.043	-0.054	-0.085
	(0.100)	(0.031)	(0.058)	(0.022)	(0.084)
$\sum$ Deaths	0.006	-0.001	0.000	-0.001	-0.010
	(0.007)	(0.002)	(0.002)	(0.002)	(0.008)
Unemp Rate 2005	0.018	-0.011	0.035	-0.012	-0.006
	(0.049)	(0.020)	(0.036)	(0.009)	(0.028)
GDP PC 2005	-0.006	-0.005	-0.004	-0.005	-0.012
	(0.004)	(0.003)	(0.003)	(0.003)	(0.009)
Observations	653	666	651	645	683
Avg. Outcome	0.004	0.003	0.003	0.002	0.006
Controls	X	X	X	X	X
Province FE	X	X	X	X	X

Notes: OLS estimates, each coefficient is from a separate regression. The sample is composed of 686 local labor markets (LLMs). The dependent variable is the share of pre-2000 emigrants to each of the 5 top destination regions described in each column, relative to the LLM population in 2000. The independent variables are the main LLMs observable characteristics, namely the change in stock, cumulative entry and exit of firms between 2005-2008, unemployment rate and value added per capita in 100,000 euros in 2005. All regressions include 110 province FEs. Standard errors are clustered at the province level (110 clusters).

A.V Additional robustness checks

One may be concerned that our instrument is correlated with internal migration flows. While these should not be correlated with pull factors from abroad, the network of emigrants may be correlated with the internal flows and with local push factors. For instance, LLMs with high emigration rates to foreign countries could also exhibit substantial emigration to other Italian LLMs, and the latter may reduce firm creation, violating the exclusion restriction. We thus test whether our estimates are robust to this potential threat. In Table A7, columns (1) and (2), we report the results of a placebo first stage regression where we regress internal migration outflows and inflows on our emigration Pull IV. The effects are not statistically significant, suggesting that the instrument does not predict internal emigration to or immigration from other LLMs in Italy. In column (3), we test whether there is a direct substitution effect by regressing foreign immigration inflows on the instrument. Reassuringly, the instrument does not have a statistically significant effect on foreign immigration flows. Furthermore, as shown in Table 11, our main estimates are robust to the inclusion of immigration as a control variable.

Another concern is that our estimates may be capturing the effect of trade linkages, which may be correlated to migration flows. For this reason, in Table A8 we report the results of our main regression on firm entry by firm creation between tradable and non-tradable sectors. The largest impact of emigrants on firm creation is for non-tradable sector firms. This indicates that the emigration flows we are analyzing do not seem to be particularly linked to international trade activity.

In Tables A9 and A10, we estimate the first stage and the main specification using the 1992 (rather than 2000) emigration shares when constructing the IV. Results are very similar to those of Tables 6 and 7 albeit less precise, as mentioned in Section II.D. In Tables A11, A12, A13 and A14, we include the lag of the outcome variables among

the set of controls. In all cases the main results continue to hold. In Tables A15, A16 and A17, we show the results of regressing each outcome of Table 7 with different sets of controls and fixed effects: no controls in columns (1)-(3) and controlling for LLM unemployment rate and GDP per capita in 2005 in columns (4)-(6), and no fixed effects in columns (1) and (4), 20 regions fixed effects in columns (2) and (5) and 110 province fixed effects in columns (3) and (6). All results are qualitatively similar across the different specifications as long as we include at least some controls or fixed effects.

