The Chosen: Researcher Visas and Foreign PhD graduates' careers in France

Andriy Romanyuk[1,2], Alberto Corsini[3], Ina Ganguli[4,5] and Francesco Lissoni[1,6]

[1] Bordeaux School of Economics, Université de Bordeaux
[2]Dipartimento di Economia, Università dell'Insubria (Varese)
[3]Institute of Public Goods and Policies, CSIC (Madrid)
[4]Department of Economics, University of Massacchusetts Amherst
[5]National Bureau of Economic Research (NBER)
[6]Dipartimento di Politica Economica, Università Cattolica (Milano)

EEA 2025, Bordeaux - August 26

Motivation: international students and innovation

International students are:

- increasing worldwide (decades-long trend)
- **over-represented in STEM**, esp. in graduate and doctoral studies (Hunt and Gauthier-Loiselle, 2010; UNESCO, 2016)
- a (potential) resource for innovation in host countries:
  - US: extensive evidence (Chellaraj et al., 2008; Hunt, 2011; Ganguli et al., 2020)
  - Europe: sparse ad weaker (Zheng and Ejermo, 2015; Buenstorf et al., 2022)

Key issue: retention rate and determinants of stay decisions

- Relative conditions of home/host countries (Grogger and Hanson, 2015)
- Host countries' migration policies $\rightarrow$ **transition from student to work visa** & family visas (Czaika, 2018; Kahn and MacGarvie, 2020)
  - US: Evidence on H-1B visa (Kerr and Lincoln, 2010; Doran et al., 2022; Roach and Skrentny, 2021)
  - Europe: No evidence so far, but policy relevance: EU Scientific Visa and Blue Card (2005, 2009), plus national policies ...

Policy focus & research questions

Focus: Carte de sejour "scientifique-chercheur"

- France's main scientific visa (plus residence permit)
- First introduced in 1998 and later reformed

Research questions

- Has the new *Carte de sejour "scientifique-chercheur"* improved the stay rate of foreign STEM doctoral graduates?
  - → Already in 1998?
  - → Through its successive reforms?
- (Not yet in the paper) Were stayers positively or negatively selected?
  - → Do we find them in the lower or upper tails of scientific productivity distribution?
  - → Do the best foreign graduates move to other advanced countries (US, UK, elsewhere in the EU)?

Background / I

International PhD students in France

- Top destination worldwide (4th in 2021)
- Key role: 37% of total enrollments (>50% for ICT and Engineering)
- Highly specific catchment area (former colonies), although less and less so (rise of East Asians)

(Sources: OECD, 2000, 2022; Campus France, 2024)

Dual Immigration Policy

1. Freedom of Movement for EU (EEA) citizens
2. Severe entry restrictions for non-EU ones:
   - 1974: Labour immigration suspended, then OSE ("opposabilité de la situation de l'emploi": labour market test) and minimum wage requirements
   - 1977: Limitation of foreign students' stay rights
   - mid-1980s: "zero immigration" policy goal

Background / II

Loi Reseda (1998): fewer restrictions for the highly skilled → *Carte de sejour "scientifique-chercheur"*

- Beneficiaries: master and doctoral graduates, if employed at universities or public research labs
- Exemption from labor market tests (OSE) and from minimum wage requirements
- 2006 & 2016: further simplifications → part of today's "passport for talents" system
- *Exceptions* due to bilateral treaties with former colonies:
  - *Algeria*: No access to the *carte* until 2001
  - *Togo, Gabon and Centrafrican Republic*: Default exemtption from any labour entry test until 2000

No data for 1990s-2000s → 2018-2022: avg of **4K applications** per year, around **24% of foreign researchers** working in France

(Sources: Slama, 2001; GISTI, 2020; OECD, 2017; d'Albis and Boubtane, 2021)

Identification Strategy / I

Quasi-experimental setting: *Carte de sejour "scientifique-chercheur"* as exogenous treatment of specific *doctoral graduates' groups and cohorts*:

Groups *Non-EU* vs *EU* graduates (incl. French)

Cohorts Graduates who **enrolled before** the administration of treatment ($<$1998) **but graduated after** it ($\geq$1998)

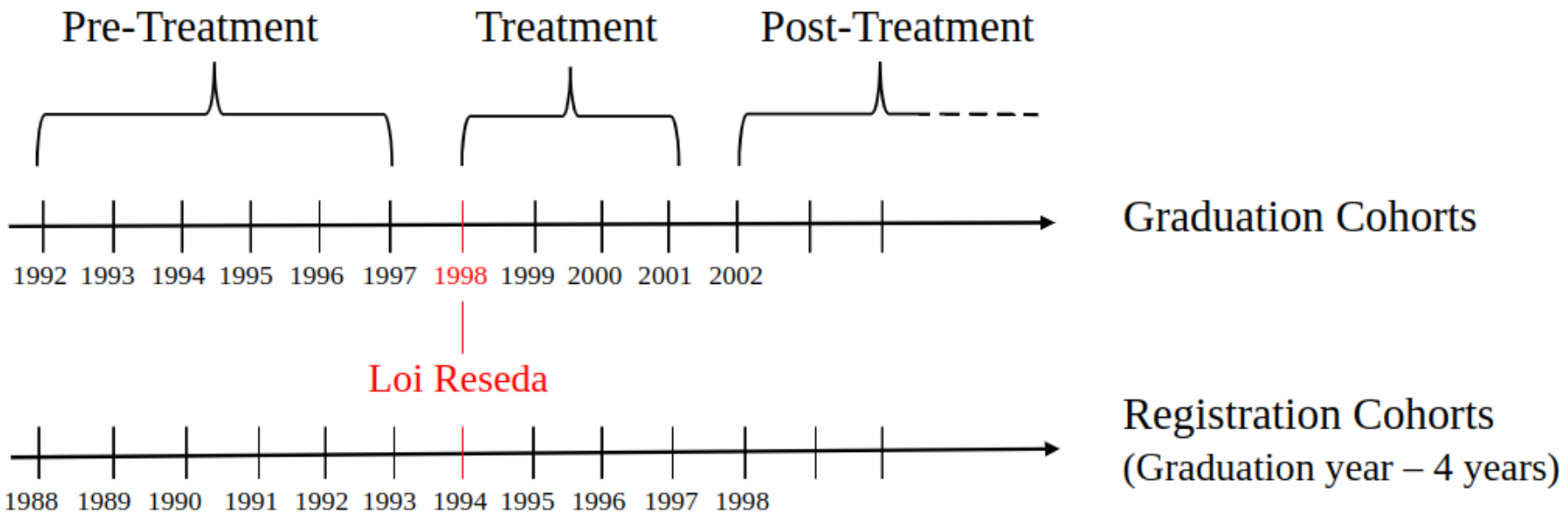
vs

Graduates who graduated before it ($<$1998)

NB Graduates who enrolled after 1998: endogenous selection into treatment

Difference-in-differences with repeated cross-sections, as in Waldinger (2010)

Identification Strategy / II



Doctoral graduates' cohorts

Identification Strategy / III

$$\Pr(Y_{i,c+t} = 1 \mid X_{i,c}) = \alpha + \gamma_i \cdot Non\_EU_i + \sum_{\substack{c \in Cohorts \\ c \neq 1997}} \theta_c \cdot Cohort_c +$$

$$+ \sum_{\substack{c \in Cohorts \\ c \neq 1997}} \delta_{i,c} \cdot (Non\_EU_i \times Cohort_c) + \phi_i \cdot \mathbf{X}_i + FE_j + \epsilon_{i,c}$$

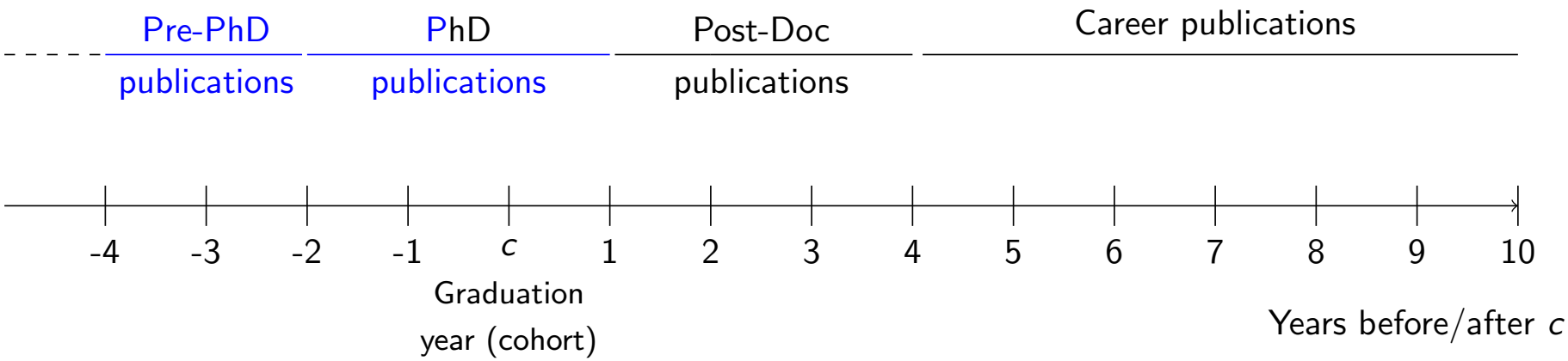
where $\Pr(Y_{i,c+t} = 1 \mid X_{i,c})$: probability of graduate i in cohort c to be active in France t years after graduating $\rightarrow$ 3 proxies:

