Science Advances

AAAS

advances.sciencemag.org/cgi/content/full/3/4/e1602232/DC1

Supplementary Materials for

Quantifying the negative impact of brain drain on the integration of European science

Omar A. Doria Arrieta, Fabio Pammolli, Alexander M. Petersen

Published 12 April 2017, *Sci. Adv.* **3**, e1602232 (2017)
DOI: 10.1126/sciadv.1602232

This PDF file includes:

Supplementary Materials and Methods

Additional Data Description

World Bank country-level R&D data: We used researcher population, government spending, and gross domestic product (GDP) data from the World Bank data repository (*41*):

1. "Researchers in R&D (per million people)", given by $Spc_{i,t}$, with mean $\pm$ standard deviation $= 2{,}900 \pm 1{,}700$;

2. The total number of researchers in R&D, given by $S_{i,t}$ (calculated using Population data in combination with $Spc_{i,t}$), with mean $\pm$ std. dev. $= 47{,}000 \pm 72{,}000$;

3. "Research and development expenditure (% of GDP)", given by $e_{i,t}$, with mean $\pm$ std. dev. $= 1.47 \pm 0.89$; We then use GDP data to convert $e_{i,t}$ to the total R&D expenditure, $E_{i,t}$;

4. "GDP (current US$)", given by $GDP_{i,t}$, with mean $\pm$ std. dev. of $\log_{10} GDP_{i,t} = 11 \pm 0.75$; and

5. "GDP per capita (current US$)", given by $GDPpc_{i,t}$, with mean $\pm$ std. dev. of the log value ($\log_{10} GDPpc_{i,t}$) $= 4.4 \pm 0.36$.

We deflated all dollar amounts to 2010 USD$. Averaging across 32 EU and 57 large non-EU countries, the average annual growth rate of $S_{i,t}$ is 4–6%, and the average annual growth rate of the total R&D expenditure is between 8–9%; over this period, there is little difference between the EU and non-EU growth rates of total R&D expenditure.

Mobility data (EU High-skilled): Competitiveness in the global economy is increasingly becoming linked to the high-skilled "knowledge" economy (*43*). And while Europe is certainly producing a large number of high-skilled laborers, it is also home to large stocks of high-skilled emigrants (*22*, *36*), in particular scientists (*12*, *17*). The study of researcher mobility has been aided by large publication datasets (*44–46*), facilitating new studies into the supply-demand for researchers, which can oftentimes be linked to specific policies and programmes. However, the availability of comprehensive researcher career data, as well technical (name disambiguation) problems that exist when attempting to extract researcher trajectories from raw publication metadata, mean that researcher mobility data is difficult to acquire and certainly not comprehensive in its coverage of all scientists.

As a proxy for researcher mobility trends, we used official EU Commission "Professionals moving abroad (Establishment)" data from The EU Single Market Regulated professionals database. This database tracks the number of (high-skilled) professionals who obtained official certification in a given country of qualification (source country), and then applied for official recognition of their professional certification in a particular host country (destination country) (*14*). Lacking the mobility outcome data, we assume that the actual number of migrating professionals is highly correlated with the number of positive decisions to recognize the professional certification in a given destination country – i.e. we assume that if an individual has their application approved then they move with high probability. As such, we also assume that the information captured by the high-skilled mobility data is highly correlated to scientific mobility trends over the same period. The database covers a variety of certification "Recognition Regime" categories (e.g. "Pharmacist", "Doctor in basic and specialized medicine both listed in Annex V", etc.). We aggregated the data for all professions using the option "Recognition Regime =All". For more specific description of their counting methods and the outcome statistics, see the data description page. The data are grouped into 13 periods indexed here by $t = 1...13$ corresponding to 1997/1998, 1999/2000, 2001/2002, 2003/2004, 2005/2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014. We did not include the final 2 years of data in our analysis because the mobility data was either incomplete or still being updated and because the World Bank R&D data is incomplete for many countries after 2012. It is also worth explicitly stating that we divided the mobility headcount variables for periods in $t \leq 2006$ by a factor of two so that these count values refer to mean annual rates. As such, in order to combine observations across these three datasets, was also aggregated the count data for publications and country-level economic indicators across the specified 2-year periods and then divided by a factor of 2, resulting in 2-year annual averages.

Thus, for each year period t we recorded $M_{ij,t}$, the total number of high-skilled migrations ("Total positive decisions") from country i ("Country of qualification") to country j ("Host country"). In all, the total mobility (headcounts) for a given time period x, $M_x = \sum_{ij} M_{ij,x}$, are 315,888 (1997–2012), 43,075 (1997–2004), and 272,813 (2005–2012). We also recorded the number of "Total negative decisions", $N_{ij,t}$, corresponding to those applications which were denied (for a variety of reasons). The total number of negative decisions by period are 24,046 (1997–2012), 4,734 (1997–2004), and 19,312 (2005–2012), representing roughly 7% of the total (positive and negative) decisions made.

We used this data to analyze the intra-EU mobility rates before (<) and after (>) the 2004 EU enlargement. The total incoming mobility before and after are given by $I^{+}_{i,<} = \sum_{j,t\leq 2004} M_{ji,t}$ and $I^{+}_{i,>} = \sum_{j,t\geq 2005} M_{ji,t}$, respectively; the total outgoing mobility before and after are then given by $O^{+}_{i,<} = \sum_{j,t\leq 2004} M_{ij,t}$ and $O^{+}_{i,>} = \sum_{j,t\geq 2005} M_{ij,t}$, respectively. Furthermore, the negative decisions can also be aggregated by country: $I^{-}_{i,<} = \sum_{j,t\leq 2004} N_{ji,t}$, $I^{-}_{i,>} = \sum_{j,t\geq 2005} N_{ji,t}$, $O^{-}_{i,<} = \sum_{j,t\leq 2004} N_{ij,t}$, and $O^{-}_{i,>} = \sum_{j,t\geq 2005} N_{ij,t}$. At the annual level, $I^{y}_{i,t}$ and $O^{y}_{i,t}$ refers to total incoming and outgoing counts within period t and decision type $y = \pm$. Figure S3(C–H) shows the time series' of $I^{+}_{i,t}$ and $O^{+}_{i,t}$ and $I^{+}_{i,t}/O^{+}_{i,t}$ for each country.

The "success rate" of outgoing (incoming) applications contains information about the competitiveness (selectivity) of the source (host) country. We define the incoming and outgoing success rates using the relative frequency of positive (+) and negative (–) decisions, $\mathcal{P}^{in}_{i,t} = I^{+}_{i,t}/(I^{+}_{i,t} + I^{-}_{i,t})$ and $\mathcal{P}^{out}_{i,t} = O^{+}_{i,t}/(O^{+}_{i,t} + O^{-}_{i,t})$, respectively. As above, we use the notation $\mathcal{P}^{in}_{i,<}$ referring to the net success rates calculated by aggregating periods $t \leq 2004$, and $\mathcal{P}^{in}_{i,>}$ referring to the net success rates calculated by aggregating periods $t \geq 2005$. These success rates can also be generalized to country-country pairs at variable time resolution ($x = \{t, <, >\}$) according to the definitions $\mathcal{P}^{in}_{ij,x} = I^{+}_{ij,x}/(I^{+}_{ij,x} + I^{-}_{ij,x})$ and $\mathcal{P}^{out}_{ij,x} = O^{+}_{ij,x}/(O^{+}_{ij,x} + O^{-}_{ij,x})$.

We use the Gini index $G^{in}_{i,t}$ ($G^{out}_{i,t}$) to measure the concentration of the incoming (outgoing) mobility across the other EU member states. For example, $G^{out}_{i,t}$ is calculated using the 31 possible destination countries (j) of country i in the mobility network as $G^{out}_{i,t} = (\sum_{j=1}^{31} \sum_{k!=j}^{30} | M_{ij,t} - M_{ik,t}|)/(2(31 - 1)^2 \langle M^{out}_{i}(t) \rangle)$ where $\langle M^{out}_{i}(t) \rangle$ is the average outgoing mobility of i in t; $G^{in}_{i,t}$ is calculated by swiching the order of i and j, k in the matrices to represent incoming counts. $G_{i,t}$ is particularly useful in our case because it is standardized over the fixed unit interval [0,1], thus it is less sensitive to the large variations in $M_{ij,t}$: the minimum value 0 represents the case in which the mobility is dispersed evenly across all the other countries, and the maximum value 1 represents the case in which the mobility is entirely concentrated on one country with no mobility to any other countries. Thus, this quantity controls for the strong variation in the incoming and outgoing links from any given i in the mobility network (see figs. S4 and S5).

We define the 'relative' net mobility, or mobility polarization, as $B_{i,x} = (O^{+}_{i,x} - I^{+}_{i,x})/(O^{+}_{i,x} + I^{+}_{i,x}) \in [-1,1]$. This quantity measures the mobility polarization: the extreme values $B_{i,x} = -1$ corresponds to $O^{+}_{i,x} = 0$ and $I^{+}_{i,x} > 0$ (entirely incoming mobility) and $B_{i,x} = 1$ corresponds to $O^{+}_{i,x} > 0$ and $I^{+}_{i,x} = 0$ (entirely outgoing mobility). By construction, this measure is centered around zero and is useful as a relative measure to compare countries with total mobility rates that differ across several orders of magnitude, as illustrated in fig. S4.

Migration data (high-skilled + low-skilled): In order to account for underlying global migration trends, we used data from Abel & Sander (*15*), who provide estimates of the bilateral migration (high-skilled + low-skilled) between countries i and j, given by the matrix $\widetilde{M}_{ij,\tau}$, which they calculated aggregating official country statistics over three 5-year periods, $\tau = 1$ (1995–2000), $\tau = 2$ (2000–2005), and $\tau = 3$ (2005–2010). This novel dataset uses sequential population stock tables, including census data about birthplace and refugee and population statistics, to reconstruct and estimate the aggregate $\widetilde{M}_{ij,\tau}$ headcount data. For a recent study comparing the changes in high-skilled versus low-skilled labor in OECD countries see Kerr et al. (*22*).

Here we use this data to calculate the analogs of the total mobility (I/O) and diversity (G) measures described above: the total migration from (to) country i given by $\widetilde{O}_{i,\tau}$ ($\tilde{I}_{i,\tau}$) and the Gini index of the migration from (to) country i given by $\widetilde{G}^{out}_{i,\tau}$ ($\widetilde{G}^{in}_{i,\tau}$). We approximate the global migration data for $t = 2011, 2012$ using the $\widetilde{M}_{ij,\tau}$ values for 2005–2010.

As above for the high-skilled mobility, we also define the total migration polarization $\widetilde{B}_{i,\tau} = (\widetilde{O}_{i,\tau} - \tilde{I}_{i,\tau})/(\widetilde{O}_{i,\tau} + \tilde{I}_{i,\tau}) \in [-1,1]$. Moreover, fig. S6 shows the mobility ratio matrix $\rho_{ij,x} \equiv M_{ij,x}/\widetilde{M}_{ij,x} \in [0,1]$ for before ($<$, corresponding to $\tau = 1,2$) and after ($>$, corresponding to $\tau = 3$) the 2004 EU enlargement. The ratio $\rho_{ij,x}$ is a proxy for the fraction of total mobility from country i to j corresponding to high-skilled labor. These statistics indicate that the majority of migration is not high-skilled laborers, however, the ratio is increasing over time: the mean value before 2004 is $\langle \rho_{<} \rangle = 0.02$ whereas the mean value after 2004 is $\langle \rho_{>} \rangle = 0.14$.

Estimating the negative impact of joining the EU using the Synthetic Control Method

In order to explain the divergence in cross-border collaboration between Western and Eastern Europe, we use the 2004 EU enlargement as a policy experiment characterized by a large subset of 10 countries with coinciding "policy intervention" (treatment) year $t^{*} = 2004$.[1] A naive assumption might be that

[1] Bulgaria and Romania serve as a second policy experiment with lagged "treatment" (entry) year $t^{*} = 2007$.

the 2004 entrants would produce more cross-border publications ($\chi_{i,t}^{s}$) after entry into the EU because of increased access to EU framework programme funding and collaborative opportunities facilitated by the "integrated" EU R&D system. However, we find the contrary to be true, that new entrants *would have produced more publications* – both in frequency $f_{i,t}^{s}$ per publication and total number $\chi_{i,t}^{s}$ – had they not entered the EU. This result provides a partial explanation for why the EU cross-border collaboration rate grew no faster than international rates during this period, representing a "stagnation" of the EU integration process (*8*).