Table A7: Placebo first stage regression on internal migration flows and immigration

	(1)	(2)	(3)
VARIABLES	Internal Emig	Internal Immig	Immig Rate 05-08
Pull IV	0.078	-0.087	-0.022
	(0.069)	(0.052)	(0.040)
Observations	686	686	686
R-squared	0.405	0.451	0.710
F-excl. instrument	1.249	2.839	0.306
Mean Outcome	10.084	9.166	2.222
S.d. Outcome	2.377	3.233	1.255
Mean Pull IV	0.046	0.046	0.046
S.d. Pull IV	0.049	0.049	0.049
Controls	X	X	X
FE	Province	Province	Province

Notes: OLS estimates. The sample is composed of 686 local labor markets (LLMs). In columns (1) and (2), the dependent variable is the cumulative emigration and immigration rate of Italian citizens 25-64 years old to and from different LLMs in Italy between 2008-2015 respectively, while in column 3 the dependent variable is the cumulative immigration rate of foreign citizens 25-64 years old from abroad between 2005-2008. All the outcomes are as a fraction of the LLM population 25-64 years old (average 2005-2008), and normalized to have mean zero and unit variance. The independent variable is the predicted emigration rate based on the shares of pre-2000 emigrants to different countries relative to the LLM population in 2000 interacted with GDP growth of each country between 2008-2015, $Pull_l = \sum_c NTWK_{l,c} * G_c$, and normalized to have mean zero and unit variance. Controls include unemployment rate and value added per capita in 100,000 euros in 2005 at the LLM level. Column (1) includes no fixed effects while columns (2) and (3) include region (20) and province (110) FEs respectively. Standard errors are clustered at the province level (110 clusters).

Table A8: Effect of emigration rates on cumulative firm entry, in tradable and non tradable sectors

VARIABLES	(1) Tradable $\sum$ Births 2008-15	(2) Non Tradable $\sum$ Births 2008-15
Emig Rate	-0.069	-0.363
	(0.028)	(0.178)
Observations	686	686
R-squared	0.547	0.513
F-excl. instr.	14.851	14.851
Avg. Baseline Outcome	0.664	8.414
Mean Emig Rate	2.648	2.648
S.d. Emig Rate	1.696	1.696
Controls	X	X
Province FE	X	X

Notes: 2SLS estimates. The sample is composed of 686 local labor markets (LLMs). The dependent variable is the cumulative firm entry in tradable (1) and nontradable (2) sectors between 2008-2015 as a fraction of LLM population 25-64 years old (average 2005-2008) times 100. The independent variable is the cumulative emigration rate between 2008-2015, i.e. the number of Italian citizens aged 25-64 migrating abroad between 2008-2015 as a fraction of the 25-64 years old population in the origin LLM (average 2005-2008), and normalized to have mean zero and unit variance. The instrument is the predicted emigration rate based on the shares of pre-2000 emigrants to different countries relative to the LLM population in 2000 interacted with GDP growth of each country between 2008-2015, $Pull_l = \sum_c NTWK_{l,c} * G_c$. Controls include unemployment rate and value added per capita in 100,000 euros in 2005 at the LLM level as well as 110 province FEs. The average baseline outcomes are the change in firm stock, cumulative firm entry and exit in the pre-period (2005-2008) as a fraction of population 25-64 years old in the LLM (average 2005-2008) times 100, annualized (i.e., divided by 3 years) and multiplied by 7 years. Standard errors are clustered at the province level (110 clusters).

Table A9: Robustness check: first stage regressions, IV based on 1992 shares

VARIABLES	(1) Emig Rate	(2) Emig Rate	(3) Emig Rate
Pull IV (1992 shares)	0.311	0.323	0.305
	(0.069)	(0.073)	(0.091)
Unemp Rate 2005	-2.210	1.996	3.926
	(1.710)	(2.420)	(3.449)
GDP PC 2005	0.863	1.040	1.201
	(0.255)	(0.176)	(0.310)
Observations	686	686	686
R-squared	0.079	0.194	0.362
F-excl. instrument	20.460	19.444	11.248
Mean Emig Rate	2.648	2.648	2.648
S.d. Emig Rate	1.696	1.696	1.696
Mean Pull IV	0.028	0.028	0.028
S.d. Pull IV	0.031	0.031	0.031
FE	-	Region	Province