- To publish at least one scientific article with a French affiliation
  - $(c+2, c+4)$ years after graduation $\rightarrow$ *Post-Doc Stay*
  - $(c+5, c+10)$ years after graduation $\rightarrow$ *Career Stay*
- To appear as supervisor of a French doctoral dissertation, by 2025 $\rightarrow$ *Supervisor*

Identification Strategy / IV

Control variables (**X**):

- *Male*
- *Double_degree* ; *Foreign_affiliation_during_PhD*
- *Published_before_PhD* ; *Nr_publications_during_PhD*
- Fixed effects (FE) for universities, disciplines (and supervisors)



Doctoral graduates' publications: explanatory and outcome variables

Data / I

Theses.fr

EDT repository maintained by the French Bibliographic Agency for Higher Education (ABES)

- Catalogue of all doctoral dissertations in French universities since 1985 (printed or electronic), plus archive for the electronic ones
- Name, surname, and unique identifier for both doctoral graduates and their supervisors, along with details such as the dissertation title, abstract, field of study (Dewey decimal), and graduation year c (but no info on registration years $\rightarrow c - 4$)

Data / II

Scopus

Key info for our purposes:

- Authors' unique identifiers $\rightarrow$ graduate i's match to all of an author's publication, with no need of disambiguation
- Authors' affiliation $\rightarrow$ French vs foreign, in $(c+2, c+4)$ and $(c+5, c+10)$ years after graduation

Data extraction and augmentation:

- From **Theses.fr**: all the dissertations in STEM (1992-2002) $\rightarrow$ **80,903** obs
- **Gender** assignment based on first names (Corsini et al., 2022)
- Country or more broadly defined region of origin based on name analysis (see next)

Region of Origin Assignment / I

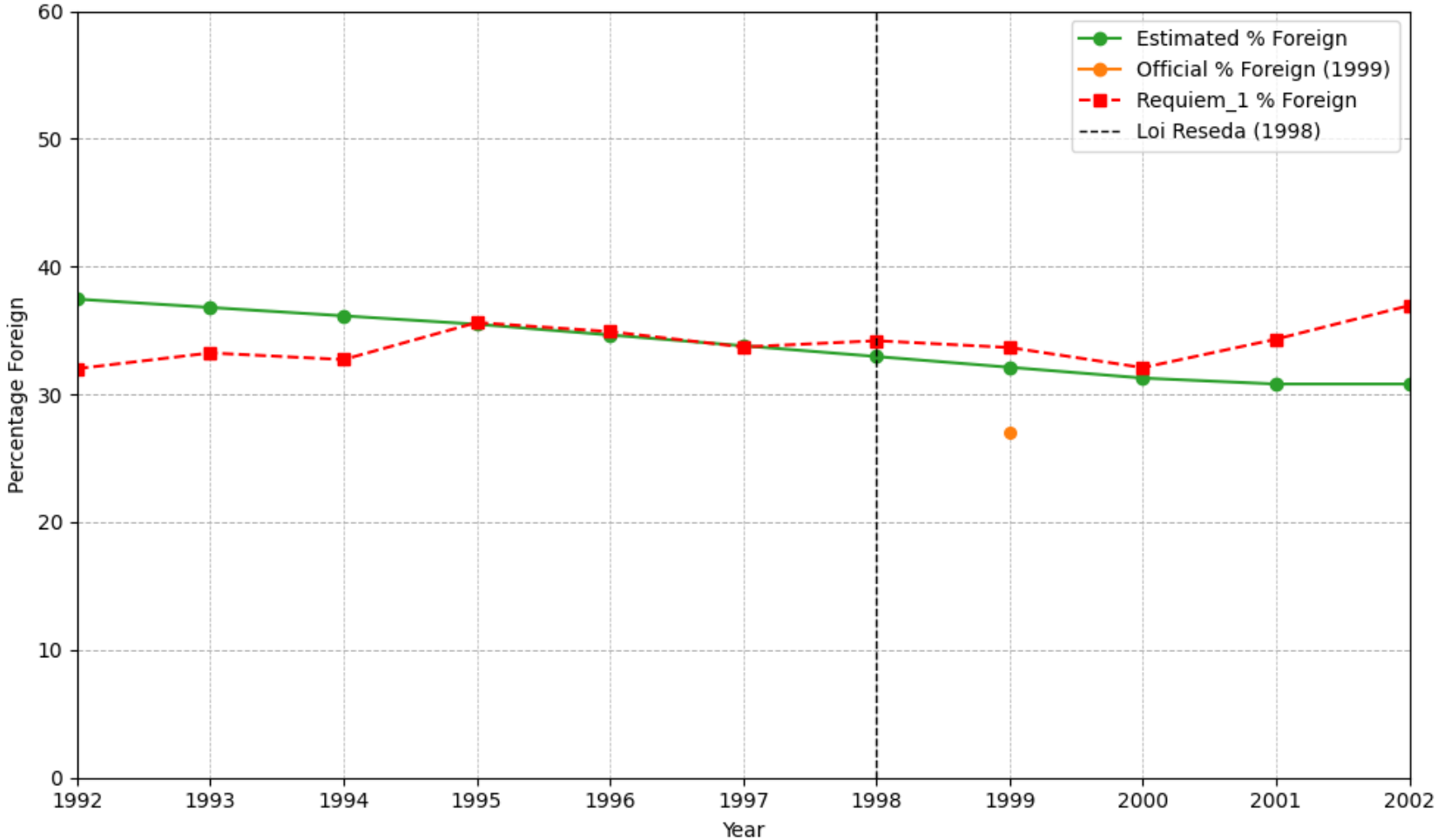
Problem: In *theses.fr*, no reliable info on doctoral graduates' country of birth and/or nationality

- Limited coverage (1990-99: missing for 30–40% obs)
- Inaccurate ("French" by default:only 5–7% identified as foreign vs 27% in official stats for 1999; Cohen, 2001)

Remedy: Name Analysis

- Based on Niggli' (2023) **LSTM machine learning model**
- Trained on the **Fichier des personnes décédées:** name, surname, country & date of birth, of all persons deceased in France since 1900
- Two sequential algorithms (*Requiem_1* and *Requiem_2*):
  1. French vs non-French graduates
  2. For non-French: 12 ethnolinguistic regions $\leftrightarrow$ foreign graduates' top countries of origin in 2011-12, plus Turkey and Iran(Campus France, 2017)

Region of Origin Assignment / II



Percentage of foreign doctoral graduates in France from 1992 to 2002:
Requiem_1 vs interpolation of official statistics

Region of Origin Assignment / III

Table 1: Taxonomy of Regions of origin and countries of origin of foreign PhD (2011-2012)

Regions of origin	Countries
Anglo_Saxon	United States
Arabic	Tunisia, Algeria, Lebanon, Moroc
Balto_Slavic	Russia
East_Asian	China, Vietnam
Italian	Italy
Niger_Congolese	Senegal, Ivory Coast, Cameroon, Gabon, Madagascar
Portuguese	Brazil, Portugal
Romanian	Romania
Spanish	Spain
West_North_Germanic	Germany, Belgium
Indo_Iranian	Iran
Turkic	Turkey

Region of Origin Assignment / IV

Regions of Origin	Count
French	55,007
Arabic	10,531
Niger_Congolese	2,466
West_North_Germanic	2,441
East_Asian	1,836
Spanish	1,564
Italian	1,384
Balto_Slavic	1,125
Indo_Iranian	967
Anglo_Saxon	861
Portuguese	861
Romanian	843
Turkic	385
Total	**80,271**

Requiem_1&2: classifications results

Region of Origin Assignment / V

Regions of Origin	Count
French	55,007
Arabic	10,531
Niger_Congolese	2,466
West_North_Germanic	2,441
East_Asian	1,836
Spanish	1,564
Italian	1,384
Balto_Slavic	1,125
Indo_Iranian	967
Anglo_Saxon	861
Portuguese	861
Romanian	843
Turkic	385
Total	**80,271**