We demonstrate this counterintuitive outcome on cross-border collaboration within the EU using the Synthetic Control Method (SCM) (*24*, *26*, *47*). This method estimates the effect of the counterfactual outcome – that each EU entrant country *had not participated in the EU enlargement* – on our two measures of cross-border integration: the fraction $\hat{f}_{i,t}^{s}$ and total number $\hat{\chi}_{i,t}^{s}$ of cross-border publications. We used a control group of $N_c = 26$ non-EU countries $\{j\}$ ={AR, AM, AZ, BY, CA, CN, CO, CU, IN, IL, JP, KZ, KW, KG, MG, MX, MN, PA, RU, RS, SG, KR, TT, TR, UA, US} to estimate the counterfactual cross-border trends $\hat{f}_{i,t}^{s}$ and $\hat{\chi}_{i,t}^{s}$ for $t \geq 2004$. Thus, the difference δ between the synthetic outcome and the real outcome corresponding to the "EU Entry Effect". Because none of the control group countries belong to the EU, the implicit assumption of no interference between units is satisfied – i.e. enlargement of the EU should not be significantly correlated to international collaboration rates in Japan, for example.

The SCM produces an optimal representation of the actual time series of interest, $Z_{i,t}(= \log_{10}\chi_{i,t}^{s}$ or $f_{i,t}^{s})$, based upon best-fit weights calculated using the control country data for the time period before EU entry ($t < 2004$).[2] The covariate data ($X_{i,t}$) we used to model $Z_{i,t}$ are the total number of publications ($\log_{10}D_{i,t}^{s}$), the normalized citations ($R_{i,t}^{s}$), the per-capita GDP ($\log_{10}GDPpc_{i,t}$), and government expenditure on R&D as % of GDP, $e_{i,t}$.[3] The factor model representation of the dependent variable is given by

$$Z_{i,t} = \gamma_t + \theta_t X_i + \lambda_t \mu_i + \epsilon_{it} \qquad (S1)$$

where γ_t represents global factors affecting all countries equally, θ_t is a vector representing the covariate effects associated with the vector of observed covariates X_i, λ_t generalizes the model to include a vector of unobserved common factors and their loadings μ_i, and ϵ_{it} is the country-specific error term. We abbreviate the SCM algorithmic procedure using the representation of a multi-dimensional projection of $Z_{i,t}$ onto the complementary vector space of control time series given by $Z_{j,t}$. In this way, the normalized weights can be conceptualized as

[2] For the total number of cross-border documents, we estimated the model using $\log_{10}\chi^{s}_{i,t}$ which is less sensitive to large deviations in scale across the control countries as well as the EU countries. We then exponentiated the SCM results in order to estimate the difference δ(%) and plot the results in Figs. 2 and fig. S3.

[3] Because the World Bank data for researcher population data is incomplete for many of the control countries, we were unable to include it without severely reducing the number of control countries (N_c).

$$w_j = \frac{\langle Z_{i,t}, X_{i,t} | Z_{j,t}, X_{j,t} \rangle}{\sum_j \langle Z_{i,t}, X_{i,t} | Z_{j,t}, X_{j,t} \rangle} \in [0,1] \qquad (S2)$$

which satisfy $\sum_{j=1}^{N_c} w_j = 1$. The SCM algorithm then finds the optimal weight vector $\boldsymbol{w}^*$ that sufficiently satisfies the following equalities

$$\begin{aligned} \sum_{j=1}^{N_c} w_j^* \, Z_{j,t} &= Z_{i,t}, \text{for } t \in [1996,2003] \\ \sum_{j=1}^{N_c} w_j^* \, X_j &= X_i \qquad (S3) \end{aligned}$$

This method is reliable as long as the number of number of periods prior to 2004 (i.e. 7 years in our case) is large with respect to the timescale of ϵ_{it}. For the longhand description and derivation of the SCM, with application to the 1988 California tobacco control program (Proposition 99) in the USA, we refer the interested reader to Abadie, Diamond, and Hainmueller (*26*).

Using the optimal weighted coefficients w^* which best reproduce the actual $Z_{i,t}$ for $t < 2004$, the weighted linear combination is extrapolated for $t \geq 2004$, thereby producing the counterfactual time series $\hat{Z}_{i,t}$. This method is well-suited for this policy intervention scenario because it accounts for the global trends in cross-border collaboration already existing before and persisting after 2004, as captured by γ_t (implicit in the non-EU global control set).

We now return to the two scenarios of interest, first where the outcome variable is the fraction of publications that involved cross-border collaboration, $Z_{i,t} \equiv f_{i,t}^s$, and in the second case where the outcome variable is total number of cross-border publications $Z_{i,t} \equiv \chi_{i,t}^s$. In both cases we measure the "EU entry effect" by computing the difference in the post-2004 totals, $Z_i^> = \sum_{t=2005}^{2012} Z_{i,t}$ and $\hat{Z}_i^> = \sum_{t=2005}^{2012} \hat{Z}_{i,t}$. In the case of $f_{i,t}^s$ we define the post-entry difference as a difference in means, $\delta = (\hat{f}^> - f^>)/(2012 - 2005 + 1)$, and in the case of $\chi_{i,t}^s$ we define the post-entry difference as a percent difference, $\delta(\%) = 100 \times (\hat{\chi}^> - \chi^>)/\chi^>$.

For the case of $f_{i,t}^s$, we observe opposite effects for the new and incumbent EU countries. Figure 2 shows $\delta > 0$ values for the 2004 entrant EU countries and $\delta < 0$ values for the incumbent EU countries. This pattern is robust for three different estimations: for all subject areas aggregated ($s =$All), as well as for the individual subject areas $s = 1300$ representing "Biochemistry, Genetics, and Molecular Biology" (Biology), and $s = 3100$ representing "Physics and Astronomy" (Physics), the two most collaborative subject areas. The diverging trends provide a key insight into the substitution effect due to high-skilled mobility: had there been no enlargement, the counterfactual number of intra-border publications $(D_{i,t}^s - \chi_{i,t}^s)$ would have decreased relative to $\chi_{i,t}^s$ for the incumbent EU countries because there would have been more researchers to potentially collaborate with abroad. However, since the net flow of high-skilled mobility was towards the pre-2004 EU countries – contributing to their stock of internationally reputable scientists along with their international connections – this left the new 2004 EU entrant countries at a loss of international collaboration opportunities.

Figure S2 shows the SCM applied to each new EU country individually, with $\delta > 0$ for 8 of the 12 countries; the mean and standard deviation of the individual values are $\langle\delta\rangle \pm \sigma_\delta = 0.044 \pm 0.068$. This value is consistent with the EU Entry effect coefficient $\beta_T = -0.058$ estimated in the Difference-in-Difference model (see table S1).

The case of $\chi_{i,t}^{s}$ further demonstrates negative effect on the intensity of Europe's science integration, as measured by cross-border collaboration. For both incumbent and new EU countries, there would have been more cross-border publications had there been no EU enlargement. For example, for all subject areas aggregated ($s =$All), we calculated a $\delta(\%) = 15$ counterfactual effect for the average incumbent EU country, and a $\delta(\%) = 9$ percent effect for the average entrant country (see Fig. 2). These results were also consistent when applying the method to just the Biology and Physics subject area data.

Figure S3 shows the SCM applied to each new EU country individually, with $\delta(\%) > 0$ for 11 of the 12 countries; the mean and standard deviation of the individual values are $\langle\delta(\%)\rangle \pm \sigma_{\delta(\%)} = 22 \pm 29$ percent. Interestingly, the only country with $\delta(\%) < 0$ is Cyprus, which our mobility analysis revealed as one of the countries with the largest relative inflow of high-skilled labor after the 2004 enlargement.

High-skilled mobility in Europe: 1997–2012

The rate of international collaboration has been increasing as a result of globalization, with a large contributor to this trend being the countries with smaller science programs which integrate with large R&D hubs (*3*, *8*, *9*). As such, over the 1997–2012 period of analysis, we also observe an increase in the per-publication cross-border collaboration rate $f_{i,t}$, especially during the early 2000s. For the incumbent EU countries, the mean (averaged over countries and 14 subject areas) cross-border collaboration rate before and after 2004 were $\langle f_< \rangle = 0.41$ and $\langle f_> \rangle = 0.53$ (significantly different mean values, with difference-in-means Student T-test p-value = 10^{-14}); for the 2004 non-EU countries, the mean cross-border collaboration rates before and after 2004 were $\langle f_< \rangle = 0.42$ and $\langle f_> \rangle = 0.46$ (significant difference-in-means, Student T-test p-value = 0.001). The increase in $f_{i,t}^{All}$ is stronger for the incumbent EU countries (fig. S4A), whereas the trend is significantly weaker for the EU enlargement countries (fig. S4B). In this latter case, if the non-EU countries are excluded from the difference between pre- and post-2004 levels, there is less evidence of any increase over the two periods for the new EU entrants. Interestingly, the notable increase between 2002 and 2003 may be attributable to EU Framework Programme (FP6) funding initiatives introducing explicit cross-border collaboration requirements.

Meanwhile, the rate of high-skilled mobility between EU members also increased over the same period (see (*18, 48*) for an in-depth review of the impact of EU enlargement on labor mobility). However, the incoming and outgoing rates for each country are typically not equal, representing a large-scale reorganization of high-skilled labor in Europe. While the in-to-out mobility ratio $I_{i,t}/O_{i,t}$ for the incumbent EU countries was distributed more evenly above and below unity (fig. S4G), $I_{i,t}/O_{i,t}$ is mostly less than unity for the 2004 non-EU countries (fig. S4H), representing the mobility imbalance between eastern and western Europe.

Figure 3 shows the high-skilled mobility matrix, before ($M_{ij,<}$) and after ($M_{ij,>}$) the 2004 enlargement. The countries are ordered according to decreasing B_i calculated across the entire period 1997–2012: the country with highest rate of outgoing mobility (largest B_i) was HR, and contrariwise, the country with smallest B_i was CY. In order to visualize the pairwise mobility counts, which can range across several orders of magnitude, we show $\log_{10} M_{ij}$. Comparing the periods 1997–2004 to 2005–2012, the total mobility across all countries increased roughly 7-fold, from $M_{<} = 43{,}075$ (1997–2004) to $M_{>} = 272{,}813$ (2005–2012). The significant increase in high-skilled labor mobility was distributed across all the European countries, thereby resulting in a reorganization of the entire mobility network, as some countries transitioned from being major sources to major sinks of high-skilled labor (e.g. CH). One constant across the two time periods is the role played by the United Kingdom as the major mobility hub, which benefited from the EU enlargement, going from a relatively small sink before the enlargement ($B_{UK,<} \approx -0.1$), to a relatively large sink afterwards ($B_{UK,>} \approx -0.6$).

In order to better visualize the sources and the sinks, in Fig. 4 we plot the net mobility $\Delta_{ij} = M_{ij} - M_{ji}$. This matrix visualization only shows the $\Delta_{ij} > 0$ entries, thereby facilitating the visual inspection of the significant net emigration ('brain drain') sources and the significant net immigration ('brain gain') sinks. For example, Norway was a major immigration sink, before and after the enlargement, drawing largely from her Scandinavian neighbors, as well as Poland, Germany, the United Kingdom, and Ireland. Interestingly, Cyprus also stands out as an immigration sink, after the enlargement, with the total incoming mobility growing by a factor of $I_{CY,>}/I_{CY,<} \approx 100$. This growth is likely due to Cyprus' pre-financial crisis status as a tax haven, making it attractive for high-skilled professionals. We calculated the minimum spanning tree representation of Δ_{ij} (fig. S9 bottom) which further emphasizes the central role played by UK as a major sink in the mobility network.