Notes: OLS estimates. The sample is composed of 686 local labor markets (LLMs). The dependent variable is the cumulative emigration rate between 2008 and 2015, i.e. the number of Italian citizens aged 25-64 migrating abroad between 2008-2015 as a fraction of the 25-64 years old population in the origin LLM (average 2005-2008), and normalized to have mean zero and unit variance. The independent variable is the predicted emigration rate based on the shares of pre-1992 emigrants to different countries relative to the LLM population in 1992 (source: Istat 2003) interacted with GDP growth of each country between 2008-2015, $Pull_l = \sum_c NTWK_{l,c} * G_c$, and normalized to have mean zero and unit variance. Controls include unemployment rate and value added per capita in 100,000 euros in 2005 at the LLM level. Column (1) includes no fixed effects while columns (2) and (3) include region (20) and province (110) FEs respectively. Standard errors are clustered at the province level (110 clusters).

Table A10: Robustness check: effect of emigration rates on change in stock and flows of firms, IV based on 1992 shares

VARIABLES	(1) All Firms Δ Stock 2008-15	(2) All Firms $\sum$ Births 2008-15	(3) All Firms $\sum$ Deaths 2008-15
Emig Rate	-0.507	-0.631	-0.124
	(0.208)	(0.302)	(0.258)
Observations	686	686	686
R-squared	0.170	0.478	0.241
F-excl. instr.	11.248	11.248	11.248
Avg. Baseline Outcome	0.790	9.078	8.288
Mean Emig Rate	2.648	2.648	2.648
S.d. Emig Rate	1.696	1.696	1.696
Controls	X	X	X
Province FE	X	X	X

Notes: 2SLS estimates. The sample is composed of 686 local labor markets (LLMs). The dependent variable is the change in firm stock (1), cumulative firm entry (2) and exit (3) between 2008-2015 as a fraction of LLM population 25-64 years old (average 2005-2008) times 100. The independent variable is the cumulative emigration rate between 2008-2015, i.e. the number of Italian citizens aged 25-64 migrating abroad between 2008-2015 as a fraction of the 25-64 years old population in the origin LLM (average 2005-2008), and normalized to have mean zero and unit variance. The instrument is the predicted emigration rate based on the shares of pre-1992 emigrants to different countries relative to the LLM population in 1992 interacted with GDP growth of each country between 2008-2015, $Pull_l = \sum_c NTWK_{l,c} * G_c$. Controls include unemployment rate and value added per capita in 100,000 euros in 2005 at the LLM level as well as 110 province FEs. The average baseline outcomes are the change in firm stock, cumulative firm entry and exit in the pre-period (2005-2008) as a fraction of population 25-64 years old in the LLM (average 2005-2008) times 100, annualized (i.e., divided by 3 years) and multiplied by 7 years. Standard errors are clustered at the province level (110 clusters).

Table A11: Robustness check: effect of emigration rates on change in stock and flows of firms (2008-2015) controlling for lagged outcomes (2005-08)

VARIABLES	(1) All Firms Δ Stock 2008-15	(2) All Firms $\sum$ Births 2008-15	(3) All Firms $\sum$ Deaths 2008-15
Emig Rate	-0.294	-0.403	-0.121
	(0.181)	(0.152)	(0.135)
Δ Stock	1.129		
	(0.007)		
$\sum$ Births		1.659	
		(0.078)	
$\sum$ Deaths			1.157
			(0.018)
Observations	686	686	686
R-squared	0.963	0.839	0.965
F-excl. instr.	14.853	14.865	14.843
Avg. Baseline Outcome	0.790	9.078	8.288
Mean Emig Rate	2.648	2.648	2.648
S.d. Emig Rate	1.696	1.696	1.696
Controls	X	X	X
Province FE	X	X	X