Control group "French"

Control group "EU"

Treatment group non-EU

Requiem_1&2: classifications results AND treatment/control groups

Scopus Matching

Two steps:

1. **Name-based fuzzy matching** (`pybliometrics`; Rose and Kitchin, 2019)
   - 58,306 matches (72.1% of graduates)
2. **Filtering**: Random Forest algorithm trained on manually curated dataset
   - Predictors: supervisor co-authorship; institutional or national affiliation; similarity in title, discipline and author names
   - 52,893 valid matches (8% 1-to-m matches $\rightarrow$ 1-to-1 highest-score) $\rightarrow$ **"Publishing" graduates** (whether pre-, during and/or post-PhD)
   - 27,378 non-publishing graduates (no or non-valid match)
   - 632 dropped obs (>65 matches, mostly with East-Asian names)

Scopus' key feature: Author ID $\rightarrow$ validated match from one publication can be extended to all publications by the same author

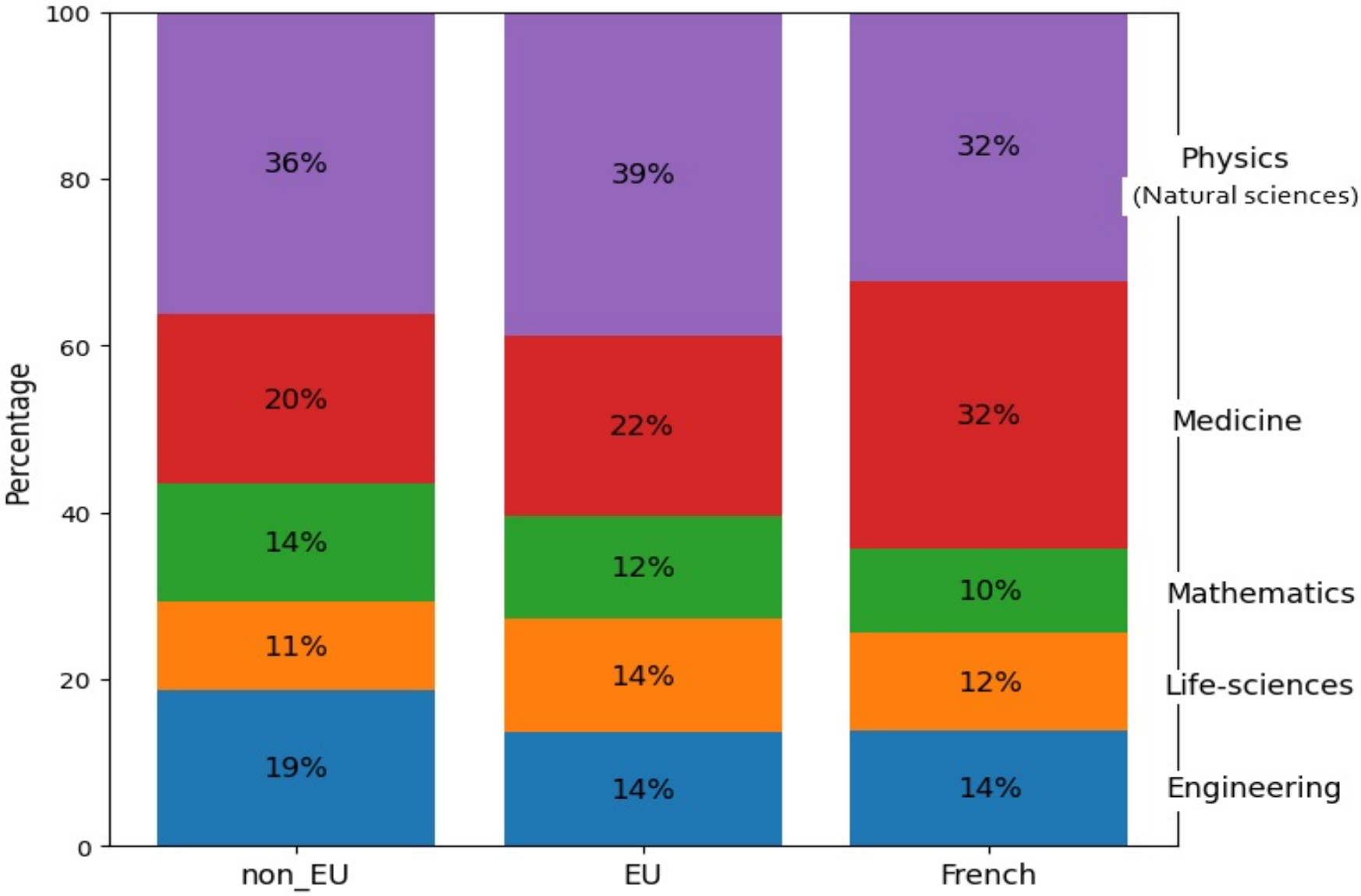
Descriptive statistics / I

Table 4: Summary Statistics, for selected areas of origin

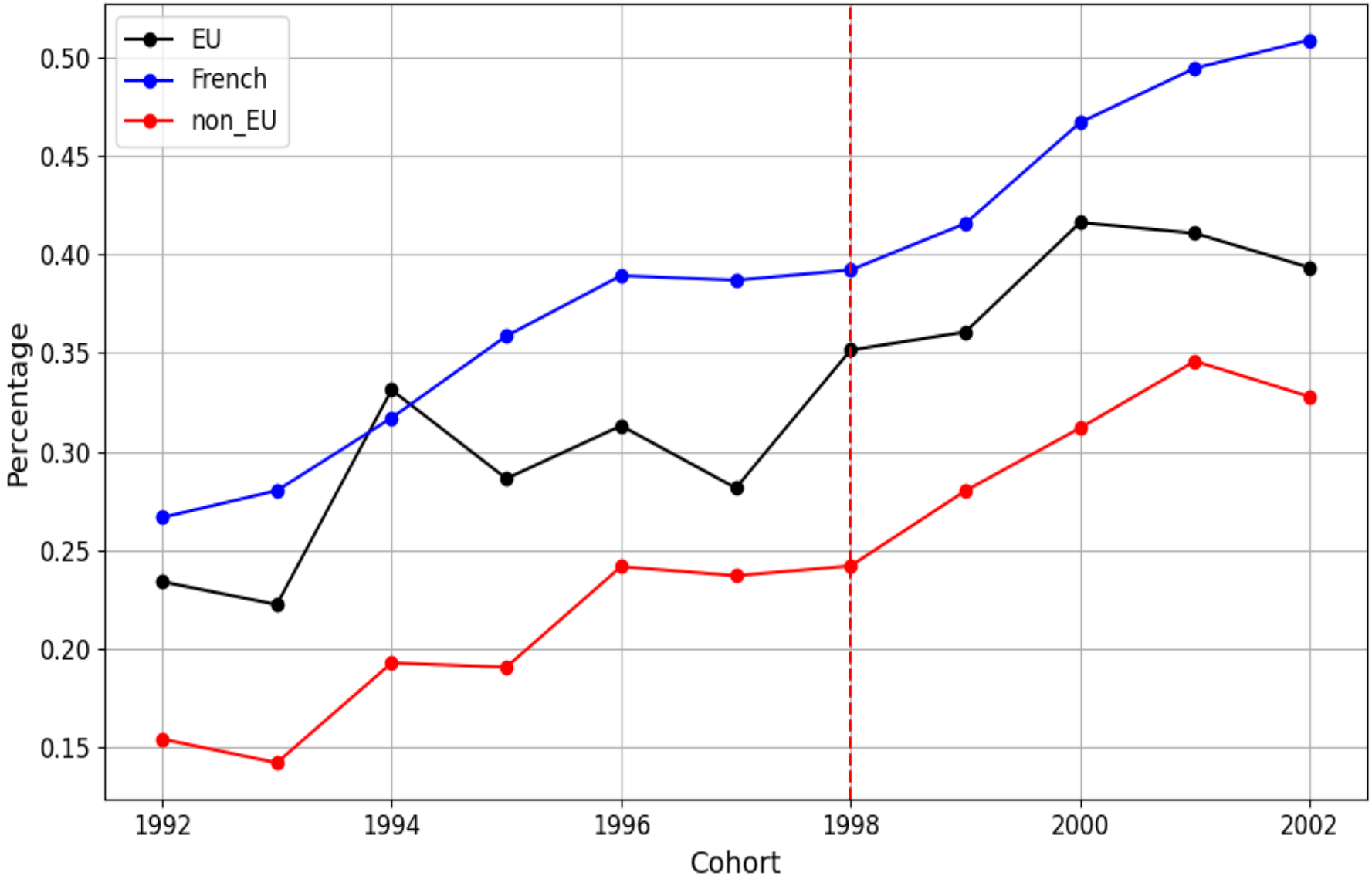
Variable	non_EU (1)	EU (2)	French (3)	(1 vs 2)	(1 vs 3)
Post Doc Stay	0.24	0.34	0.38	***	***
Career Stay	0.20	0.30	0.36	***	***
Supervisor	0.12	0.16	0.19	***	***
Female	0.38	0.44	0.39	***	
Double Degree	0.02	0.01	0.00	*	***
Published before PhD	0.10	0.16	0.12	***	***
Nr of Publications during PhD	1.66	2.66	2.19	***	***
Foreign Affiliation dur. PhD	0.12	0.19	0.08	***	***

Columns (1) to (3) report mean values for the indicated area of origin.
Columns (1 vs 2) and (1 vs 3) report significance levels for mean differences
(***: *0.01*, **: *0.05*, *: *0.1*).