Identifying national communities according to empirical migration networks provides insight into the role of geographic and cultural proximity within Europe. In order to cluster the countries into groups, we aggregated the data across all years (i.e. 1997–2012, due to the sparsity of the non-EU flows before 2004) and we then applied the modularity maximization algorithm (*49*) to both mobility networks M_{ij} and Δ_{ij}. Figure S6 shows that in both cases, there were 3 communities identified. Moreover, there is little variation between the similar communities in the clustered M_{ij} and Δ_{ij} networks, indicating a level of satisfactory robustness in the clustering outcome. Interestingly, most of the EU enlargement countries are contained in the yellow group, with Germany as its central hub. Thus, DE appears as a major entry point for high-skilled mobility from the new Eastern-Europe EU members, possibly reflecting its historical role as the entry point during the era of the Eastern bloc, whereas UK draws mostly from its northern neighbors (for closer visual inspection also see also fig. S5).

We also measured the High-skilled mobility relative to total migration rates estimated by Abel & Sander (*15*) over the same periods. Figure S8 shows the matrix of mobility ratios, with each element representing the pairwise ratio $\rho_{ij} \equiv M_{ij,x} / \widetilde{M}_{ij,x} \in [0,1]$. The $\rho_{ij,<}$ matrix indicates that Spain had a relatively large outward migration of high-skilled labor before as well as after 2004. Other countries with large mean outgoing ρ_{ij} values after 2004 are EE, PL, HU, MT, PT, and IT. Similarly, countries with large incoming ρ_{ij} values after 2004 are IE, IS, UK, NO, BE, DE, and DK. At the aggregate level, the probability distribution $P(\rho_{ij})$ indicates that the ratios span a wide range, with average value 0.022 before and 0.14 after 2004, indicating a 6-fold increase in the fraction of total migration attributable to high-skilled laborers in the latter period.

The regulated professionals mobility data also contains the application success rates for professional license transfer. Because application approval is a precondition for migration, it serves as an additional quantitative indicator of each country's competitiveness ($\mathcal{P}_i^{out}$, as in the case of outgoing mobility) and selectivity ($\mathcal{P}_i^{in}$, as in the case of incoming mobility). Figure S10 shows the mobility polarization before ($B_{i,<}$) and after ($B_{i,>}$), and net rates $\mathcal{P}_{i,x}^{in/out}$, before and after the 2004 enlargement. Interestingly, Cyprus and the Czech Republic, two of the wealthiest countries over the entire study period in terms of per capita $GDP(PPP)$ (constant 2005 international dollars), are the only two enlargement countries with $B_i < 0$ before and after 2004. In the case of CZ, this is largely owing to its relatively high incoming success rate, $\mathcal{P}_{>,<}^{in}$. Countries with a notable decrease in their "labor import" selectivity, corresponding to a significant increase in $\mathcal{P}_i^{in}$, are GR, DE, and PL. Countries with a notable increase in their "labor export" competitiveness, corresponding to a significant increase in $\mathcal{P}_i^{out}$, are AT, CH, FR, IT, LT, LV, PT and SI; CY, BG and RO are two countries with a notable decrease in competitiveness as "high-skilled labor exporters".

To further identify dyadic relations in the labor export selectivity and competitiveness of the countries, fig. S11 shows the mobility acceptance rate matrix $\mathcal{P}_{ij}$ before and after the 2004 enlargement. This representation indicates that Norway, Italy, and Poland's low incoming success rate is largely due to just a few countries, Greece's low incoming success rate after the enlargement is low across the board – possibly indicative of bureaucratic inefficiencies. The difference in outgoing and incoming success rates, $\overline{\mathcal{P}}_{ij}^{out} - \overline{\mathcal{P}}_{ij}^{in}$, identifies countries with mismatches in competitiveness and selectivity, indicative of labor market inefficiencies within the European's "single market" (*6*).

Panel regression model for measuring the "EU Entry Effect"

We implement a difference-in-difference (DiD) identification strategy similar to studies measuring the impact of economic or political regime change (e.g. liberalization in the former, or democratization in the latter case) on a country's economic growth (*50*, *51*). Specifically, we use a panel regression to estimate the impact of cross-border mobility and EU enlargement on the per-publication rate of cross-border activity, $f_{i,t}^S$, a proxy for European science integration.

The DiD interaction represents the cross-term between EU membership and the country's entry year, thereby measuring the impact of a change in EU membership status on a new member state's rate of international collaboration. In this way, the control group consists of the countries that did not change their EU membership status over 1997–2012 (members of country groups $g_{EU,i} = 1$ and 4), and the treated group are those that did change their EU membership status over 1997–2012 (members of country groups $g_{EU,i} = 2$ and 3). We estimated the parameters of the following linear panel data model with country fixed-effects, which controls for scientific productivity and impact, R&D investment, high-skilled mobility, total migration in particular, and research subject area

$$
\begin{aligned}
f_{i,t}^{s} \quad = & \quad \beta_t t + \beta_T T_{EU,i,t} + \{\beta_D \log_{10} D_{i,t}^{s} + \beta_R R_{i,t}^{s} \\
+ & \quad \beta_E \log_{10} E_{i,t} + \beta_{GDPpc} \log_{10} GDPpc_{i,t} + \beta_{Spc} \log_{10} Spc_{i,t} + \\
+ & \quad \beta_B B_{i,t} + \beta_I \log_{10} I_{i,t}^{+} + \beta_{P(in)} \mathcal{P}_{i,t}^{in} + \beta_O \log_{10} O_{i,t}^{+} + \beta_{P(out)} \mathcal{P}_{i,t}^{out} + \beta_{G(in)} G_{i,t}^{in} + \beta_{G(out)} G_{i,t}^{out} \\
+ & \quad \beta_{\widetilde{B}}\, \widetilde{B}_{i,\tau} + \beta_{\widetilde{O}} \log_{10} \widetilde{O}_{i,\tau} + \beta_{\tilde{I}} \log_{10} \tilde{I}_{i,\tau} + \beta_{\widetilde{G}(in)}\, \widetilde{G}_{i,\tau}^{in} + \beta_{\widetilde{G}(out)}\, \widetilde{G}_{i,\tau}^{out} \} \\
+ & \quad \vec{\beta}_s \cdot \overrightarrow{SA}\,(s) + \beta_{i,0} + \epsilon_{i,t} \\
= & \quad \beta_t t + \beta_T T_{EU,i,t} + \{\vec{\beta} \cdot \vec{x}_{s,i,t}\} + \vec{\beta}_s \cdot \overrightarrow{SA}\,(s) + \beta_{i,0} + \epsilon_{i,t} \qquad (S4)
\end{aligned}
$$

The “EU Entry” effect is estimated using the indicator value $T_{EU,i,t}$ capturing the EU-vs-nonEU and before-vs-after cross-term: it is 1 for countries belonging to the EU in year t and 0 otherwise. Thus, there are three groups of countries: (i) the incumbent EU countries with $T_{EU,i,t} = 1$ for all t, (ii) the group of new entrants with a transition from $T_{EU,i,t} = 0$ to $T_{EU,i,t} = 1$ in $t = 5$ for the ten 2004 entrants (CY, CZ, EE, HU, LT, LV, MT, PL, SK ,SI), and $t = 8$ for the two 2007 entrants (BG, RO), and (iii) the three Eurozone countries (CH, HR, and NO) that were not part of the EU as of the end of 2012 with $T_{EU,i,t} = 0$ for all t.

Figure 1 provides a first justification for our Difference-in-Difference approach, as the f_t for Western and European are approximately equal up to a constant offset (i.e. equal slopes) for $t < 2004$. In order to provide an additional consistency check and to justify our identification strategy, we define the EU entry group as $g_{ENTRY} = \delta_{g_{EU,i},2} \cup \delta_{g_{EU,i},3} \cup \delta_{i,HR}$, which is a variable equal to 1 if the country is in EU country group 2, 3 or is HR. We then ran the model in Eq. S4 with an additional interaction term between g_{ENTRY} and the year, given by $X_{i,t} = g_{ENTRY} \times dummy_t$. The coefficient ψ_t of the interaction term, shown in fig. S7, estimates the annual impact of (eventual) entry on the countries belonging to g_{ENTRY}. With the exception of the value $\psi_{2001/2002}$, the ψ_t values are only significant negative after the baseline year $t \equiv 2005/2006$. Thus, this result rules out other factors prior to 2004 that could have also contributed to our estimation of the mobility and EU enlargement effects.

Results of partial regression models as robustness check

Table S1 shows the parameter estimates for partial models (A–E) that do not include one or more of the data types (Scientific productivity and impact, R&D investment, High-skilled mobility, Total migration). We also ran the same regression as the Full model, however restricting the data to the two periods before and the two periods after 2004 (4-period model F). This 4-period model better satisfies the difference-in-difference model condition that the number of countries (units) be much larger than the number of time periods analyzed, $31 = |i| \gg |t| = 4$. In all, the coefficients estimated across all model estimates shown in table S1 are consistent in magnitude, sign, and significance, demonstrating the full model’s robustness.

In addition to T_{EU} and $B_{i,t}$, there are several other parameters which are of particular interest. First, an increasing total incoming mobility I_t^{+} is related to smaller f_t ($\beta_I < 0$, $p \leq 0.026$ in all regressions), consistent with the mobility mechanism whereby countries receiving foreign high-skill labor are at the same time losing the cross-border activity that was previously being channeled across the same foreign collaborator. This effect was also observed for the total mobility data ($\beta_{\tilde{I}} < 0$, $p \leq 0.005$ in all regressions). If, however, the foreign collaboration channel is maintained, then the cross-border activity is sustained. Thus, we observe that countries with higher outgoing mobility O_t^{+} have higher f_t ($\beta_O > 0$, $p \leq 0.034$ in regressions A–E). Thus, an important caveat is whether or not the cross-border mobility results in the termination of cross-border activities, a causal effect that we are not able to estimate given the limitations of our data.

Second, the model indicates that more concentrated (non-uniform) distribution of outgoing mobility (larger G_t^{out}) is related to larger f_t ($\beta_{G(out)} > 0, p \leq 0.050$ in all regressions). This effect is consistent with the maintenance and investment in the cross-border activities among the core of more selective countries, principally the old EU members, who are characterized by a less-dispersed outward mobility (see Fig. 3 and fig. S9).

Third, among the two scientific productivity and impact covariates we included, we observe a negative relation between the quantity of scientific output ($\beta_D < 0, p \leq 0.027$ in all regressions) implying a saturation effect in the capability to collaborate internationally. More importantly, we confirm the prestige effect represented by the citation impact R_t of each country ($\beta_R > 0, p \leq 0.001$ in all regressions), capturing the positive feedback between reputation and the formation of collaborative activities across countries.

And finally, the subject area controls indicate that “Biochemistry, Genetics, and Molecular Biology” (1300) and “Physics and Astronomy” (3100) are the most collaborative domains, with “Agricultural and Biological, Sciences” (1100), “Chemistry” (1600), “Materials Science” (2500), and “Medicine” (2700) forming a middle group, and the rest of the subject areas comprising a third relatively “low-collaboration” subset. The high-f_t group of biology and physics is largely due to the emergence of large team science stemming from globalizing endeavors (e.g. European Organization for Nuclear Research – CERN) and initiatives (e.g. the Human Genome Project, ENCODE) (*3*).