Notes: 2SLS estimates. The sample is composed of 686 local labor markets (LLMs). The dependent variable is the change in firm stock (1), cumulative firm entry (2) and exit (3) between 2008-2015 as a fraction of LLM population 25-64 years old (average 2005-2008) times 100. The independent variable is the cumulative emigration rate between 2008-2015, i.e. the number of Italian citizens aged 25-64 migrating abroad between 2008-2015 as a fraction of the 25-64 years old population in the origin LLM (average 2005-2008), and normalized to have mean zero and unit variance. The instrument is the predicted emigration rate based on the shares of pre-2000 emigrants to different countries relative to the LLM population in 2000 interacted with GDP growth of each country between 2008-2015, $Pull_l = \sum_c NTWK_{l,c} * G_c$. Controls include unemployment rate and value added per capita in 100,000 euros in 2005 at the LLM level as well as 110 province FEs. Additionally, we control for the change in stock, cumulative entry and exit of firms between 2005-2008 as a fraction of LLM population 25-64 years old (average 2005-2008). The average baseline outcomes are the change in firm stock, cumulative firm entry and exit in the pre-period (2005-2008) as a fraction of population 25-64 years old in the LLM (average 2005-2008) times 100, annualized (i.e., divided by 3 years) and multiplied by 7 years. Standard errors are clustered at the province level (110 clusters).

Table A12: Robustness check: effect of emigration rates on change in stock and flows of Young-owned firms (2008-15) controlling for lagged outcomes (2005-08)

VARIABLES	(1) Young Firms Δ Stock 2008-15	(2) Young Firms $\sum$ Births 2008-15	(3) Young Firms $\sum$ Deaths 2008-15
Emig Rate	-0.267	-0.226	0.050
	(0.147)	(0.120)	(0.132)
Δ Stock	0.736		
	(0.065)		
$\sum$ Births		1.496	
		(0.083)	
$\sum$ Deaths			1.068
			(0.194)
Observations	686	686	686
R-squared	0.658	0.810	0.792
F-excl. instr.	15.018	14.928	14.856
Avg. Baseline Outcome	-0.316	6.493	6.809
Mean Emig Rate	2.648	2.648	2.648
S.d. Emig Rate	1.696	1.696	1.696
Controls	X	X	X
Province FE	X	X	X

Notes: 2SLS estimates. The sample is composed of 686 local labor markets (LLMs). The dependent variable is the change in stock (1), cumulative entry (2) and exit (3) of firms owned and managed by under 45 ("Young firms") between 2008-2015 as a fraction of LLM population 25-64 years old (average 2005-2008) times 100. The independent variable is the cumulative emigration rate between 2008-2015, i.e. the number of Italian citizens aged 25-64 migrating abroad between 2008-2015 as a fraction of the 25-64 years old population in the origin LLM (average 2005-2008), and normalized to have mean zero and unit variance. The instrument is the predicted emigration rate based on the shares of pre-2000 emigrants to different countries relative to the LLM population in 2000 interacted with GDP growth of each country between 2008-2015, $Pull_l = \sum_c NTWK_{l,c} * G_c$. Controls include unemployment rate and value added per capita in 100,000 euros in 2005 at the LLM level as well as 110 province FEs. Additionally, we control for the change in stock, cumulative entry and exit of firms owned and managed by under 45 between 2005-2008 as a fraction of LLM population 25-64 years old (average 2005-2008). The average baseline outcomes are the change in firm stock, cumulative firm entry and exit in the pre-period (2005-2008) as a fraction of population 25-64 years old in the LLM (average 2005-2008) times 100, annualized (i.e., divided by 3 years) and multiplied by 7 years. Standard errors are clustered at the province level (110 clusters).

