I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17906

'Based on Admin Data!': How Administrative Data Fosters Young Economists' Career

Anthony Lepinteur
Roberto Nisticò

MAY 2025

IZA Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17906

'Based on Admin Data!': How Administrative Data Fosters Young Economists' Career

Anthony Lepinteur
University of Luxembourg and IZA

Roberto Nisticò
University of Naples Federico II and IZA

MAY 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

'Based on Admin Data!': How Administrative Data Fosters Young Economists' Career*

This paper examines whether access to administrative data mitigates or reinforces inequalities in academic careers. We study the VisitINPS program, which grants researchers access to rich administrative records, and construct a longitudinal dataset covering the quasi-universe of applicants. Using a Two-Way Fixed Effects model complemented by a Regression Discontinuity Design, we find that administrative data access improves research visibility and career progression but does not increase overall publication volume. However, these gains are unequal and our findings suggest that administrative data access may magnify, rather than reduce, existing disparities in the academic economics community.

JEL Classification: J01, J60, J40

Keywords: administrative data, career progression, two-way fixed-effects

Corresponding author:
Roberto Nisticò
University of Naples Federico II
Department of Economics and Statistics
via Cintia Complesso Monte S. Angelo
80126, Napoli
Italy
E-mail: roberto.nistico@unina.it

* We would like to thank Maria De Paola, Giorgia Menta, Gabriele Letta and all participants of the 36th SIEP General Conference for useful discussions and comments. We also would like to thank Nisha Vernekar for excellent research assistance. We gratefully acknowledge support from INPS. Data access was provided as part of the "Visitinps Scholars" program. The findings and conclusions expressed are solely those of the authors and do not represent the views of INPS.

1. Introduction

Economic research has increasingly recognised the role of structural inequalities in shaping academic careers. A growing body of evidence shows that access to elite networks, institutional prestige, funding, and gender dynamics systematically influence hiring, promotion, and research recognition in economics (Waldinger, 2010; Jacob and Lefgren, 2011; Bosquet *et al.*, 2019; Tabakovic and Wollmann, 2019; Bransch and Kvasnicka, 2022; Ajzenman *et al.*, 2023; Ebehardt *et al.*, 2023; Hale *et al.*, 2023; Carrell *et al.*, 2024; Gualavisi *et al.*, 2024; Jones and Sloan, 2024). Scholars trained at top-ranked PhD programs benefit from stronger mentorship, wider professional connections, and greater visibility, which in turn increases their likelihood of publishing in high-impact journals and securing competitive faculty positions (Zinovyeva and Bagues, 2015; Heckman and Moktan, 2020).

At the same time, women and researchers from non-elite institutions face persistent disadvantages, including higher standards in research evaluation (Card *et al.*, 2020; Hengel, 2022), lower citation rates (Koffi, 2021), and exclusion from key professional opportunities such as high-profile conferences and influential co-authorships (Ductor *et al.*, 2021). These patterns not only shape individual careers but also have broader implications for the discipline, potentially leading to talent misallocation and limiting the diversity of perspectives in economic research (Bayer and Rouse, 2016; Lundberg and Stearns, 2019; Bayer *et al.*, 2020; Schultz and Stansbury, 2022).

The rise of administrative data as a cornerstone of empirical research introduces a new and largely unexplored dimension to this debate. As access to these datasets becomes increasingly central to applied economics (Card *et al.*, 2010; Einav and Levin, 2014; Connelly *et al.*, 2016; Figlio *et al.*, 2017; Currie *et al.*, 2020), an urgent question emerges: does working with administrative data help level the playing field, or does it further entrench existing inequalities?

By investigating this question, this paper connects the debate on structural inequalities in academia to the evolving nature of research production, by providing the first analysis of how access to administrative data shapes academic careers, and shedding light on whether this new methodological advance can serve as an equalising force or merely reinforce the status quo. In parallel, governments and public institutions increasingly act as custodians of administrative data, with access often mediated by publicly funded programs. As such, the design, reach, and equity of these programs are matters of public policy. As administrative data become essential inputs into high-impact research – and, by extension, into academic career success – the rules governing who can access these resources take on new significance. In this context, understanding whether public data access mechanisms mitigate or reproduce structural inequalities is crucial not only for academia but also for the fair and effective governance of research infrastructure.