Supplementary Figures and Tables

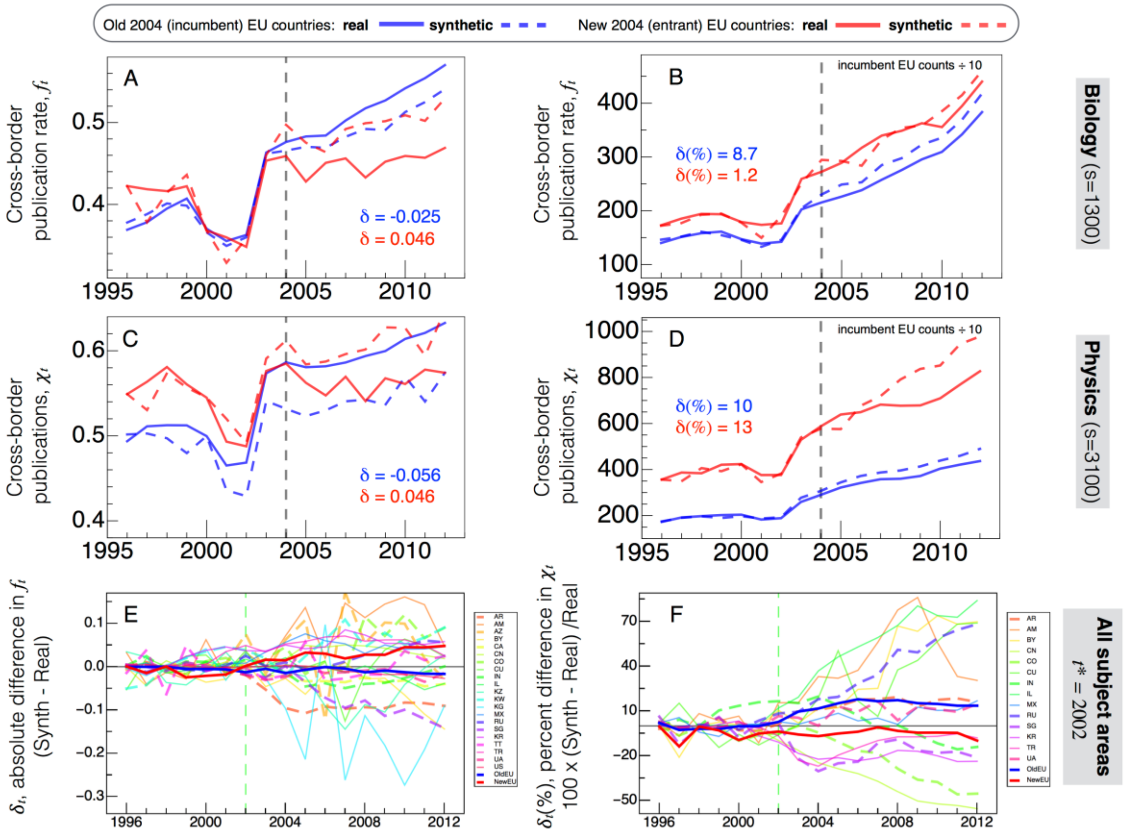


fig. S1. Supplementary SCM results. (**A**–**D**) Synthetic control method applied to data for two highly collaborative disciplines, biology and physics. Direct comparison to Fig. 2(A, B) shows that the values of the counterfactual difference δ and $\delta(\%)$ are similar in sign and magnitude, with the exception of $\delta(\%)$ for biology, which is smaller for both entrant and incumbents, and marginally nonzero for the entrants. (**E**, **F**) Results of the "permutation test" in time for all subject areas pooled together.
(E) Reproduction of the SCM "permutation test" shown in panel Fig. 2C using instead a "placebo" intervention year $t^* = 2002$. Because f_t is an intensive variable, plotted for each country is the time series representing the absolute difference $\hat{f}_t - f_t$. In this case, the entrant's curve has the 9th-most positive difference, thereby failing to demonstrate a significant counterfactual difference with respect to the results of the other control countries. (F) Reproduction of the SCM "permutation test" shown in panel Fig. 2D using instead a "placebo" intervention year $t^* = 2002$. Because χ_t is an extensive variable, plotted for each country is the time series representing the percent difference $100(\hat{\chi}_t - \chi_t)/\chi_t$. The results of this case also fail to indicate a significant counterfactual difference relative to the control countries. To be clear, each curve in (E, F) corresponds to the SCM result estimated for each individual country using the remaining control countries. The countries that are not shown in each panel failed to pass a SCM goodness-of-fit criteria for $t < t^*$ based upon the mean squared error between the synthetic and real curve.

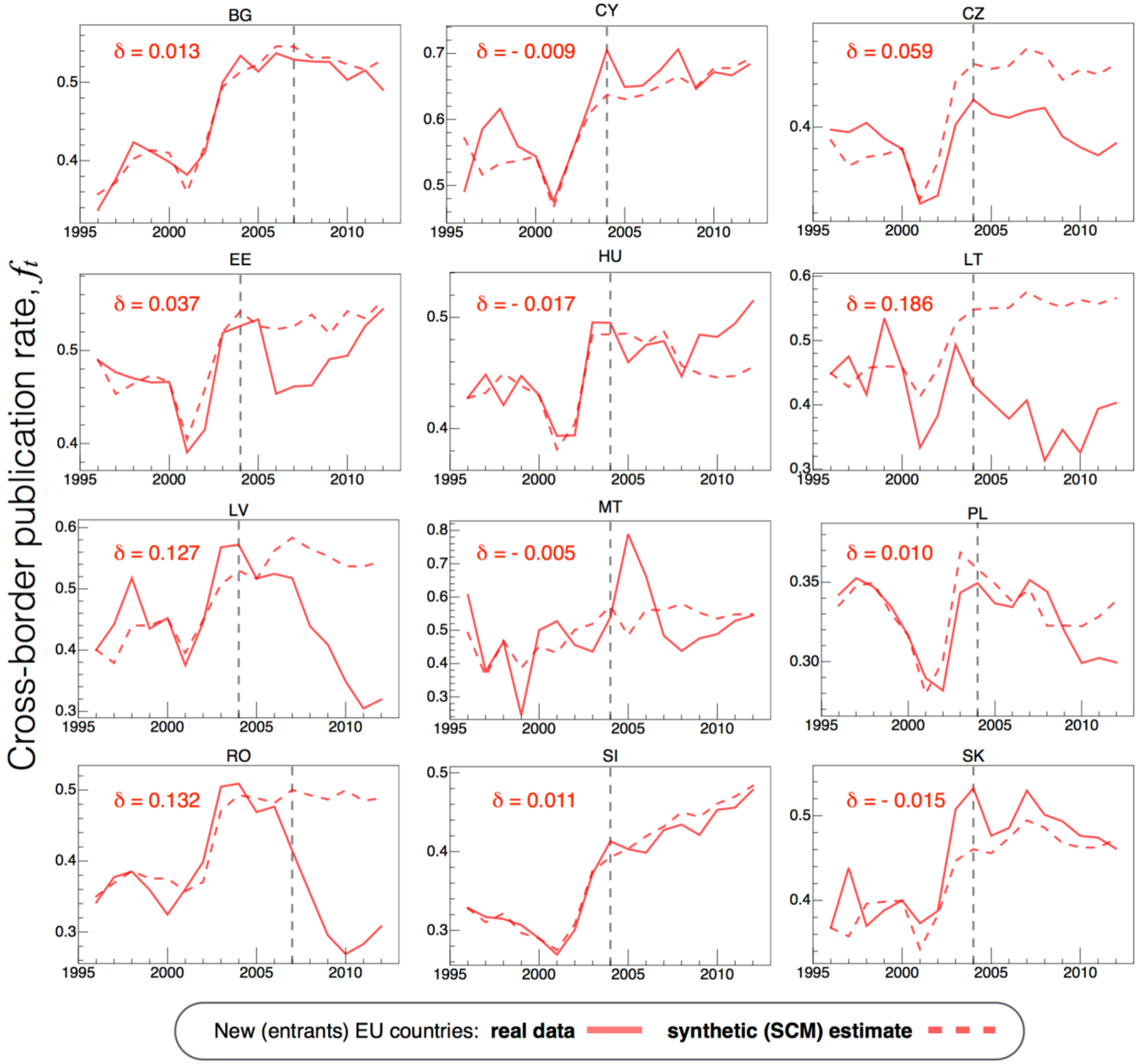


fig. S2. SCM: cross-border publication rate *f*. Counterfactual estimates of the fraction of the publications that are cross-border, $\hat{f}_i$, for each of the 2004 and 2007 new EU entrants. Each solid line indicates the observed number of cross-border publications by year. Each dashed line indicates the synthetic estimates had the country *not entered* the EU. The synthetic control group is comprised of 26 non-European countries. The SCM explanatory variables used to estimate $\hat{f}_i$ are the total number of publications ($\log_{10} D_{i,t}^{All}$), the normalized citations ($R_{i,t}^{All}$), the per-capita GDP ($\log_{10} GDPpc_{i,t}$), and government expenditure on R&D, $e_{i,t}$; "All" indicates the total across all subject areas (s). δ is the mean difference between the curves after the entry year (indicated by each dashed vertical line), serving as a basic estimates of the net impact of the 2004 enlargement on each country. $\delta > 0$ for 8 yearsof the 12 countries; the mean and standard deviation of the individual values are $\langle \delta \rangle \pm \sigma_\delta = 0.044 \pm 0.068$.

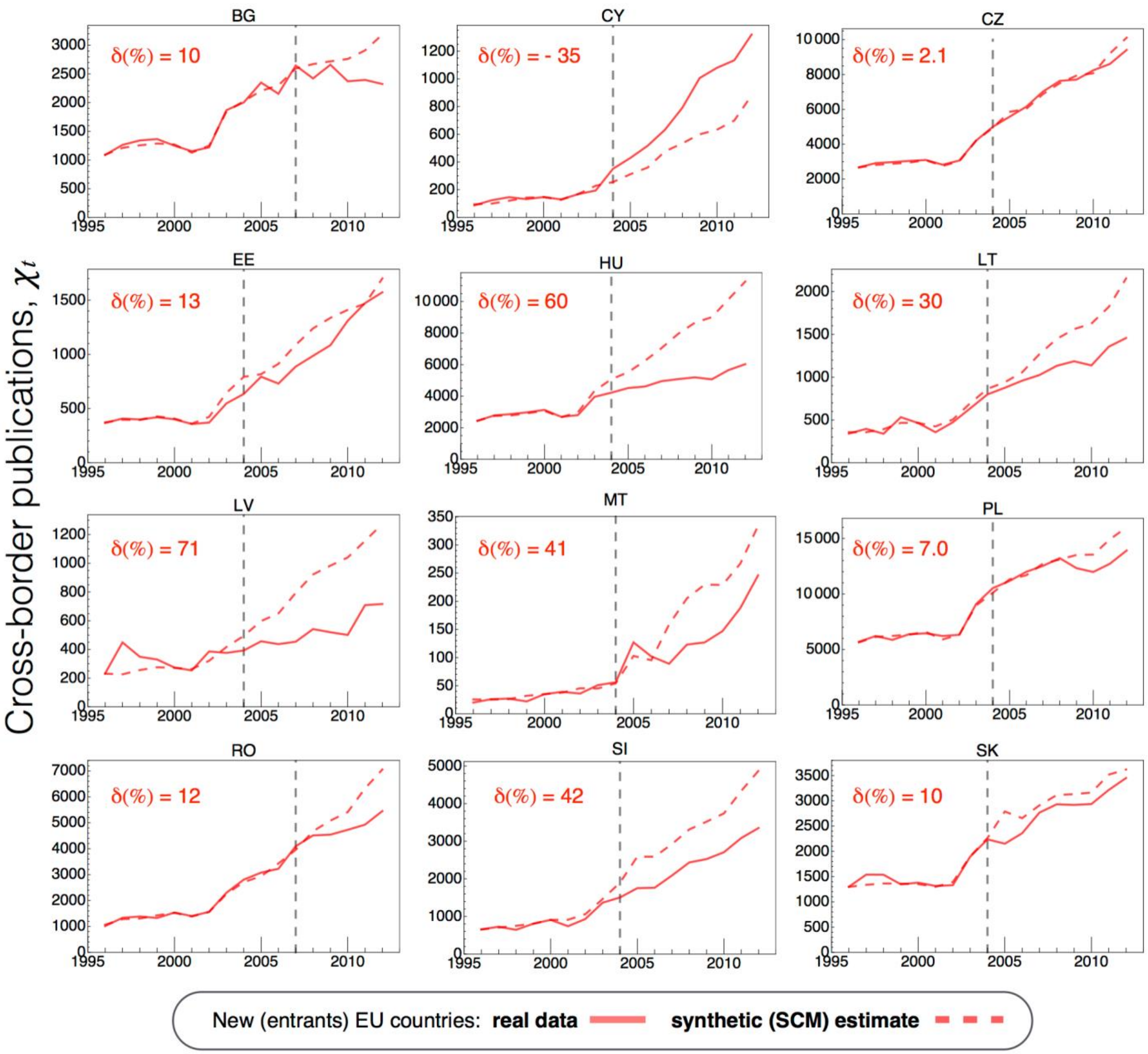


fig. S3. SCM: total cross-border publications χ. Counterfactual estimates of the number of the publications that are cross-border, $\hat{\chi}_i$, for each of the 2004 and 2007 new EU entrants. Each solid line indicates the (real) observed number of cross-border publications by year. Each dashed line indicates the synthetic estimates had the country *not entered* the EU. The SCM explanatory variables used to estimate $\hat{\chi}_i$ are the total number of publications ($\log_{10} D_{i,t}^{All}$), the normalized citations ($R_{i,t}^{All}$), the per-capita GDP ($\log_{10} GDPpc_{i,t}$), and government expenditure on R&D, $e_{i,t}$; "All" indicates the total across all subject areas (s). $\delta(\%)$ is the percent difference between the net area under the real and synthetic curves after the entry year (indicated by each dashed vertical line), serving as a basic estimates of the net impact of the 2004 enlargement on each country. $\delta(\%) > 0$ for 11 of the 12 countries; the mean and standard deviation of the individual values are $\langle \delta(\%) \rangle \pm \sigma_{\delta(\%)} = 22 \pm 29$ percent.