Table A13: Robustness check: effect of emigration rates on LLM employment (2008-15) controlling for lagged outcome (2005-08)

VARIABLES	(1) Δ Employees 2008-15	(2) Δ Emp/Pop 2008-15	(3) Δ Avg. Size 2008-15	(4) Δ Wage Bill 2008-15
Emig Rate	-0.045	-0.024	-0.014	-0.018
	(0.020)	(0.021)	(0.025)	(0.022)
Δ Employees	0.061			
	(0.087)			
Δ Emp/Pop		0.120		
		(0.095)		
Δ Avg. Size			0.048	
			(0.115)	
Δ Wage Bill				0.018
				(0.063)
Observations	686	686	686	686
R-squared	0.197	0.216	0.242	0.264
F-excl. instr.	15.013	15.251	13.962	14.583
Avg. Outcome 2005	16709.0	0.3	5.5	348.6
Mean Emig Rate	2.648	2.648	2.648	2.648
S.d. Emig Rate	1.696	1.696	1.696	1.696
Controls	X	X	X	X
Province FE	X	X	X	X

Notes: 2SLS estimates. The sample is composed of 686 local labor markets (LLMs). The dependent variable is the change in LLM employment (1), employment to population ratio (2), average firm size (3) and total wage bill in 100,000 euros (4) between 2008-2015, as a fraction of each outcome in 2005. The independent variable is the cumulative emigration rate between 2008-2015, i.e. the number of Italian citizens aged 25-64 migrating abroad between 2008-2015 as a fraction of the 25-64 years old population in the origin LLM (average 2005-2008), and normalized to have mean zero and unit variance. The instrument is the predicted emigration rate based on the shares of pre-2000 emigrants to different countries relative to the LLM population in 2000 interacted with GDP growth of each country between 2008-2015, $Pull_l = \sum_c NTWK_{l,c} * G_c$. Controls include unemployment rate and value added per capita in 100,000 euros in 2005 at the LLM level as well as 110 province FEs. Additionally, we control for the percentage change in each outcome between 2005-2008. Standard errors are clustered at the province level (110 clusters).

Table A14: Robustness check: effect of emigration rates on LLM skills (2008-15) controlling for lagged outcome (2005-08)

	(1)	(2)	(3)
	Δ Blue Coll	Δ White Coll	Δ Managers
VARIABLES	2008-15	2008-15	2008-15
Emig Rate	-0.018	-0.055	-1.168
	(0.027)	(0.028)	(1.074)
Δ Blue Coll	0.002		
	(0.079)		
Δ White Coll		0.155	
		(0.103)	
Δ Managers			0.050
			(0.250)
Observations	686	686	584
R-squared	0.199	0.240	0.184
F-excl. instr.	15.293	14.746	6.361
Avg. Outcome 2005	8950.1	6737.4	191.7
Mean Emig Rate	2.648	2.648	2.544
S.d. Emig Rate	1.696	1.696	1.369
Controls	X	X	X
Province FE	X	X	X

Notes: 2SLS estimates. The sample is composed of 686 local labor markets (LLMs). The dependent variable is the change in LLM employment of blue collar workers (1), white collars (2) and managers (3) between 2008-2015, as a fraction of each outcome in 2005. The independent variable is the cumulative emigration rate between 2008-2015, i.e. the number of Italian citizens aged 25-64 migrating abroad between 2008-2015 as a fraction of the 25-64 years old population in the origin LLM (average 2005-2008), and normalized to have mean zero and unit variance. The instrument is the predicted emigration rate based on the shares of pre-2000 emigrants to different countries relative to the LLM population in 2000 interacted with GDP growth of each country between 2008-2015, $Pull_l = \sum_c NTWK_{l,c} * G_c$. Controls include unemployment rate and value added per capita in 100,000 euros in 2005 at the LLM level as well as 110 province FEs. Additionally, we control for the percentage change in each outcome between 2005-2008. Standard errors are clustered at the province level (110 clusters).