We study the case of VisitINPS, a program launched by the Italian Social Security Institute (INPS) that grants researchers access to its administrative data. Using a unique longitudinal dataset covering the quasi-universe of early-career researchers who applied to the program, we examine whether gaining access to administrative data improves publication records, enhances professional visibility, and accelerates career advancement. To guide our empirical analysis, we develop a simple theoretical model in which researchers signal their quality through both past publications and future research potential, and we examine how administrative data access affects the strength of these signals in hiring and promotion decisions.

We test these predictions using a Two-Way Fixed Effects (TWFE) approach, complemented by a TWFE Regression Discontinuity Design (TWFE-RDD) that exploits variation in proposal scores around the VisitINPS selection threshold. This empirical design allows us to distinguish the causal impact of data access from potential selection effects. Our findings reveal that access to INPS administrative data significantly shapes academic trajectories, but not by simply increasing overall publication volume. Instead, researchers with data access are more likely to publish in highly ranked journals, particularly top field journals in labour economics – an area where INPS data are especially valuable. Thus, the primary career benefit comes not from more publications but from stronger signals of research potential – working papers, conference participation, and networking opportunities – that improve researchers' visibility and job market prospects. Additionally, administrative data access appears to shape academic careers geographically: while it reduces the probability of securing a position in a top world-ranked economics department, it increases the likelihood of employment in a top Italian economics department, suggesting that administrative data access facilitates a more domestically oriented career trajectory.

Finally, while administrative data access benefits early-career researchers, not all scholars capitalise on this opportunity equally. Those from top-ranked PhD programs appear particularly skilled at translating data access into career gains, reinforcing existing academic hierarchies. In contrast, despite stronger research output, women do not experience proportionally greater career progression, suggesting that hiring decisions may apply different standards to research signals on the basis of gender. Our findings indicate that rather than levelling the playing field, administrative data access tends to reinforce existing disparities, benefiting those already well-positioned in the academic hierarchy.

Our article makes notable contributions to several strands of the literature. First, it adds to the growing body of research examining the role of academic resources and signals in shaping the career trajectories of young researchers in Economics. Existing studies have highlighted the impact of factors such as obtaining a PhD from a prestigious institution (Waldinger, 2010; Jones and Sloan, 2024), publishing in a Top-5 journal (Heckman and Moktan, 2020), co-authoring with top scholars (Azoulay *et al.*, 2010; Azoulay *et al.*, 2019; Li *et al.*, 2019) or more generally, academic social ties (Colussi, 2018; Anderson and Richards-Shubik, 2022; Ductor and Visser, 2022). Our paper complements this evidence by demonstrating that access to administrative data – an increasingly valuable but still relatively scarce resource – can also play a significant role in career progression.

Second, our study contributes to the literature on sociodemographic disparities in academic careers. Prior work has documented the effects of gender, race, and socioeconomic background on hiring, promotion, and professional recognition in economics (Bosquet *et al.*, 2019; Bransch and Kvasnicka, 2022; Ajzenman *et al.*, 2023; Ebehardt *et al.*, 2023; Hale *et al.*, 2023; Carrell *et al.*, 2024; Gualavisi *et al.*, 2024; Jones and Sloan, 2024). By exploring whether administrative data access benefits all researchers equally or not, our paper provides new insights into how research opportunities interact with broader patterns of inequality in academia.

Third, our study speaks to the policy design and governance of research infrastructure. Administrative data are typically collected and maintained by public institutions, yet access policies often remain opaque, uneven, and resource-intensive to navigate. Public programs like VisitINPS effectively serve as gatekeepers to valuable data resources. Understanding who benefits most from such programs is therefore central to evaluating their distributional impact – not just on research outcomes but on the composition of the academic community itself. Our findings suggest that when public data access programs unintentionally reinforce pre-existing academic hierarchies, they may undermine the goals of equity and inclusivity in science policy. These findings underscore the importance of intentional design and evaluation of data access initiatives to promote inclusivity and broaden participation, particularly among underrepresented groups and institutions.