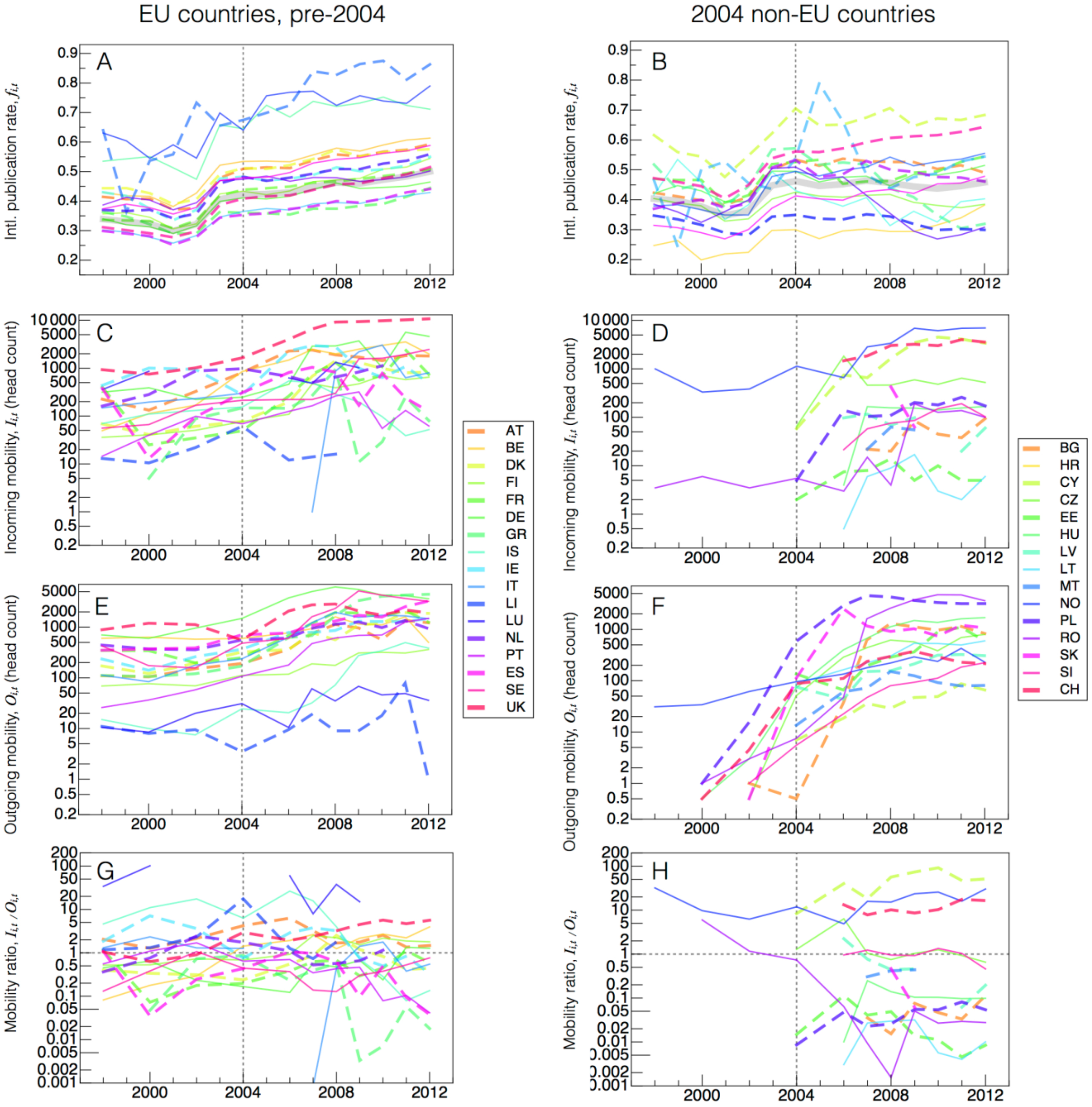


fig. S4. International collaboration rates and high-skilled labor mobility. (left column) 2004 incumbent countries and (right column) non-EU countries. (**A**, **B**) Rate of international collaboration, $f_{i,t}$ (per publication), representing a weighted mean calculated across all subject areas (s). The opaque grey curve represents the average over all countries within each group, indicating a notable post-2003 saturation in the case of the 2004 entrant countries. (**C**, **D**) Incoming mobility counts, $I^{+}_{i,t}$. (**E**, **F**) Outgoing mobility counts, $O^{+}_{i,t}$. (**G**, **H**) Ratio of incoming to outgoing mobility, $I^{+}_{i,t}/O^{+}_{i,t}$. Note that IS and LI, neither of which are EU members but rather European Economic Area members, are included with the 2004 incumbents in order for visual parity.

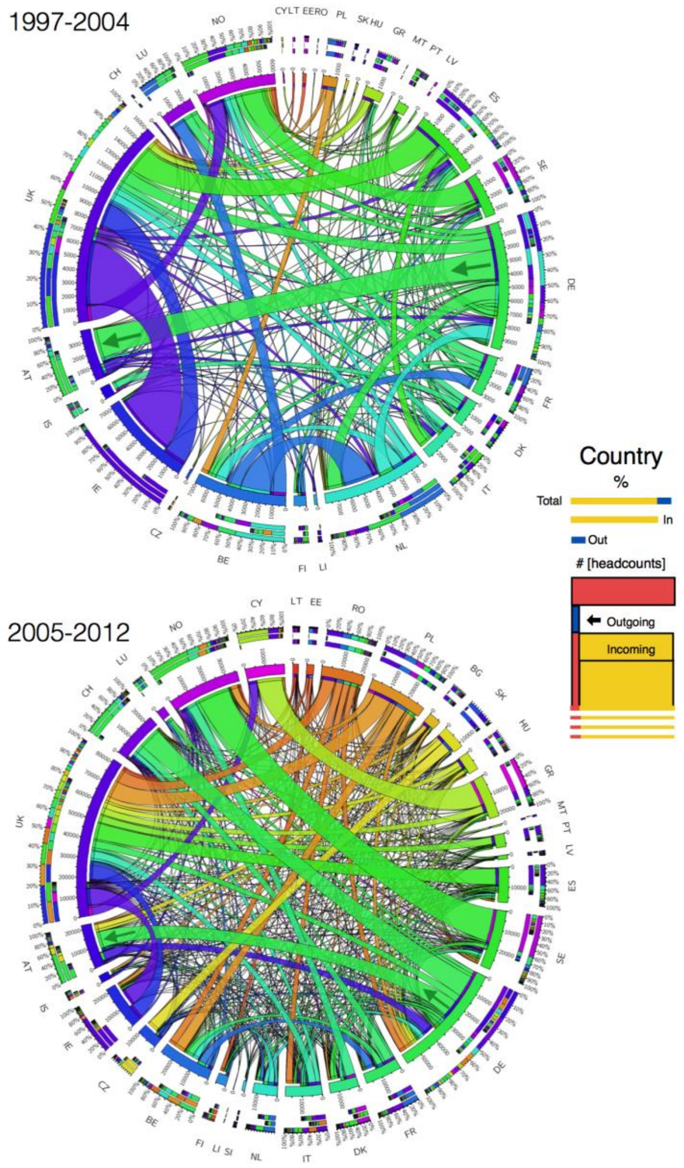
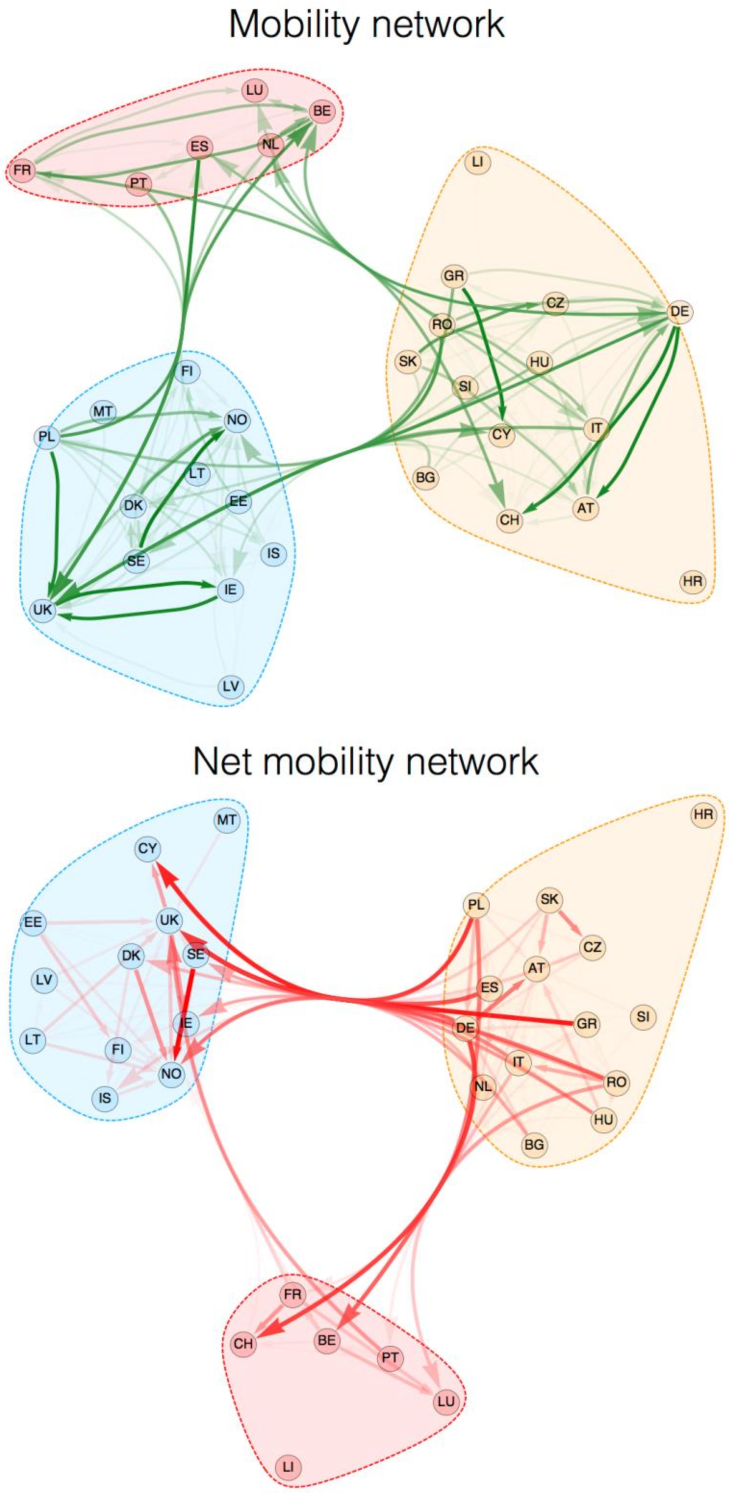


fig. S5. Country-country mobility networks before and after the 2004 enlargement. Each country is represented on the circumference by a colored arc, where the arc-length is proportional to the total incoming and outgoing mobility. The ribbons between each country are proportional to the mobility M_{ij}. The mobility direction is encoded in the color of the ribbon, which is the same as the destination country, as well as the endpoint characteristics of the ribbon, denoted by the gap between the ribbon and the termination arc. The legend provides a schematic example of a country which receives incoming mobility from just a single (yellow) country, and is the source of outgoing mobility for just a single (blue) country. Altogether, the mobility of each country can be summarized by the 3 circumscribing histograms: the outer-most arc represents the total distribution of mobility by all partner countries, the middle arc represents the distribution of incoming mobility by source country, and the inner arc represents the distribution of outgoing mobility by destination country. Shown are only the links representing 20 or more mobility events, together accounting for 97% of the total mobility before 2004 and 99% of the total mobility after 2004. We thank the developers of the open-source *Circos* layout software (*52*) used to produce this network visualization.