Descriptive statistics / II

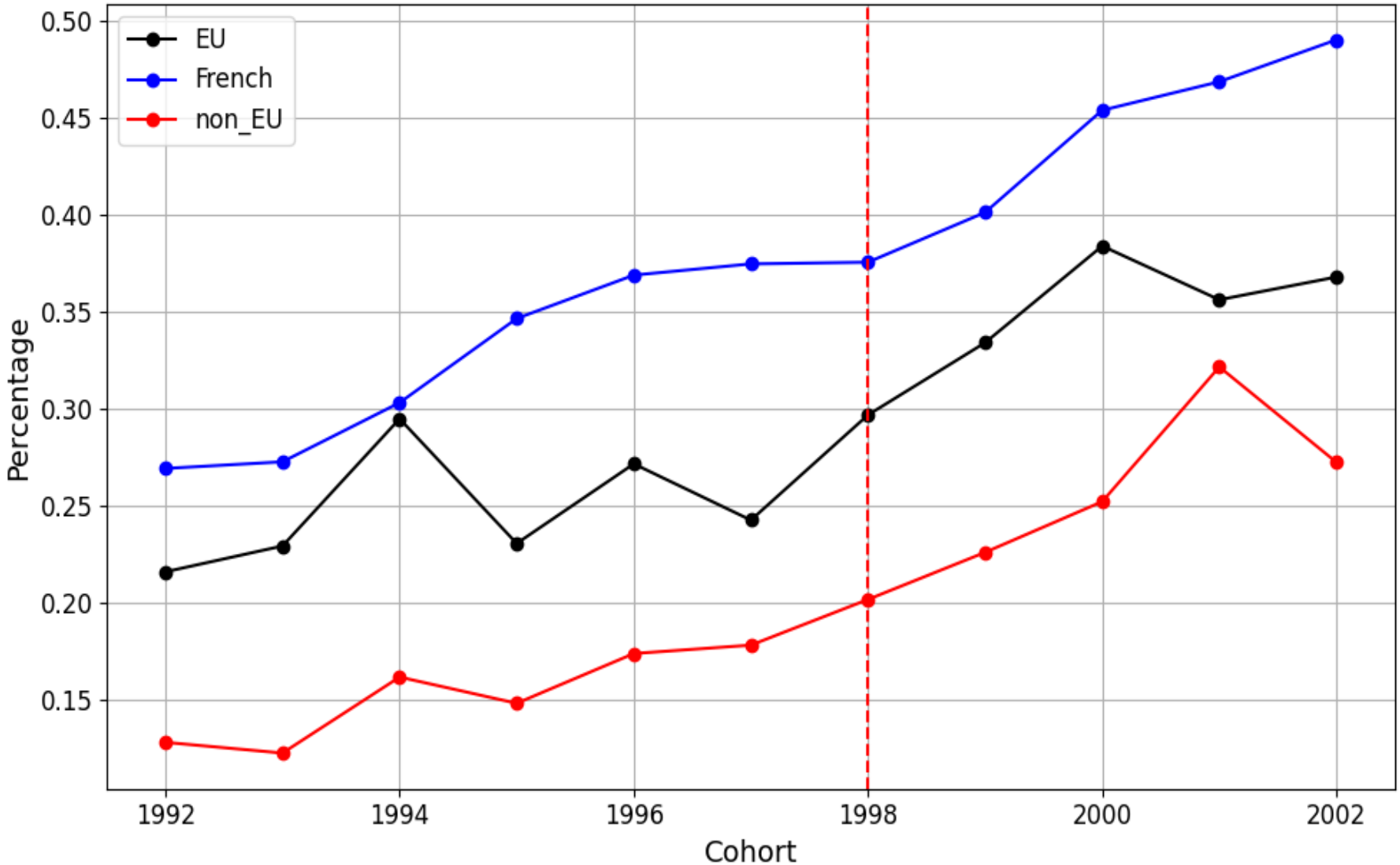


Descriptive statistics / III



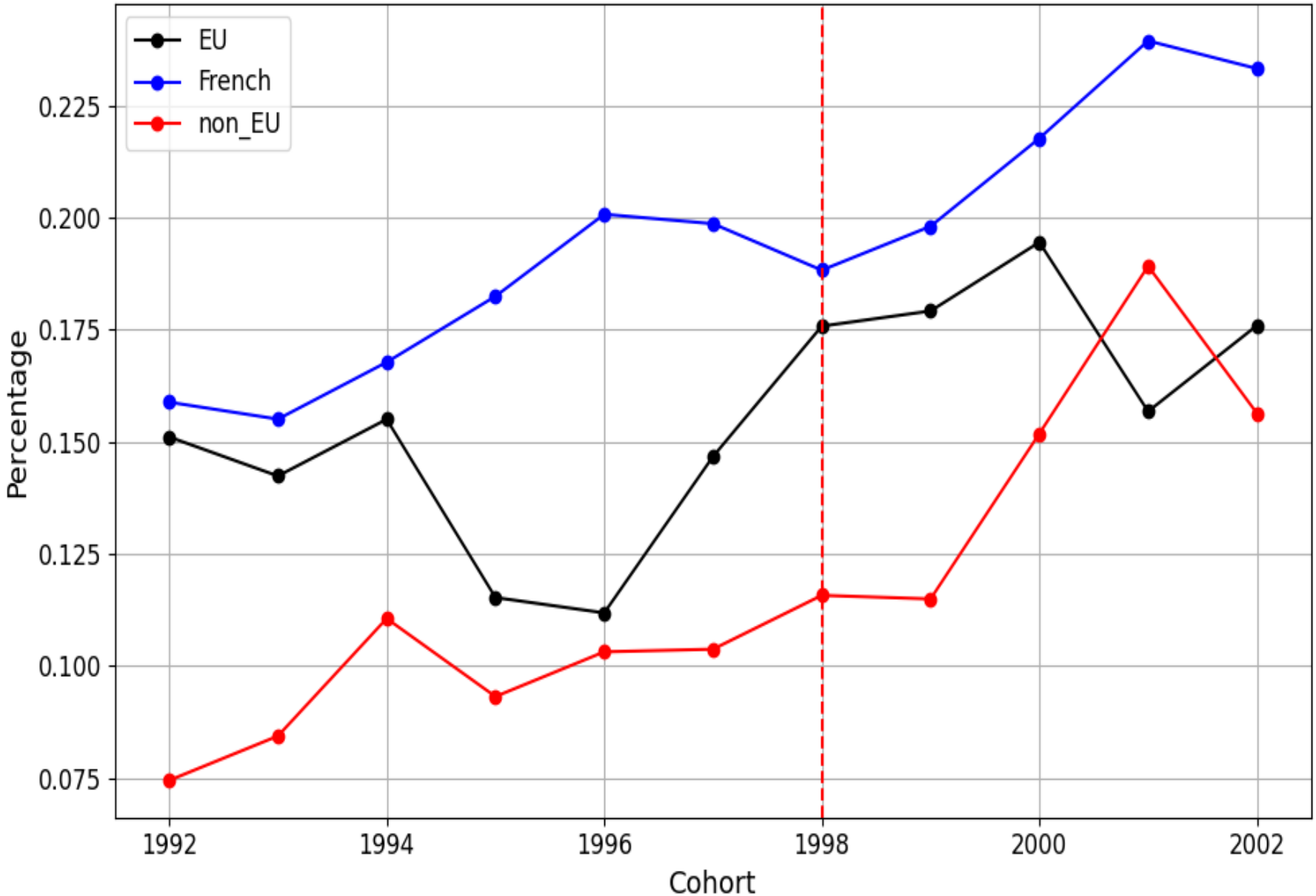
Post Doc Stay

Descriptive statistics / IV