Finally, our study contributes to the literature emphasising the importance of data access for scientific progress and evidence-based policymaking. While previous research has shown that greater access to data enhances the quality, quantity, and impact of economic research (Angrist *et al.*, 2020; Cole *et al.*, 2020; Nagaraj *et al.*, 2020; Nagaraj and Tranchero, 2023), our paper adds to this evidence by examining how access to administrative data shapes the career trajectories of young scholars.

The remainder of the paper is organised as follows. Section 2 presents a short and simple model to provide insights on the incentives to access and work with administrative data. Section 3 describes the VisitINPS program and the longitudinal dataset we put together and analyse in this paper. Section 4 presents the empirical strategy and the estimation sample. Section 5 presents and discusses the empirical findings. Section 6 concludes.

2. On the Incentives to Access Administrative Data: Some Theoretical Considerations

This section presents a model of hiring probabilities for applicants competing for an academic position. We start with a model to analyse how access to administrative data affects the perceived quality of applicants and their subsequent hiring probabilities, within a simplified winner-takes-all selection mechanism. We then extend this model to illustrate how accessing administrative data may reinforce pre-existing differences in research careers or instead act as an equaliser.

2.1. Administrative Data, Publication, Potential and Academic Promotion

Consider a set of N applicants competing for a single position. Each applicant i has a latent, true quality that is unobservable to the employer. Instead, employers e observe a perceived quality signal s_i for each applicant based on information usually included in application packages, which is a noisy measure of the applicant's true quality. This perceived quality signal s_i is the sum of two components μ_i and ϵ_i. μ_i is the employer's assessment of applicant i's quality, which is an increasing function of her published work P_i and potential V_i. The former represents past academic output and is perfectly observable while the latter captures forward-looking attributes, inferred from indicators such as working papers, conference participation or professional networks. $\epsilon_i \sim N(0, \sigma^2)$ is a random noise term capturing unobservable factors affecting the employer's perception of applicant i. The perceived quality signal s_i for applicant i is hence defined as follows:

$$s_i = \mu_i + \epsilon_i = \alpha P_i + \beta V_i + \epsilon_i \tag{1}$$

where P_i and V_i are separately additive (relaxing this assumption makes the model more complex but eventually does not alter its predictions). α and β represent the employer's valuation of published work and potential, respectively. As stated earlier, we assume both are strictly positive. For simplicity, we initially treat α and β as constant across employers and applicants, though we relax this assumption later. The employer's hiring decision follows a winner-takes-all mechanism, where only the applicant with the highest perceived quality signal s_i is selected. Consequently, the hiring probability π_i for applicant i is defined as the probability that s_i exceeds the signals of all N competing applicants:

$$\pi_i = \mathrm{P}(s_i > s_j \; \forall j \neq i)$$

For simplicity, consider the case with two applicants, i and j, both with normally distributed signals $\epsilon_i \sim N(0, \sigma^2)$ and $\epsilon_j \sim N(0, \sigma^2)$. Then, the probability that s_i is higher than s_j can be expressed as:

$$\pi_i = \mathrm{P}(s_i > s_j \; \forall j \neq i) = \Phi\left(\frac{\mu_i - \mu_j}{\sqrt{2}\sigma}\right) \tag{2}$$