Mobility network

Net mobility network

fig. S6. Community structure of the high-skilled mobility networks. Mobility data for the entire sample period, 1997–2012, are used to cluster countries into communities using Newman's modularity maximization algorithm (*49*). The modularity values of each graph are: 0.081 (top: links weights correspond to total mobility matrix M_{ij}) and 0.093 (bottom: links weights correspond to the net mobility matrix Δ_{ij}).

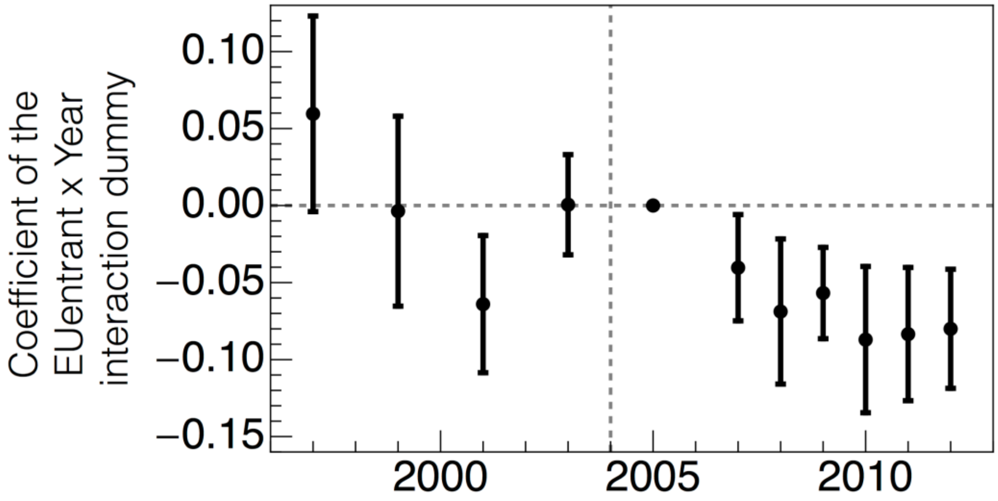


fig. S7. Consistency check for the significance of the EU entry effect. Values after the baseline year $t \equiv 2005/2006$ are significantly negative, whereas values prior to the baseline are not significant, with the exception of the value representing $t = 2001/2002$, possibly indicative of the impact of the 6th Framework Programme arising from the introduction of new cross-border collaboration requirements within the competitive grant funding program. Error bars represent the 95% confidence interval.

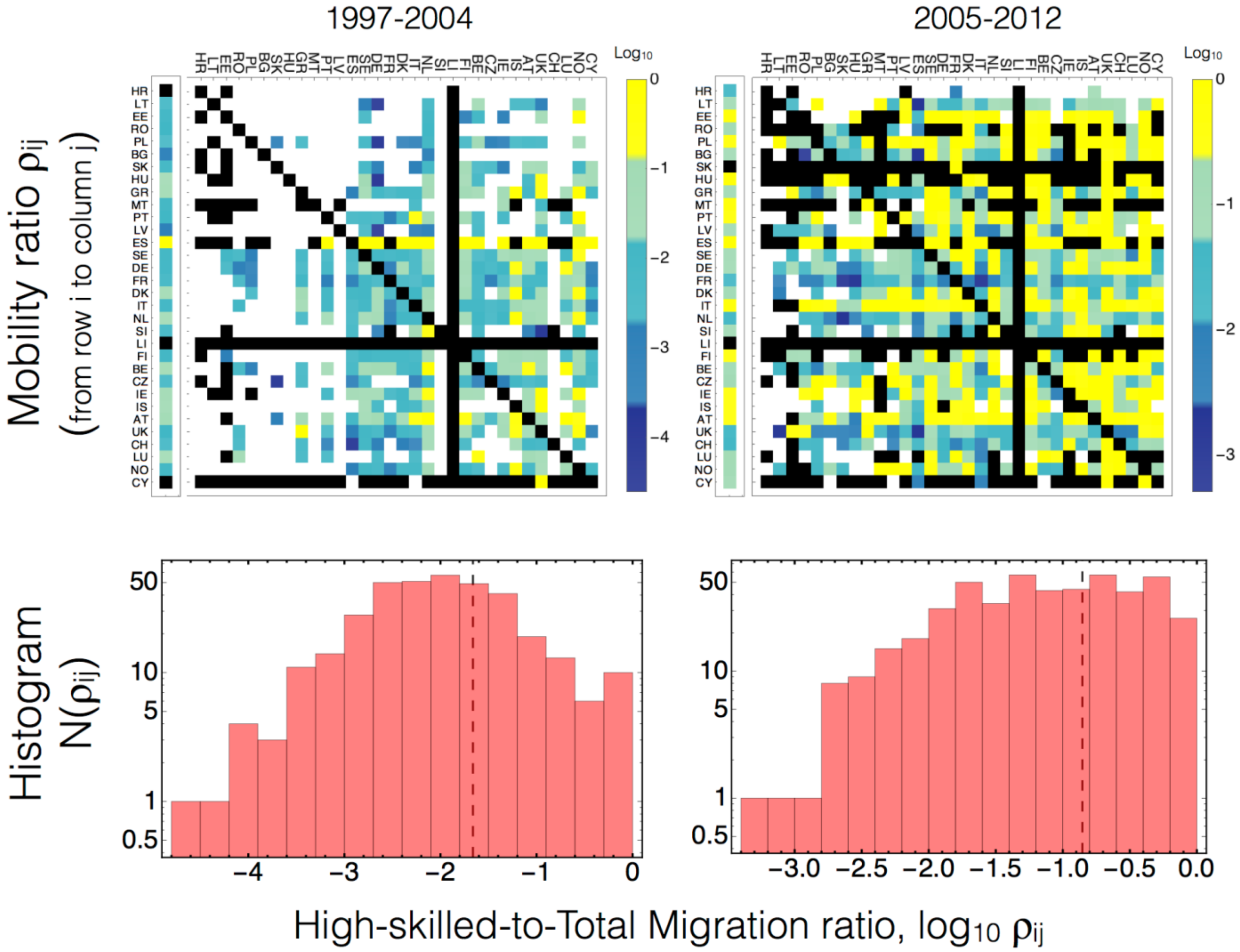


fig. S8. Comparison of high-skilled to total migration by country-country pair. Ratio of High-skilled mobility ($M_{ij,x}$) to the total migration ($\widetilde{M}_{ij,x}$) estimated by Abel & Sander: before and after the 2004 enlargement. (top) Matrix visualization of the mobility ratio $\rho_{ij,x} \equiv M_{ij,x}/\,\widetilde{M}_{ij,x} \in [0,1]$. Color scale indicated to the right separates the range of observed values into sextiles on logarithmic scale; White values indicate $\rho_{ij} = 0$ values ($M_{ij,x} = 0$ and $\widetilde{M}_{ij,x} > 0$); Black values indicate cases where $\widetilde{M}_{ij,x} = 0$; no total migration data available for LI (Liechtenstein). The color scale to the left of each matrix indicates the mean ratio value, $\langle \log_{10}\rho_{i,\tau}\rangle$, calculated for each row (only non-zero $\rho_{ij,x}$ are included in the calculation). The mean ratio value is provided as a visual aid to identify the countries with relatively large (e.g. Spain for 1997–2004) and small (e.g. France for 2005–2012) high-skilled mobility relative to total migration. (bottom) Count histogram $N(\rho_{ij,x})$ of the non-zero $\rho_{ij,x}$ values on $\log_{10}$ scale; vertical dashed line indicates the distribution average: 0.022 before and 0.14 after 2004.

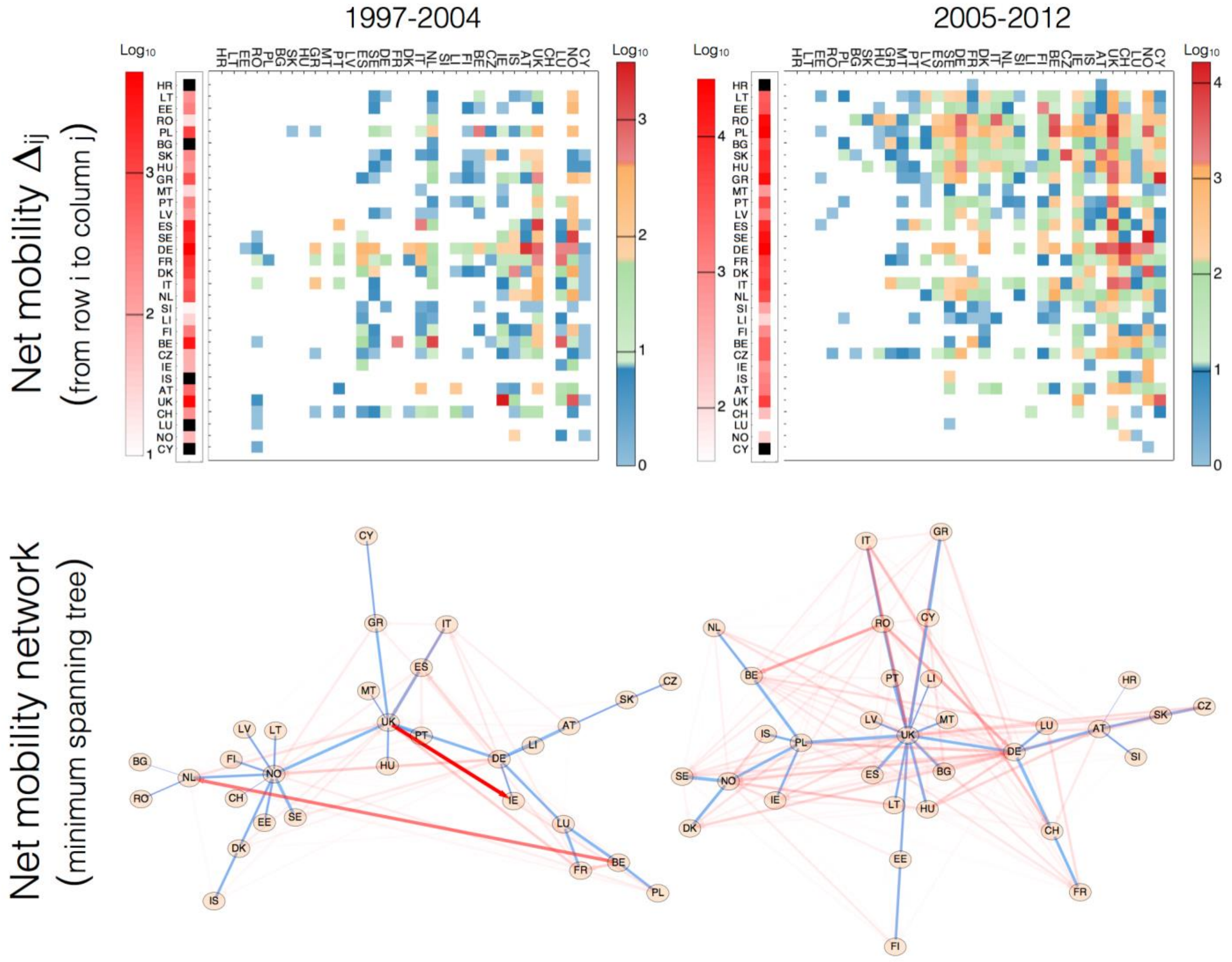


fig. S9. Net flow of high-skilled labor: before and after the 2004 enlargement. (top) The asymmetric net mobility matrix Δ_{ij} (head counts), showing only the matrix elements corresponding to net positive outflow from country (row) i to country (column) j ($\Delta_{ij} = M_{ij} - M_{ji} > 0$). The red color scale to the left of each Δ_{ij} matrix represents the net mobility out of country i calculated by summing the entries across each row (black cells indicates a net value < 10 for 1997–2004 and < 40 for 2005–2012). The color scale to the right of each Δ_{ij} matrix visualization represents the partitioning of the range of $\log_{10}\Delta_{ij}$ values into sextiles. Color values are not comparable across time periods. (bottom) Minimum spanning tree (MST) representation of each net mobility network, indicated by the blue links; the red links provide an overlay of the non-MST links. Color values are not comparable across time periods.