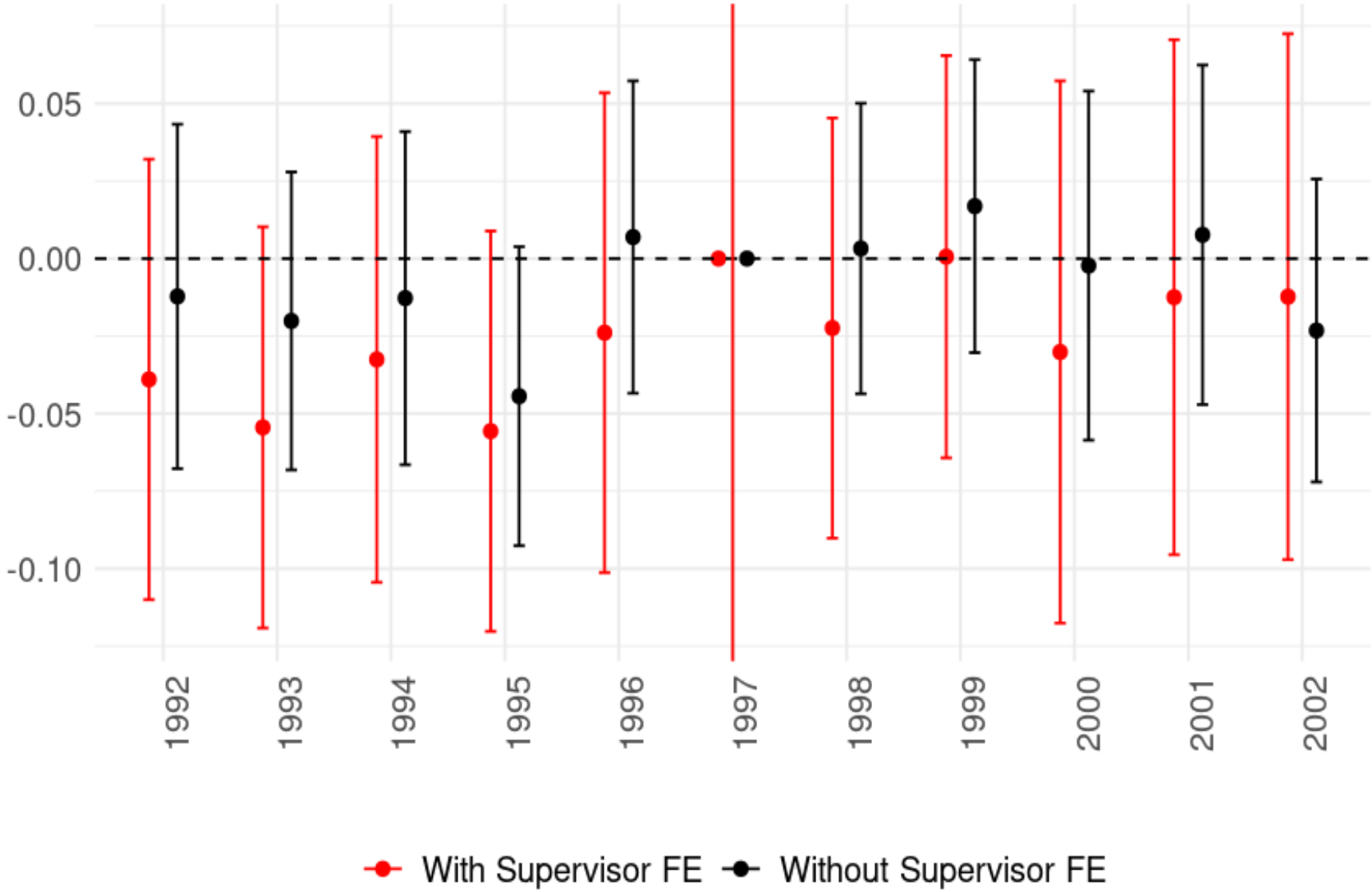
Career Stay

Descriptive statistics / V



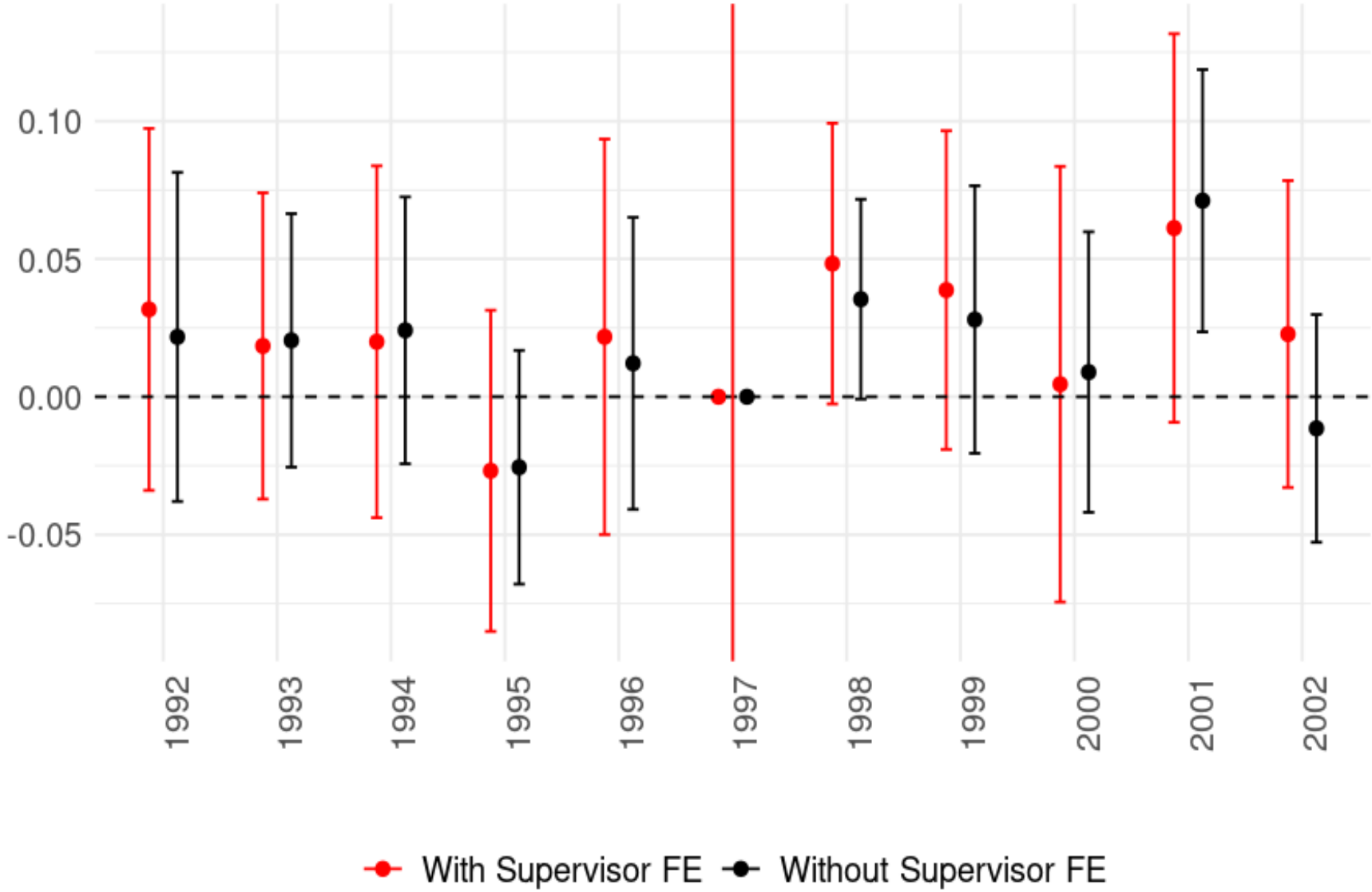
Supervisor

Results: *Post-Doc Stay* (non_EU vs French)