For a given applicant i, the strategy to secure the position is straightforward: achieve the highest signal s_i. While various approaches could improve this signal, we focus solely on the decision to invest in administrative data access, denoted by the dummy variable D_i, which takes a value of one if the applicant has access to administrative data, and zero otherwise, *ceteris paribus*. Administrative data comes with potential benefits and costs. On the benefit side, it can enhance the perceived quality signal by possibly improving the quality of published work P_i and/or potential V_i. We specify:

$$P_i(D_i) = P_{0i} + \delta_P D_i \tag{3}$$

$$V_i(D_i) = V_{0i} + \delta_V D_i \tag{4}$$

P_{0i} and V_{0i} are the baseline levels of published work and potential for applicant i, respectively. δ_P and δ_V capture the 'potential effects' of administrative access on published work and perceived potential, respectively. We use the term 'potential effects' because δ_P and δ_V may be positive or null (we find it implausible that they could be negative). However, administrative data entails fixed costs that are usually larger than those borne by survey data users. For instance, obtaining initial access to administrative data often involves competitive selection processes and bureaucratic hurdles. Furthermore, the analysis of administrative data is also frequently restricted to designated workspaces with limited access windows. We assume the difference in cost between administrative data and survey data to be equal to C^D and strictly positive.

The applicant's expected utility depends on whether they choose $D_i = 1$ (with administrative data access) or $D_i = 0$ (without administrative data access). In the former case, the expected utility of the applicant is:

$$E(U_i|D_i = 1) = \pi_i(D = 1) * A - C^D.$$

In the latter case, it is:

$$E(U_i|D_i = 0) = \pi_i(D = 0) * A$$

with A representing the utility derived from getting the position. An applicant will decide to access and analyse administrative data if $E(U_i|D_i = 1) > E(U_i|D_i = 0)$. Normalising $A = 1$, this inequality can be developed as follows:

$$\pi_i(D = 1) - C^D > \pi_i(D = 0)$$

$$\pi_i(D = 1) - \pi_i(D = 0) > C^D \tag{5}$$

In simple terms, the decision to work with administrative data boils down to whether the rise in hiring probability exceeds the costs. Using (2), the rise in hiring probabilities caused by administrative data can be written as:

$$\Delta\pi_i = \pi_i(D = 1) - \pi_i(D = 0) = \Phi\left(\frac{\mu_i(D=1)-\mu_j}{\sqrt{2}\sigma}\right) - \Phi\left(\frac{\mu_i(D=0)-\mu_j}{\sqrt{2}\sigma}\right). \tag{6}$$

Expanding Φ around $\frac{\mu_i(D=0)-\mu_j}{\sqrt{2}\sigma}$ using a Taylor expansion, we get:

$$\Phi\left(\frac{\mu_i(D=1)-\mu_j}{\sqrt{2}\sigma}\right) \approx \Phi\left(\frac{\mu_i(D=0)-\mu_j}{\sqrt{2}\sigma}\right) + \phi\left(\frac{\mu_i(D=0)-\mu_j}{\sqrt{2}\sigma}\right) * \frac{\Delta\mu_i}{\sqrt{2}\sigma}. \tag{7}$$

Combing (6) and (7), the approximate change in hiring probability due to administrative data is:

$$\Delta\pi_\mathrm{i} \approx \phi\left(\frac{\mu_i(D=0)-\mu_j}{\sqrt{2}\sigma}\right) * \frac{\Delta\mu_i}{\sqrt{2}\sigma} \tag{8}$$

Access to administrative data enters this equation via the expected increase in perceived quality due to data access $\Delta\mu_i$ which, according to (1), (3) and (4), is equal to:

$$\Delta\mu_i = \mu_i(D_i = 1) - \mu_i(D_i = 0) = \alpha\delta_P + \beta\delta_V \tag{9}$$

Combining (8) and (9), we get that an applicant i will work with administrative data when:

$$\Delta\pi_i \approx \phi\left(\frac{\mu_i(D=0)-\mu_j}{\sqrt{2}\sigma}\right) * \frac{\alpha\delta_P+\beta\delta_V}{\sqrt{2}\sigma} > C^D. \qquad (10)$$

In short, (10) summarises the idea that the decision of applicants to invest in administrative data access depends on whether it will improve their perceived quality signal and, consequently, their chances of being hired.