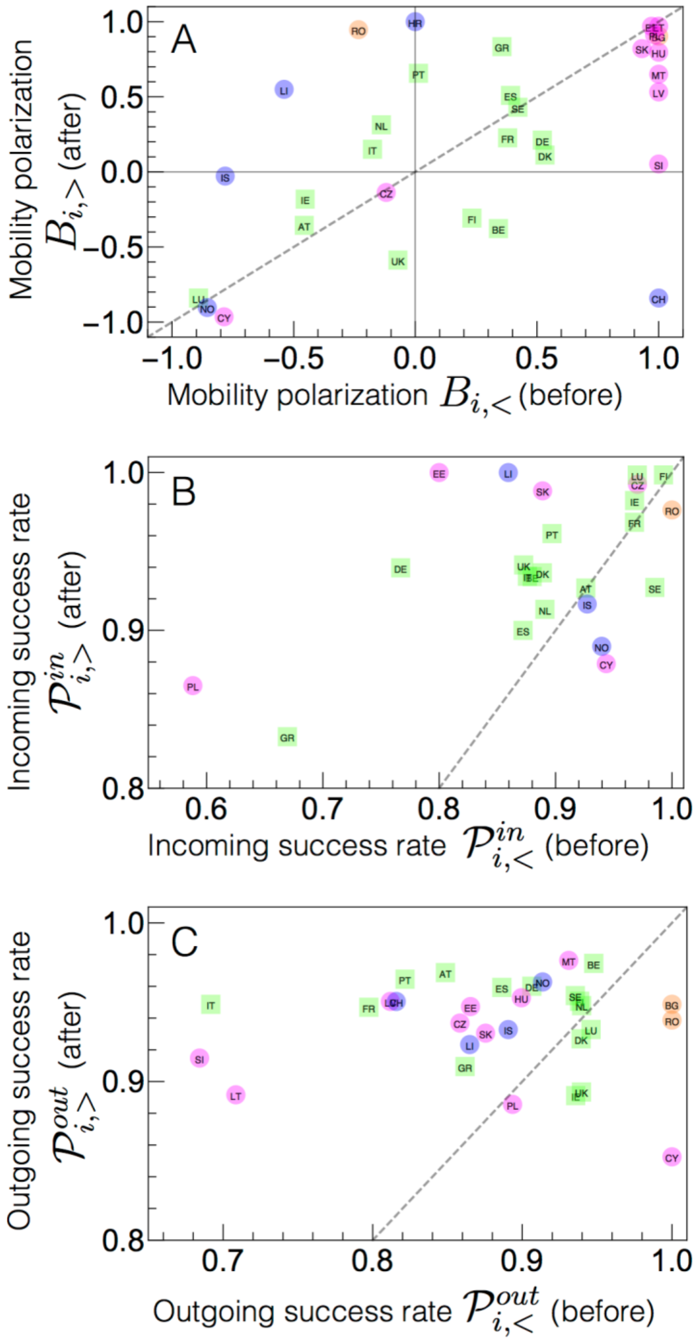


fig. S10. Mobility success rates by country: before and after the 2004 enlargement. (**A**) Mobility polarization B_i, (**B**) Incoming mobility success rate $\mathcal{P}^{in}_{i,<(>)}$, and (**C**) Outgoing mobility success rate $\mathcal{P}^{out}_{i,<(>)}$: before ($<$) and after ($>$) the 2004 enlargement. Countries are colored according to their EU membership status group $g_{EU,i}$. The dashed line corresponds to the diagonal $y = x$, and is shown to facilitate the visual inspection of countries which increased (above the line) or decreased (below the line) over the two time periods. In (B): Not shown are several of the new EU enlargement countries, which are clumped with $\mathcal{P}^{in}_{i,<} = 0$ and $\mathcal{P}^{in}_{i,>}$ between 0.9 and 1.0. In (C): Not shown is HR (Croatia), which has $\mathcal{P}^{out}_{i,<} = 0$ and $\mathcal{P}^{out}_{i,>} \approx 0.57$.

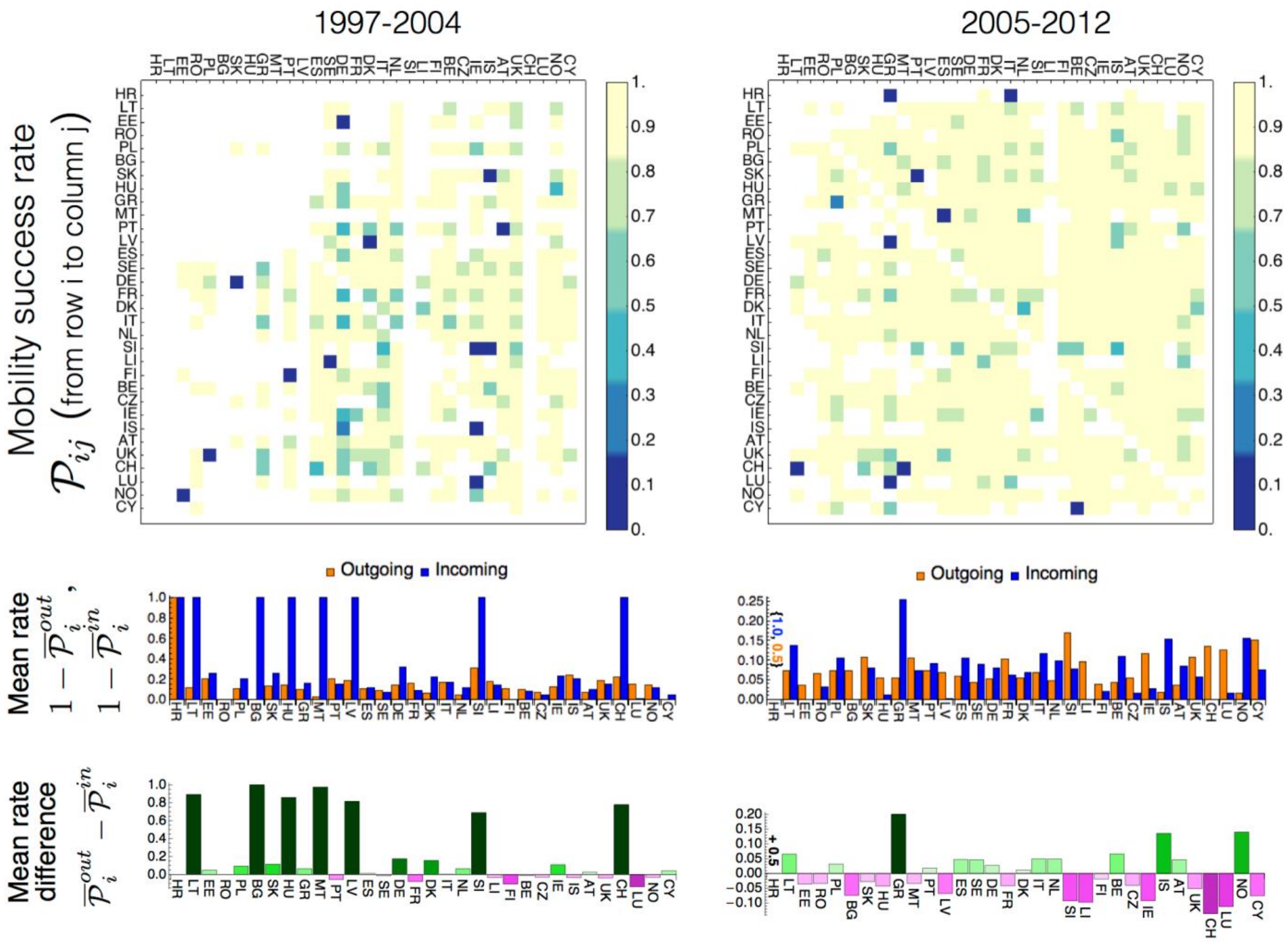


fig. S11. Mobility success rates by country-country pair: before and after the 2004 enlargement. High-skilled mobility success rates: before and after the 2004 enlargement. (top) Acceptance rate matrix $\mathcal{P}_{ij}$ (likelihood per application among those applications with either a positive or negative decision). The color scale for each matrix visualization represents a partitioning of the $\mathcal{P}_{ij}$ matrix entries into sextiles to facilitate visual inspection. (middle) The complement of the mean incoming/outgoing success rates, $\overline{\mathcal{P}}_{ij}^{out}$ $\overline{\mathcal{P}}_{ij}^{in}$, calculated as an average over the nonzero countries. (bottom) The difference in outgoing and incoming success rates $\overline{\mathcal{P}}_{ij}^{out}-\overline{\mathcal{P}}_{ij}^{in}$. Positive values indicate countries that are relatively competitive as high-skilled labor exporters. All color values are comparable across time periods.

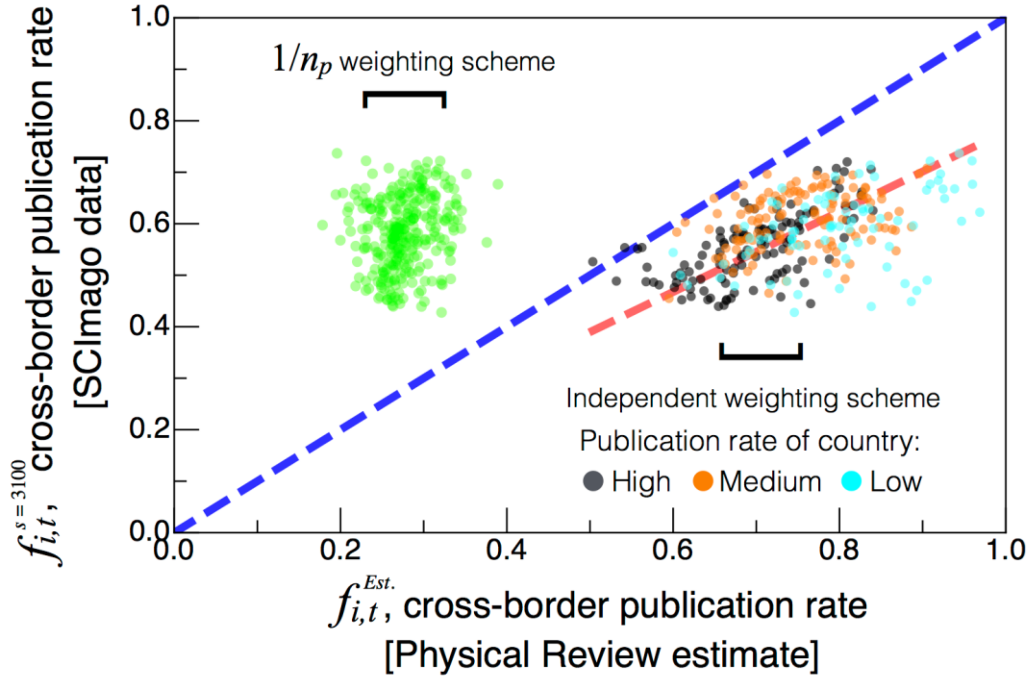


fig. S12. Validation of the SCImago cross-border counting scheme. Scatter plot, where dots correspond to international collaboration rate values for a given country in a given year. Compared are the "Physics and Astronomy (3100)" collaboration rate data, $f_{i,t}^{s=3100}$, obtained from SCImago and our estimates, $f_{i,t}^{Est.}$, using Physical Review journal data from the corresponding 13-year period 1997–2009. The diagonal dashed-blue line is the equivalence line corresponding to perfect agreement between each country-year observation. The green data corresponds to the estimates using an alternative counting scheme (method ii) where the weights are equipartitioned across the n_p countries associated with a given publication, i.e. inversely proportional to n_p. The black/orange/cyan data correspond to estimates using a n_p-independent weighting scheme (method i). In this latter case, which we confirm to be the method used by SCImago, we separated the data into three groups according to the average publication rate of each country in the Physical Review dataset: 300 or more publications per year (black); between 100 and 300 publications per year (orange); and between 20 and 100 publications year (cyan). The dashed red line corresponds to the best-fit line which we estimated by pooling these three subsets together, demonstrating how the independent weighting scheme better reproduces the cross-country variation than the n_p-dependent scheme.

table S1. Full panel regression model results. Parameter estimates for the panel data model for the collaboration rate $f_{i,t}^{s}$ (see Eq. S4), implemented with country fixed-effects and robust standard error estimates. Red and blue highlights indicate parameters significant at the $p \leq 0.05$ level. Beta coefficient are estimated using standardized variables for the non-categorical variables ($\log_{10} D_{i,t}^{s}$ thru $\widetilde{G}_{i,\tau}^{out}$).