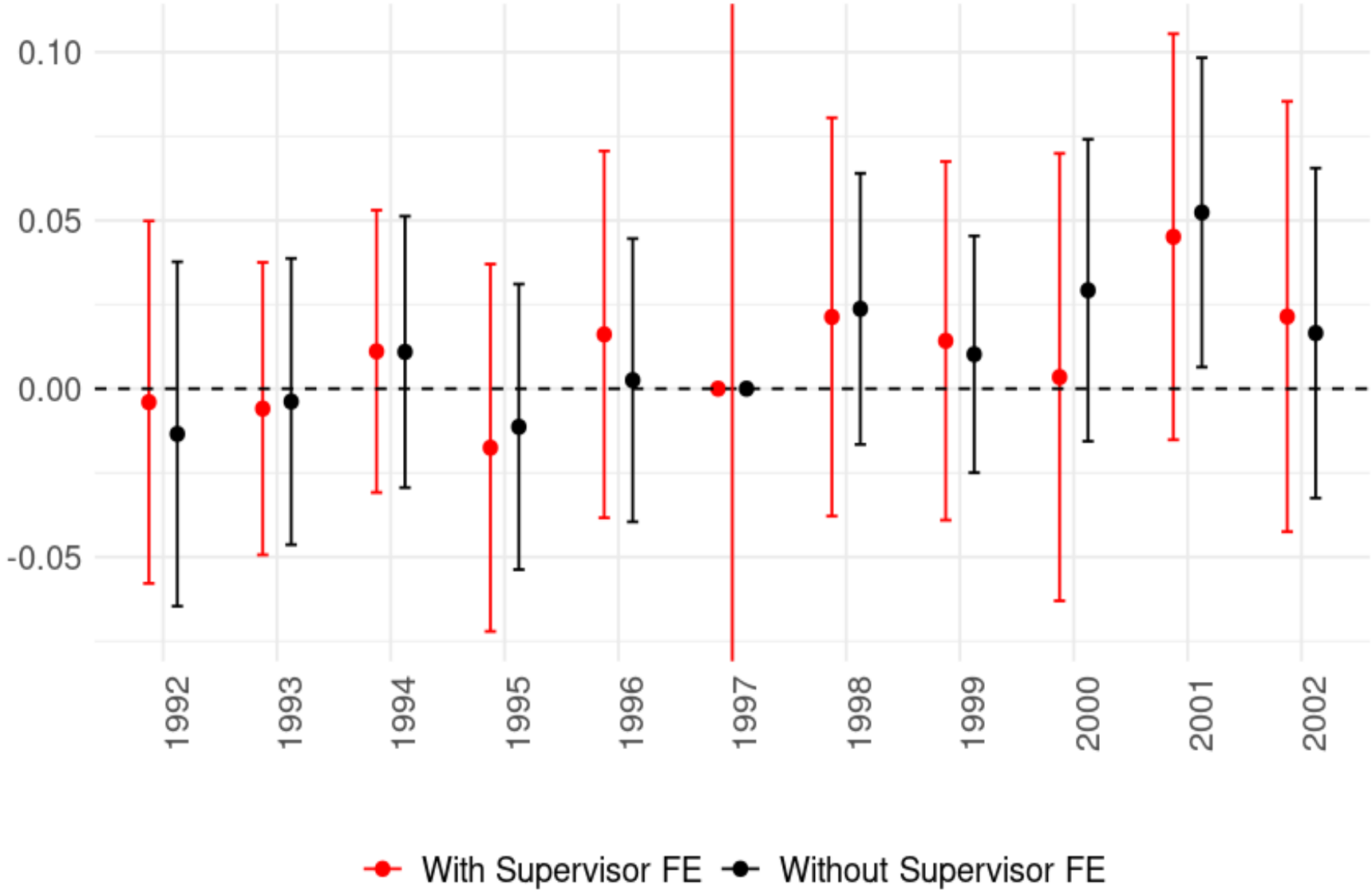
Estimated δ Coefficients (95% C.I.) by Cohort

Results: *Career_Stay* (non_EU vs French)



Estimated δ Coefficients (95% C.I.) by Cohort

Supervisor: baseline regressions (non_EU vs French)



Estimated δ Coefficients (95% C.I.) by Cohort

Conclusions & Next Steps / I

Results so far

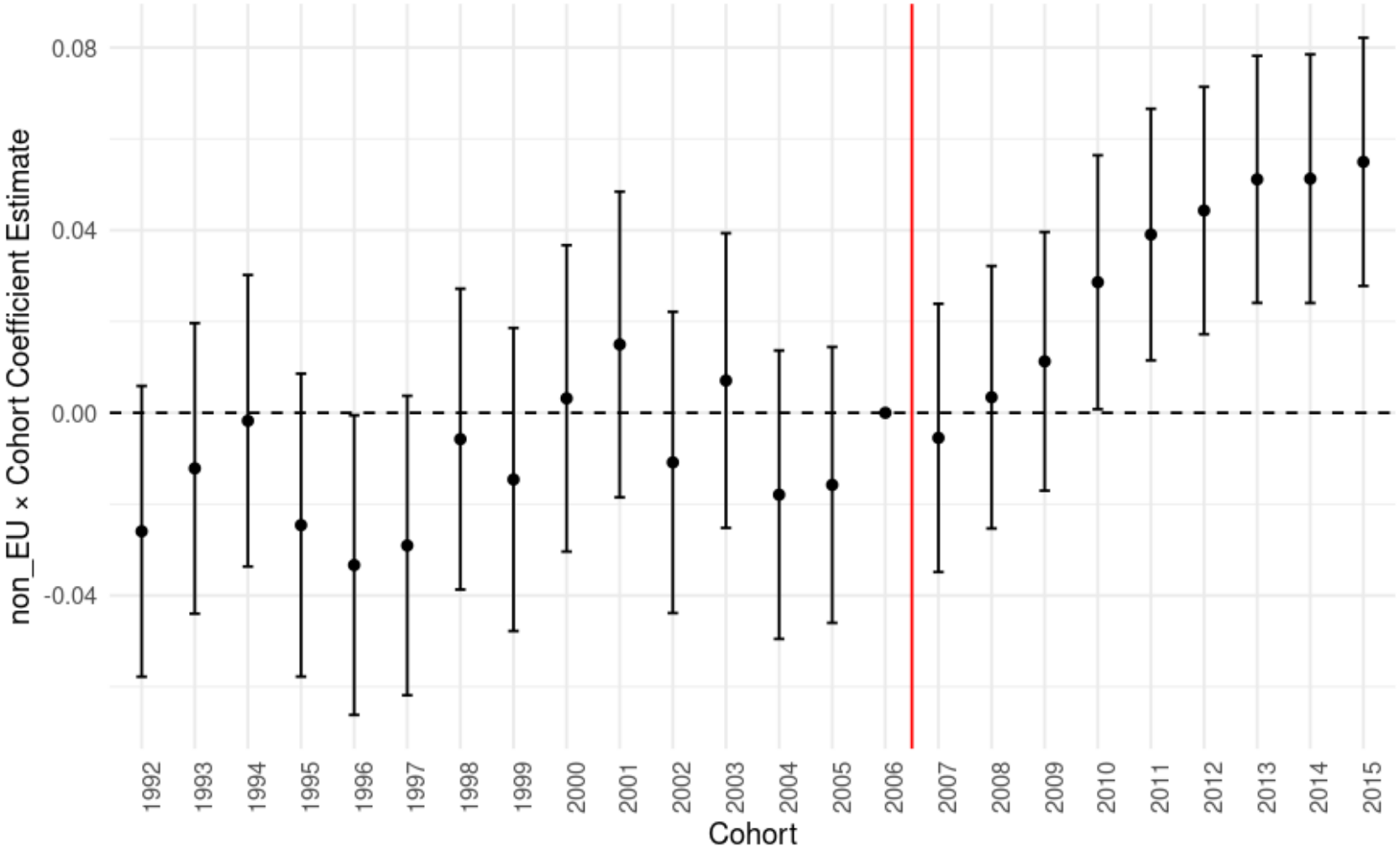
- No or negligible effects of the 1998 reform
- Potential issues with name analysis → attenuation bias, but:
  - ... Differences in stay rates as expected, and significant
  - ... Regression results robust to different specifications of both *Requiem_1* and *Requiem_2* (probability thresholds)
- Possible explanations:
  - High share of students from Togo, Gabon and Centrafrica
  - Inefficiencies in implementation
  - Much higher stay costs other than visa-related ones
  - Self-selection into non-academic careers (to be checked: stay rates conditional on publishing)

Next Steps

- Further explore the characteristics of stayers
- Replication for impact of successive reforms

2006 EU Directive 2005/71 → *carte sejour "compétence et talents"*
2016 Hollande reform → *passeport talent, mention "chercheur"*

Conclusions & Next Steps / II



Supervisor regression, 1992-2015 (ref year 2006)

Estimated δ Coefficients (95% C.I.) by Cohort (no supervisor FE)

References / I

Buenstorf, G., D. P. Heinisch, and M. Kapa (2022). International doctoral graduates as inventors in the german innovation system. *Industry and Innovation 29*(4), 488–510.

Campus France (2017, February). Les chiffres clés: 2017. Accessed: 2024-10-29.

Campus France (2019). Les doctorants à l'international - tendances de la mobilité doctorale en france et dans le monde. Agence française pour la promotion de l'enseignement supérieur, l'accueil et la mobilité internationale, `https://ressources.campusfrance.org/publications/notes/fr/note_60_fr.pdf` (last access: July 2024).

Campus France (2024). La mobilité étudiante dans le monde: Chiffres clés. Agence française pour la promotion de l'enseignement supérieur, l'accueil et la mobilité internationale, `https://ressources.campusfrance.org/noindex/chiffres_cles_2024_fr.pdf` (last access: July 2024).