A key point worth discussing is the contrast that may exist in practice between the predictability of costs and the uncertainty surrounding potential gains. The costs, captured by C^D, are typically known ex-ante or at least reasonably predictable. Applicants can factor these costs into their decision-making with a fair degree of confidence. By contrast, the gains are far less predictable, as they depend on several factors that are harder to quantify or anticipate. These include the applicant's ability to leverage administrative data to improve their signal (δ_P and δ_V), as well as the employer's valuation of enhanced published work α and potential β.

Additionally, factors such as noise in the hiring process (σ) and the quality of the competitors (μ_j) further complicate the decision to work with administrative data. However, since these factors are largely beyond the control of applicant i they are here treated as exogenous.

In our empirical setting (detailed in Section 3), all researchers have chosen to apply for administrative data access, implying that they had already assessed the expected benefits to outweigh the costs. This suggests that, despite the uncertainty surrounding the potential returns, applicants viewed administrative data as a strategic investment in their academic trajectory. However, the key question remains: do these expected benefits actually materialise? The first part of our empirical analysis examines whether access to administrative data leads to measurable improvements in publication records (P_i) and perceived potential (V_i), and, crucially, whether these enhancements translate into greater career prospects, such as securing an academic position.

2.2. Administrative Data and Academic Inequalities

Thus far, our analysis has treated applicants as essentially homogeneous, with δ_P, δ_V, α and β being equal for all applicants. In reality, they might differ along many dimensions that may lead them to extract more (or less) value from administrative data. Including these differences not only enriches the model but also speaks to the possibility of Matthew effects, whereby already-advantaged scholars reap disproportionate gains from a new resource, and to its converse, where a novel opportunity helps level the playing field.

To incorporate these possibilities, consider the framework above but allow for two types of applicants, indexed by $r \in \{A, B\}$. Substituting these types into equations (3) and (4) yields

$$P_i(D_i) = P_{0i} + \delta_{P,r} D_i \tag{11}$$

$$V_i(D_i) = V_{0i} + \delta_{V,r} D_i \tag{12}$$

where $\delta_{P,r}$, and $\delta_{P,r}$ need not be uniform across all researchers. In line with equation (9), the change in μ_i^r from gaining data access is

$$\Delta\mu_i^r = \alpha\delta_{P,r} + \beta\delta_{V,r} \tag{13}$$

If type A researchers initially benefitted from a more privileged situation because of factors orthogonal to access to administrative data and enjoy systematically larger $\delta_{P,A}$ and $\delta_{V,A}$ than do type B, the perceived quality gap expands once both acquire administrative data, in line with a Matthew effect. Conversely, if type B benefits from sufficiently large increments relative to A, disadvantaged applicants can narrow the gap.

A further extension allows α and β themselves to vary by type, say α_r and β_r . Then Equation (13) becomes

$$\Delta\mu_i^r = \alpha_r\delta_{P,r} + \beta_r\delta_{V,r} \tag{14}$$

Under this scenario, even the same incremental improvements in published output or potential can be weighted differently by potential employers. If $\alpha_r\delta_{P,A} + \beta_r\delta_{V,A}$ is significantly larger than $\alpha_r\delta_{P,B} + \beta_r\delta_{V,B}$, then type A gains more from administrative data, amplifying its advantage. If, instead, disadvantaged applicants receive higher increments or more favorable weighting, they can partially offset baseline gaps.

3. The VisintINPS Program and our Data

3.1. The VisitINPS Program

In 2016, the Italian Social Security Institute (INPS) launched the VisitINPS Scholars program (henceforth, VisitINPS) to provide researchers with access to its extensive administrative data archives. The program allows social scientists to conduct research on topics related to social security, pensions, labour markets, income support policies, and welfare systems, fostering empirical research on policy-relevant issues. Selected scholars are invited to work on-site at INPS offices, where they can directly access and analyse administrative data.