	Full model parameter estimates			Partial model parameter estimates (A-E)										4-period model (F)		3-period model (G)	
Dependent Variable: $f_{i,t}^{s}$ (fraction)	Eq. [S4] Coeff.	Stand. var. (beta)	*p*-value	Model A	*p*-value	Model B	*p*-value	Model C	*p*-value	Model D	*p*-value	Model E	*p*-value	Model F	*p*-value	Model G	*p*-value
Year, *t*	0.015 ± 0.001	0.095 ± 0.008	0.000	0.019 ± 0.001	0.000	0.009 ± 0.001	0.000	0.011 ± 0.002	0.000	0.015 ± 0.001	0.000	0.018 ± 0.002	0.000	0.016 ± 0.005	0.006	0.048 ± 0.009	0.000
T_{EU} (EU entry – treatment effect)	-0.058 ± 0.019	-0.376 ± 0.122	0.004	-0.043 ± 0.018	0.027	-0.070 ± 0.020	0.001	-0.076 ± 0.021	0.001	-0.055 ± 0.021	0.012	-0.053 ± 0.019	0.011	-0.044 ± 0.013	0.003	-0.130 ± 0.013	0.000
Scientific productivity and impact																	
$\log_{10} D_{i,t}^{s}$ (publications)	-0.223 ± 0.037	-1.253 ± 0.208	0.000	-0.216 ± 0.041	0.000					-0.224 ± 0.039	0.000	-0.214 ± 0.037	0.000	-0.292 ± 0.041	0.000	-0.294 ± 0.062	0.001
$R_{i,t}^{s}$ (normalized citations)	0.164 ± 0.023	0.935 ± 0.132	0.000	0.159 ± 0.024	0.000					0.164 ± 0.024	0.000	0.159 ± 0.023	0.000	0.196 ± 0.035	0.000	0.197 ± 0.032	0.000
R&D investment																	
$\log_{10} E_{i,t}$ (Gvt. expenditure on R&D)	-0.080 ± 0.047	-0.467 ± 0.275	0.100			-0.096 ± 0.042	0.031			-0.065 ± 0.043	0.138			0.014 ± 0.058	0.811	0.205 ± 0.059	0.005
$\log_{10} GDPpc_{i,t}$ (per capita GDP)	0.217 ± 0.058	0.505 ± 0.135	0.001			0.129 ± 0.060	0.040			0.151 ± 0.062	0.021			0.456 ± 0.110	0.000	0.198 ± 0.112	0.104
$\log_{10} Spc_{i,t}$ (per capata researchers)	0.164 ± 0.063	0.292 ± 0.113	0.015			0.171 ± 0.062	0.010			0.182 ± 0.062	0.006			0.201 ± 0.060	0.002	-0.085 ± 0.092	0.376
High-skilled mobility																	
$B_{i,t}$ (high-skilled mobility polarization)	-0.043 ± 0.013	-0.169 ± 0.049	0.002					-0.056 ± 0.012	0.000			-0.046 ± 0.011	0.000	-0.095 ± 0.024	0.000	-0.115 ± 0.021	0.000
$\log_{10} I_{i,t}^{+}$ (total incoming mobility)	-0.024 ± 0.010	-0.187 ± 0.080	0.026					-0.028 ± 0.012	0.023			-0.023 ± 0.011	0.042	-0.056 ± 0.015	0.001	-0.099 ± 0.017	0.000
$\log_{10} O_{i,t}^{+}$ (total outgoing mobility)	0.019 ± 0.008	0.130 ± 0.059	0.034					0.030 ± 0.009	0.003			0.027 ± 0.009	0.004	0.011 ± 0.011	0.307	-0.007 ± 0.012	0.574
$\mathsf{P}_{i,t}^{in}$ (incoming success rate)	-0.015 ± 0.040	-0.036 ± 0.101	0.722					-0.002 ± 0.043	0.967			-0.019 ± 0.042	0.651	-0.004 ± 0.049	0.940	-0.110 ± 0.062	0.104
$\mathsf{P}_{i,t}^{out}$ (outgoing success rate)	-0.110 ± 0.045	-0.206 ± 0.084	0.020					-0.068 ± 0.047	0.159			-0.067 ± 0.050	0.189	-0.201 ± 0.056	0.001	-0.088 ± 0.055	0.134
$G_{i,t}^{in}$ (incoming mobility Gini index)	0.011 ± 0.035	0.026 ± 0.079	0.746					0.003 ± 0.040	0.937			0.017 ± 0.038	0.660	0.009 ± 0.043	0.828	0.145 ± 0.055	0.024
$G_{i,t}^{out}$ (outgoing mobility Gini index)	0.135 ± 0.044	0.237 ± 0.077	0.004					0.101 ± 0.048	0.044			0.103 ± 0.051	0.050	0.272 ± 0.056	0.000	0.166 ± 0.049	0.006
Total migration																	
$\widetilde{B}_{i,\tau}$ (migration polarization)	-0.040 ± 0.020	-0.169 ± 0.084	0.052					-0.027 ± 0.018	0.145			-0.036 ± 0.020	0.087	-0.060 ± 0.033	0.078	-0.514 ± 0.081	0.000
$\log_{10} \widetilde{I}_{i,\tau}$ (total incoming mobility)	-0.044 ± 0.011	-0.186 ± 0.048	0.000					-0.050 ± 0.011	0.000			-0.048 ± 0.011	0.000	-0.154 ± 0.051	0.005	-0.458 ± 0.057	0.000
$\log_{10} \widetilde{O}_{i,\tau}$ (total outgoing mobility)	0.024 ± 0.013	0.172 ± 0.092	0.071					0.016 ± 0.011	0.165			0.019 ± 0.012	0.138	0.018 ± 0.014	0.199	0.125 ± 0.024	0.000
$\widetilde{G}_{i,\tau}^{in}$ (incoming migration Gini index)	0.212 ± 0.104	0.086 ± 0.042	0.051					0.242 ± 0.119	0.051			0.264 ± 0.102	0.015	0.204 ± 0.162	0.217	-1.60 ± 0.23	0.000
$\widetilde{G}_{i,\tau}^{out}$ (outgoing migration Gini index)	-0.030 ± 0.044	-0.023 ± 0.034	0.502					-0.025 ± 0.039	0.528			-0.017 ± 0.043	0.687	-0.137 ± 0.052	0.014	-1.18 ± 0.26	0.001
Subject Area (s) (publication-level)																	
"Agricultural and Biological, Sciences" (1100)	-0.031 ± 0.032	-0.202 ± 0.208	0.339	-0.028 ± 0.034	0.419	0.043 ± 0.009	0.000	0.062 ± 0.015	0.000	-0.032 ± 0.033	0.342	-0.027 ± 0.033	0.415	-0.067 ± 0.031	0.036	0.281 ± 0.060	0.001
"Biochemistry, Genetics, and Molecular Biology" (1300)	0.044 ± 0.017	0.284 ± 0.112	0.016	0.046 ± 0.018	0.016	0.077 ± 0.013	0.000	0.097 ± 0.008	0.000	0.044 ± 0.018	0.019	0.047 ± 0.017	0.012	0.026 ± 0.017	0.127	0.422 ± 0.077	0.000
"Business Management and Accounting" (1400)	-0.359 ± 0.056	-2.324 ± 0.361	0.000	-0.350 ± 0.060	0.000	-0.126 ± 0.010	0.000	-0.106 ± 0.017	0.000	-0.360 ± 0.058	0.000	-0.348 ± 0.056	0.000	-0.456 ± 0.062	0.000	-0.133 ± 0.026	0.000
"Chemical Engineering" (1500)	-0.177 ± 0.040	-1.145 ± 0.259	0.000	-0.171 ± 0.043	0.000	-0.020 ± 0.010	0.045	-0.001 ± 0.011	0.924	-0.178 ± 0.041	0.000	-0.169 ± 0.041	0.000	-0.236 ± 0.045	0.000	0.141 ± 0.061	0.041
"Chemistry" (1600)	-0.028 ± 0.025	-0.184 ± 0.164	0.269	-0.025 ± 0.027	0.355	0.037 ± 0.012	0.003	0.056 ± 0.013	0.000	-0.029 ± 0.026	0.272	-0.025 ± 0.026	0.349	-0.058 ± 0.023	0.017	0.314 ± 0.092	0.006
"Computer Science" (1700)	-0.135 ± 0.027	-0.874 ± 0.175	0.000	-0.132 ± 0.029	0.000	-0.075 ± 0.011	0.000	-0.056 ± 0.016	0.002	-0.135 ± 0.028	0.000	-0.131 ± 0.028	0.000	-0.156 ± 0.029	0.000	0.227 ± 0.058	0.002
"Decision Sciences" (1800)	-0.315 ± 0.062	-2.043 ± 0.404	0.000	-0.305 ± 0.068	0.000	-0.023 ± 0.014	0.125	-0.003 ± 0.017	0.848	-0.317 ± 0.065	0.000	-0.303 ± 0.063	0.000	-0.393 ± 0.066	0.000	(omitted)	
"Energy" (2100)	-0.248 ± 0.050	-1.604 ± 0.325	0.000	-0.240 ± 0.054	0.000	-0.034 ± 0.016	0.039	-0.015 ± 0.019	0.450	-0.249 ± 0.052	0.000	-0.238 ± 0.051	0.000	-0.321 ± 0.060	0.000	0.064 ± 0.056	0.276
"Engineering" (2200)	-0.068 ± 0.020	-0.439 ± 0.129	0.002	-0.066 ± 0.021	0.003	-0.060 ± 0.009	0.000	-0.040 ± 0.015	0.010	-0.068 ± 0.020	0.002	-0.066 ± 0.020	0.003	-0.060 ± 0.020	0.006	0.331 ± 0.078	0.001
"Environmental Science" (2300)	-0.118 ± 0.038	-0.762 ± 0.243	0.004	-0.113 ± 0.040	0.008	(omitted)		0.019 ± 0.014	0.181	-0.118 ± 0.039	0.005	-0.112 ± 0.038	0.006	-0.164 ± 0.038	0.000	0.182 ± 0.057	0.008
"Materials Science" (2500)	0.003 ± 0.022	0.017 ± 0.145	0.906	0.006 ± 0.024	0.818	0.056 ± 0.011	0.000	0.075 ± 0.014	0.000	0.002 ± 0.023	0.922	0.006 ± 0.023	0.791	-0.004 ± 0.023	0.871	0.394 ± 0.088	0.001
"Medicine" (2700)	(omitted)	baseline Subj. Area		(omitted)		-0.035 ± 0.021	0.097	-0.016 ± 0.011	0.158	(omitted)		(omitted)		(omitted)		0.386 ± 0.096	0.002
"Pharmacology, Toxicology, and Pharmaceutics" (3000)	-0.191 ± 0.038	-1.240 ± 0.246	0.000	-0.185 ± 0.041	0.000	-0.019 ± 0.014	0.180	(omitted)		-0.192 ± 0.040	0.000	-0.183 ± 0.038	0.000	-0.254 ± 0.040	0.000	0.116 ± 0.058	0.069
"Physics and Astronomy" (3100)	0.151 ± 0.019	0.976 ± 0.126	0.000	0.152 ± 0.020	0.000	0.164 ± 0.013	0.000	0.183 ± 0.017	0.000	0.150 ± 0.020	0.000	0.153 ± 0.020	0.000	0.147 ± 0.019	0.000	0.544 ± 0.087	0.000
Constant	-29.1 ± 2.5	-190 ± 17	0.000	-36.2 ± 2.6	0.000	-16.9 ± 2.6	0.000	-20.8 ± 3.3	0.000	-29.1 ± 2.6	0.000	-34.5 ± 3.5	0.000	-32.2 ± 10.5	0.004	-94.4 ± -18.0	0.000
Adjusted R^2	0.66	0.66		0.65		0.61		0.61		0.65		0.66		0.67		0.60	
Number of observations	4494	4494		4494		4494		4494		4494		4494		1680		504	
Number of countries	31	31		31		31		31		31		31		31		12	