Chellaraj, G., K. E. Maskus, and A. Mattoo (2008). The contribution of international graduate students to us innovation. *Review of International Economics 16*(3), 444–462.

Cohen, E. (2001). Un plan d'action pour améliorer l'accueil des étudiants étrangers en france: Diagnostic et propositions. *Rapport au ministre de l'Éducation nationale et au ministre des Affaires étrangères 124*, 24.

Corsini, A., M. Pezzoni, and F. Visentin (2022). What makes a productive ph. d. student? *Research Policy 51*(10), 104561.

Czaika, M. (2018). *High-skilled migration: Drivers and policies*. Oxford University Press.

Doran, K., A. Gelber, and A. Isen (2022). The effects of high-skilled immigration policy on firms: Evidence from visa lotteries. *Journal of Political Economy 130*(10), 2501–2533.

d'Albis, H. and E. Boubtane (2021). L'immigration professionnelle en france depuis 2000. *Annales des Mines-Realités industrielles 2 (Mai)*, 40–43.

Ganguli, I., S. Kahn, and M. J. MacGarvie (Eds.) (2020). *The Role of Immigrants and Foreign Students in Science, Innovation, and Entrepreneurship*. National Bureau of Economic Research - University of Chicago Press.

GISTI (2020). Statut des algériennes et des algériens en france. Groupe d'information et de soutien des immigrés, `https://www.gisti.org/spip.php?article6450` (last access: July 2024).

References / II

Grogger, J. and G. H. Hanson (2015). Attracting talent: Location choices of foreign-born phds in the united states. *Journal of Labor Economics 33*(S1), S5–S38.

Hunt, J. (2011). Which immigrants are the most innovative and entrepreneurial? Distinctions by entry visa. *Journal of Labor Economics 29*(3), 417–457.

Hunt, J. and M. Gauthier-Loiselle (2010, April). How much does immigration boost innovation. *American Economic Journal: Macroeconomics 2*(2), 31–56.

Kahn, S. and M. MacGarvie (2020). The impact of permanent residency delays for stem phds: Who leaves and why. *Research Policy 49*(9), 103879.

Kerr, W. R. and W. F. Lincoln (2010). The supply side of innovation: H-1b visa reforms and us ethnic invention. *Journal of Labor Economics 28*(3), 473–508.

Niggli, M. (2023). 'moving on'—investigating inventors' ethnic origins using supervised learning. *Journal of Economic Geography 23*(4), 921–947.

OECD (2000). *Education at a Glance. OECD Indicators*. Organisation for Economic Co-operation and Development.

OECD (2017). *Le recrutement des travailleurs immigrés : France*. Organisation for Economic Co-operation and Development.

OECD (2022). *Education at a Glance. OECD Indicators*. Organisation for Economic Co-operation and Development.

Roach, M. and J. Skrentny (2021). Rethinking immigration policies for stem doctorates. *Science 371*(6527), 350–352.

Rose, M. E. and J. R. Kitchin (2019). pybliometrics: Scriptable bibliometrics using a python interface to scopus. *SoftwareX 10*, 100263.

Slama, S. (2001). Tapis rouge pour les élites. *Plein droit 47-48*(1), 36–39.

UNESCO (2016). *UNESCO Science Report: Towards 2030*. United Nations Educational, Scientific and Cultural Organization.

Waldinger, F. (2010). Quality matters: The expulsion of professors and the consequences for phd student outcomes in nazi germany. *Journal of political economy 118*(4), 787–831.

Zheng, Y. and O. Ejermo (2015). How do the foreign-born perform in inventive activity? evidence from sweden. *Journal of Population Economics 28*, 659–695.

BACK-UP SLIDES

Background /1: International students in France

Key role of international students

- High share of foreign students in OECD countries, although declining: 11% (1998) $\rightarrow$ 4% (2021)
- Foreign share of enrollments in France on the rise: 7% $\rightarrow$ 9%
- Especially attractive for doctoral studies: ranked 4th globally in 2021 (37% of total enrollments; >50% for ICT and Engineering)

Highly specific catchment area, although less and less so

- Countries of origin, 2021 (2011):
  - MENA: 29% (40%)
  - Sub-Saharan Africa: 16% (13%)
  - China: 9% (12%)
  - Italy: 8% (12%)
- Around 1998: MENA even higher, China much lower

(Sources: OECD, 2000, 2022; Campus France, 2024)

Background /2: France's Dual Immigration Policy

1. **Freedom of Movement for EU citzens**:
   - EU citizens (plus Iceland, Norway, Switzerland) can move freely in France since 1957 (right formalized in 1992)

2. **Entry restrictions for non-EU citizens**:
   - Since the mid-1970s: **strong restrictions to labour immigration**
     - 1974: Labour immigration suspended
     - **1975: Labour market test** ("opposabilité de la situation de l'emploi": OSE) introduced
     - 1977: Limitation of foreign students' stay rights
     - mid-1980s: "zero immigration" policy goal
   - 1998: **selectivity turn** (Loi Reseda) $\rightarrow$ **fewer restrictions for the highly skilled**, including the creation of a *Carte de sejour "scientifique-chercheur"*

Carte de sejour "scientifique-chercheur"

Key features

- Beneficiaries: master and doctoral graduates, if employed at universities or public research labs
- Exemption from labor market tests (OSE) and from minimum wage requirements
- 2006 & 2016: further simplifications → part of today's "passport for talents" system

Exceptions due to bilateral treaties with former colonies:

- *Algeria*: No access to the *carte* until 2001
- *Togo, Gabon and Centrafrican Republic*: Default exemtption from any labour entry test until 2000

Relevance

- 2018-2022: avg of **4K applications** per year, around **24% of foreign researchers** working in France

(Sources: Slama, 2001; GISTI, 2020; OECD, 2017)

Legislation memo

- Loi n° 98-349 du 11 mai 1998 relative à l'entrée et au séjour des étrangers en France et au droit d'asile, dite loi RESEDA ou loi Chevènement → *carte de sejour "scientifique-chercheur"*
- Code de l'entrée et du séjour des étrangers et du droit d'asile ou CESEDA, parfois surnommé "code des étrangers" (entré en vigueur le 1er mars 2005; partie réglementaire publiée le 15 novembre 2006)
- Loi n° 2006-911 du 24 juillet 2006 relative à l'immigration et à l'intégration → *carte de sejour "compétence et talents"*, valable 3 ans et renouvlable accordée à l'étranger "susceptible de participer, du fait de ses compétences et de ses talents, de façon significative et durable au développement économique ou au rayonnement, notamment intellectuel, scientifique, culturel, humanitaire ou sportif de la France", selon les critères fixés chaque année par la Commission nationale des compétences et des talents
- La loi du 7 mars 2016 relative au droit des étrangers en France → *passeport talent, mention "chercheur"*, titulaires d'au moins un diplôme de master et chargés d'une mission de recherche ou d'enseignement universitaire dans le cadre d'une convention de pas plus de 4 ans

LSTM model in essence

- It leverages the sequential structure of characters in names and surnames $\rightarrow$ e.g. in *Roman-yuk* **higher weight to suffix "-yuk"** (distinctively Ukrainian)