Table A15: Robustness check: effect of emigration rates on change in stock of firms, different controls and fixed effects

VARIABLES	(1) All Firms Δ Stock 2008-15	(2) All Firms Δ Stock 2008-15	(3) All Firms Δ Stock 2008-15	(4) All Firms Δ Stock 2008-15	(5) All Firms Δ Stock 2008-15	(6) All Firms Δ Stock 2008-15
Emig Rate	-0.384	-0.715	-0.573	-0.550	-0.647	-0.414
	(0.369)	(0.206)	(0.191)	(0.213)	(0.196)	(0.155)
Observations	686	686	686	686	686	686
R-squared		0.019	0.156		0.030	0.175
F-excl. instr.	26.223	31.313	13.469	28.311	29.564	14.851
Avg. Baseline Outcome	0.790	0.790	0.790	0.790	0.790	0.790
Mean Emig Rate	2.648	2.648	2.648	2.648	2.648	2.648
S.d. Emig Rate	1.696	1.696	1.696	1.696	1.696	1.696
Controls	-	-	-	X	X	X
FE	-	Region	Province	-	Region	Province

Notes: 2SLS estimates. The sample is composed of 686 local labor markets (LLMs). In columns (1)-(3) no controls are included, while in columns (4)-(6) we control for unemployment rate and value added per capita in 100,000 euros in 2005 at the LLM level. In columns (1) and (4) no fixed effects or controls are included, in columns (2) and (5) we include 20 region fixed effects, and in columns (3) and (6) we include 110 province fixed effects. The dependent variable is the change in firm stock between 2008-2015 as a fraction of LLM population 25-64 years old (average 2005-2008) times 100. The independent variable is the cumulative emigration rate between 2008-2015, i.e. the number of Italian citizens aged 25-64 migrating abroad between 2008-2015 as a fraction of the 25-64 years old population in the origin LLM (average 2005-2008), and normalized to have mean zero and unit variance. The instrument is the predicted emigration rate based on the shares of pre-2000 emigrants to different countries relative to the LLM population in 2000 interacted with GDP growth of each country between 2008-2015, $Pull_l = \sum_c NTWK_{l,c} * G_c$. The average baseline outcomes are the change in firm stock, cumulative firm entry and exit in the pre-period (2005-2008) as a fraction of population 25-64 years old in the LLM (average 2005-2008) times 100, annualized (i.e., divided by 3 years) and multiplied by 7 years. Standard errors are clustered at the province level (110 clusters).

Table A16: Robustness check: Effect of emigration rates on cumulative firm entry, different controls and fixed effects

VARIABLES	(1) All Firms $\sum$ Births 2008-15	(2) All Firms $\sum$ Births 2008-15	(3) All Firms $\sum$ Births 2008-15	(4) All Firms $\sum$ Births 2008-15	(5) All Firms $\sum$ Births 2008-15	(6) All Firms $\sum$ Births 2008-15
Emig Rate	-1.301	-0.631	-0.514	-0.769	-0.577	-0.432
	(0.368)	(0.242)	(0.222)	(0.263)	(0.240)	(0.196)
Observations	686	686	686	686	686	686
R-squared		0.207	0.493		0.240	0.527
F-excl. instr.	26.223	31.313	13.469	28.311	29.564	14.851
Avg. Baseline Outcome	9.078	9.078	9.078	9.078	9.078	9.078
Mean Emig Rate	2.648	2.648	2.648	2.648	2.648	2.648
S.d. Emig Rate	1.696	1.696	1.696	1.696	1.696	1.696
Controls	-	-	-	X	X	X
FE	-	Region	Province	-	Region	Province