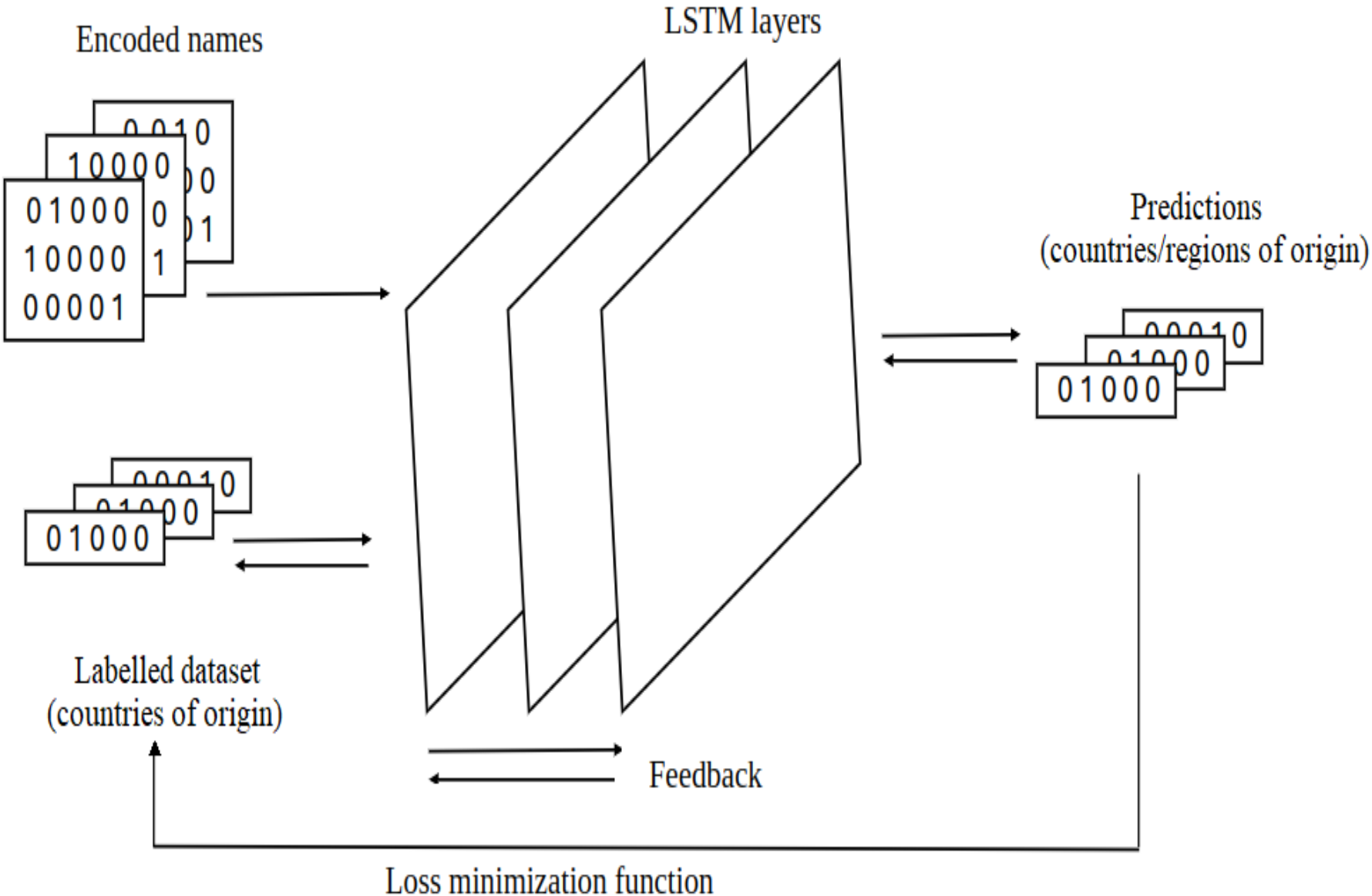
- 2 key inputs:
  1. **Labelled dataset** (60K obs from the *Fichier des personnes décédées*)
  2. **Pre-determined "classes"**: *linguistic & geographic* aggregation based on 2010/2011 Cohort - first available info (Campus France, 2019)

- F1 stat around 90% for each class (*external validation* ongoing)

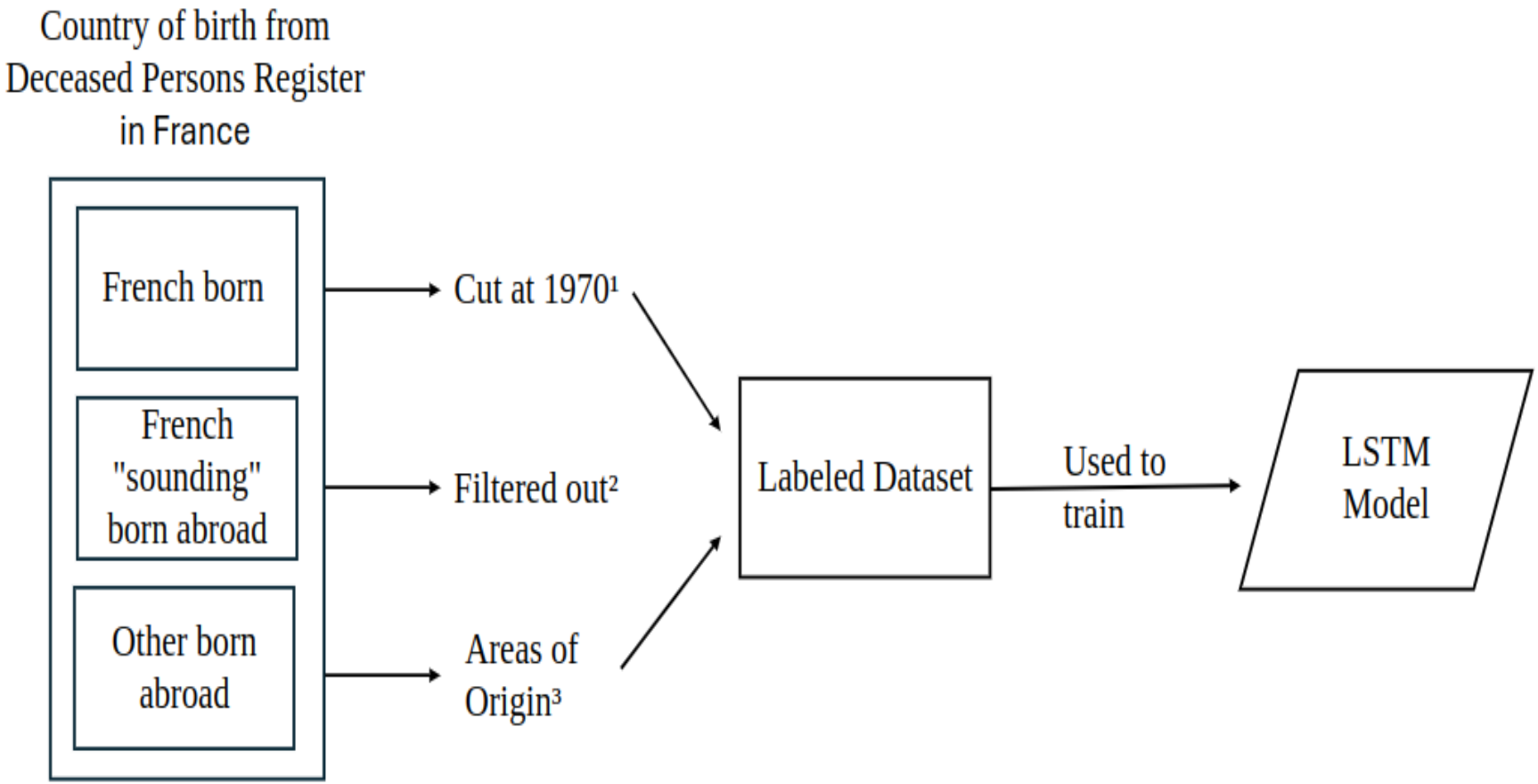
Encoding name example: Romain Gary

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z		END
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
7	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

LSTM Model Workflow



Requiem_1&2: Labelled dataset



1. To avoid confounding from subsequent migration waves
2. Using LSTM trained by Niggli (2023) on Olympics and Inventor data
3. From top foreign PhD sending countries

Requiem_1&2: Diagnostics / I

Requiem_1 in-sample performance:

Table B.3: Classification Report - **Requiem_1**

Class	Precision	Recall	F1-Score	Support
non_French	0.958	0.968	0.963	22733
French	0.968	0.957	0.962	22733
Accuracy			0.963	45466
Macro Avg	0.963	0.963	0.963	45466
Weighted Avg	0.963	0.963	0.963	45466

Requiem_1 out-of-sample performance (Linkedin):

Table B.8: Classification Report - **Requiem_1**: Using Observed Location Information, 1992-2001

Info Type	Precision		Recall		Accuracy	Support
	Non-French	French	Non-French	French		
Country of Birth	0.77	0.95	0.89	0.88	0.88	146
High School	0.91	0.98	0.88	0.90	0.90	164
Bachelor	0.59	0.97	0.93	0.78	0.82	282
Master	0.13	0.99	0.78	0.74	0.74	191

Requiem_1&2: Diagnostics / II

Requiem_2 in-sample performance:

Table B.4: Classification Report - **Requiem_2**

Class	Precision	Recall	F1-Score	Support
Anglo_Saxon	0.792	0.903	0.844	824
Arabic	0.969	0.907	0.937	8316
Balto_Slavic	0.943	0.903	0.923	1627
East_Asian	0.932	0.919	0.925	1045
Italian	0.777	0.960	0.859	1005
Niger_Congolese	0.867	0.819	0.842	2988
Portuguese	0.936	0.955	0.945	3072
Romanian	0.853	0.874	0.863	318
Spanish	0.857	0.859	0.858	1309
West_North_Germanic	0.765	0.843	0.802	1064
Indo_Iranian	0.596	0.824	0.692	369
Turkic	0.844	0.960	0.898	796
Accuracy			0.898	22733
Macro Avg	0.844	0.894	0.866	22733
Weighted Avg	0.905	0.898	0.900	22733

Requiem_1&2: Diagnostics / III