Notes: 2SLS estimates. The sample is composed of 686 local labor markets (LLMs). In columns (1)-(3) no controls are included, while in columns (4)-(6) we control for unemployment rate and value added per capita in 100,000 euros in 2005 at the LLM level. In columns (1) and (4) no fixed effects or controls are included, in columns (2) and (5) we include 20 region fixed effects, and in columns (3) and (6) we include 110 province fixed effects. The dependent variable is the cumulative firm entry between 2008-2015 as a fraction of LLM population 25-64 years old (average 2005-2008) times 100. The independent variable is the cumulative emigration rate between 2008-2015, i.e. the number of Italian citizens aged 25-64 migrating abroad between 2008-2015 as a fraction of the 25-64 years old population in the origin LLM (average 2005-2008), and normalized to have mean zero and unit variance. The instrument is the predicted emigration rate based on the shares of pre-2000 emigrants to different countries relative to the LLM population in 2000 interacted with GDP growth of each country between 2008-2015, $Pull_l = \sum_c NTWK_{l,c} * G_c$. The average baseline outcomes are the change in firm stock, cumulative firm entry and exit in the pre-period (2005-2008) as a fraction of population 25-64 years old in the LLM (average 2005-2008) times 100, annualized (i.e., divided by 3 years) and multiplied by 7 years. Standard errors are clustered at the province level (110 clusters).

Table A17: Robustness check: effect of emigration rates on cumulative firm exit, different controls and fixed effects

VARIABLES	(1) All Firms $\sum$ Deaths 2008-15	(2) All Firms $\sum$ Deaths 2008-15	(3) All Firms $\sum$ Deaths 2008-15	(4) All Firms $\sum$ Deaths 2008-15	(5) All Firms $\sum$ Deaths 2008-15	(6) All Firms $\sum$ Deaths 2008-15
Emig Rate	-0.917	0.083	0.059	-0.219	0.070	-0.018
	(0.462)	(0.258)	(0.187)	(0.293)	(0.256)	(0.189)
Observations	686	686	686	686	686	686
R-squared		0.095	0.238	0.016	0.095	0.241
F-excl. instr.	26.223	31.313	13.469	28.311	29.564	14.851
Avg. Baseline Outcome	8.288	8.288	8.288	8.288	8.288	8.288
Mean Emig Rate	2.648	2.648	2.648	2.648	2.648	2.648
S.d. Emig Rate	1.696	1.696	1.696	1.696	1.696	1.696
Controls	-	-	-	X	X	X
FE	-	Region	Province	-	Region	Province

Notes: 2SLS estimates. The sample is composed of 686 local labor markets (LLMs). In columns (1)-(3) no controls are included, while in columns (4)-(6) we control for unemployment rate and value added per capita in 100,000 euros in 2005 at the LLM level. In columns (1) and (4) no fixed effects or controls are included, in columns (2) and (5) we include 20 region fixed effects, and in columns (3) and (6) we include 110 province fixed effects. The dependent variable is the cumulative firm exit between 2008-2015 as a fraction of LLM population 25-64 years old (average 2005-2008) times 100. The independent variable is the cumulative emigration rate between 2008-2015, i.e. the number of Italian citizens aged 25-64 migrating abroad between 2008-2015 as a fraction of the 25-64 years old population in the origin LLM (average 2005-2008), and normalized to have mean zero and unit variance. The instrument is the predicted emigration rate based on the shares of pre-2000 emigrants to different countries relative to the LLM population in 2000 interacted with GDP growth of each country between 2008-2015, $Pull_l = \sum_c NTWK_{l,c} * G_c$. The average baseline outcomes are the change in firm stock, cumulative firm entry and exit in the pre-period (2005-2008) as a fraction of population 25-64 years old in the LLM (average 2005-2008) times 100, annualized (i.e., divided by 3 years) and multiplied by 7 years. Standard errors are clustered at the province level (110 clusters).

A.VI Additional tables

Table A18 presents reduced form estimates by regressing the 2008-2015 change in the main outcomes of interest directly on our IV. These reduced form results are a useful benchmark to compare the magnitude of the main effects with the pre-trends shown in Table 3.

Table A18: Reduced-form: Effect of the instrument on change in stock and flows of firms

	(1)	(2)	(3)
	All Firms	All Firms	All Firms
	Δ Stock	$\sum$ Births	$\sum$ Deaths
VARIABLES	2008-15	2008-15	2008-15
Pull IV	-0.180	-0.188	-0.008
	(0.053)	(0.084)	(0.090)
Observations	686	686	686
R-squared	0.186	0.572	0.241
Avg. Outcome	-0.123	7.867	7.990
Mean Pull IV	0.046	0.046	0.046
S.d. Pull IV	0.049	0.049	0.049
Controls	X	X	X
Province FE	X	X	X

Notes: OLS estimates. The sample is composed of 686 local labor markets (LLMs). The dependent variable is the change in firm stock (1), cumulative firm entry (2) and exit (3) between 2008-2015 as a fraction of LLM population 25-64 years old (average 2005-2008) times 100. The independent variable is the predicted emigration rate based on the shares of pre-2000 emigrants to different countries relative to the LLM population in 2000 interacted with GDP growth of each country between 2008-2015, $Pull_l = \sum_c NTWK_{l,c} * G_c$. Controls include unemployment rate and value added per capita in 100,000 euros in 2005 at the LLM level as well as 110 province FEs. The average baseline outcomes are the change in firm stock, cumulative firm entry and exit in the pre-period (2005 to 2008) as a fraction of average pre-period population 25-64 years old in the LLM times 100, annualized (i.e., divided by 3 years) and multiplied by 7 years. Standard errors are clustered at the province level (110 clusters).

References

Adão, Rodrigo, Michal Kolesár, and Eduardo Morales. 2019. "Shift-share designs: Theory and inference." *The Quarterly Journal of Economics* 134 (4): 1949–2010.

BFS. 2017. "Gross domestic product per canton and region." Switzerland Federal Statistical Office [publisher] `https://www.bfs.admin.ch/bfs/en/home/statistics/national-economy/national-accounts/gross-domestic-product-canton.html` (accessed September 15, 2019).

———. 2018. "International Migrations of the Permanent Resident Population by Citizenship, Sex, and Age, 1991-2016." Switzerland Federal Statistical Office (BFS) [publisher] (accessed July 2, 2018).

Chernozhukov, Victor, and Christian Hansen. 2008. "The reduced form: A simple approach to inference with weak instruments." *Economics Letters* 100 (1): 68–71.

Eurostat. 2019. "Regional gross domestic product (PPS per inhabitant) by NUTS 2 regions [TGS00005]." European Commission [publisher] `https://ec.europa.eu/eurostat/databrowser/view/TGS00005/default/table` (accessed September 15, 2019).

Fouka, Vasiliki, Soumyajit Mazumder, and Marco Tabellini. 2020, October. "From Immigrants to Americans: Race and Assimilation During the Great Migration." Harvard Business School Working Paper, No. 19-018.

Goldsmith-Pinkham, Paul, Isaac Sorkin, and Henry Swift. 2020. "Bartik Instruments: What, When, Why, and How." *American Economic Review* 110 (8): 2586–2624.

Istat. 2003. "Inter-census Population of Italian municipalities by age and sex, 1992-2001." Istituto Nazionale di Statistica (Istat) [publisher] `https://demo.istat.it` (accessed September 15, 2019).

NINo. 2018. "National Insurance Number (NINo) Registrations To Adult Overseas Nationals Entering The UK (XI)." UK Department for Work and Pensions [publisher] (accessed April 10, 2018).

Ruggles, Steven, Sarah Floodand, Ronald Goeken, Erin Meyer, Jose Pacas, and Matthew Sobek. 2019. "Integrated Public Use Microdata Series (IPUMS USA): Version 9.0 [dataset]." Minneapolis, MN: IPUMS, 10, D010 (accessed September 4, 2018) URL: http://doi.org/10.18128/D010.V9.0.

U.S., Census Bureau. 2018. "2005-2016 American Community Survey." U.S. Census Bureau [publisher] https://www.census.gov/programs-surveys/acs (accessed September 4, 2018).