Central to the program's framework is the submission and evaluation of research proposals by prospective scholars, acting as Principal Investigator (PI). Each year, PIs submit proposals outlining their research objectives, methodologies, and broader significance of their work within the context of welfare studies. For each call, these proposals undergo a rigorous evaluation process conducted by a Committee comprising the scientific coordinator of the VisitINPS program and two expert reviewers appointed by INPS, who assess each submission based on predefined criteria. Each proposal is then assigned a numerical score ranging from 0 to 100, reflecting both the quality of the research proposal and the applicant's professional and academic experience. The research proposal itself accounts for 70 percent of the score and is evaluated based on scientific rigour, feasibility, contribution to knowledge, and policy relevance. The remaining 30 percent is determined by the applicant's research track record, including prior experience and publication history.

To be eligible for selection, a proposal must meet a minimum score threshold, initially set at 50 in 2016 and raised to 60 in subsequent years. However, reaching this threshold does not guarantee selection, as the final cutoff fluctuates annually due to variations in the number of applications and workspace availability at INPS offices. This means that a proposal with a qualifying score in one year may fall below the selection threshold in another, introducing an element of unpredictability into the process.

Moreover, the VisitINPS program offers three distinct avenues of support: Fellowships, Scholar Type A, and Scholar Type B. These avenues allow scholars to tailor their engagement with the program according to their specific needs and circumstances, whether seeking financial assistance for accommodation or additional resources. While the program grants access to administrative data through a PI role, it is important to note that young researchers can also gain access by serving as 'Collaborators' on accepted projects. We will address any potential confounding effects of this alternate route in our robustness checks.

Appendix Figure A1 presents a network analysis of the institutional affiliations of PIs of the VisitINPS applicants, illustrating the program's reach beyond Italy. Each node represents an institution, with larger nodes indicating greater centrality in the collaboration network, while edge thickness reflects the strength of institutional connections. Appendix Figure A1 confirms that, while a large share of applicants comes from Italian universities (blue nodes), VisitINPS has also attracted researchers from prominent international institutions (green nodes), including IZA, CEPR, NBER, and leading European and U.S. universities.

3.2. Our data: the academic biographies of the applicants

With the support of INPS, we were granted access to the list of all VisitINPS proposals since the program's inception in 2016. For each proposal, we were given information on the name of the applicant (or PI), the title of the project, the score assigned by the reviewers, the submission year, and the specific program to which the proposal was submitted, and the final status of the project (i.e. 'non-eligible', 'eligible but not selected' or 'eligible and selected'). Note that very few selected applicants ended up not enrolling in the VisitINPS program. These few cases were excluded from our analysis.

To complement this dataset, we undertook an extensive data collection effort, obtaining the most recent resumes of all PIs and web-scraping professional and institutional websites. This process involved parsing a diverse array of information about their academic life and output, which we systematically aggregated at the yearly level. For each PI, we reconstructed a panel database spanning from 2010 to 2023. This database, which we refer to as the *academic biographies*, includes the PI's main job title, the number of peer-reviewed articles published, the specific journals where these articles were published, and the number of co-authors on their published works at the end of each calendar year. Furthermore, it tracks when and where the PIs got their PhD degree and the institutions where PIs had worked throughout their careers.

We further enriched this dataset by reaching out to the organising committees of the European Association of Labour Economists (EALE) and the Italian Association of Labour economics (AIEL) conferences to check whether PIs had presented their work at their annual meetings. Furthermore, we examined whether PIs had published a working paper in the VisitINPS working paper series. Using Google Scholar and Scopus, we also recorded all types of manuscript, whether peer-reviewed or otherwise, based on administrative INPS data published by the PIs as well as yearly citations counts.

This comprehensive data collection effort culminated in a database comprising the quasi-universe of PIs who submitted a VisitINPS proposal (419 out of 443 unique PIs).[1] This invaluable resource provides unprecedented depth of insight into the academic biographies, or professional trajectories, of VisitINPS applicants, that we will use to conduct our empirical analysis.

4. Identification Strategy and Estimation Sample

Our identification strategy relies on a Two-Way Fixed-Effect (TWFE) approach, with our main regression model formulated as follows:

$$Y_{it} = \beta * VisitINPS_{it} + \alpha_i + \lambda_t + \epsilon_{it}. \tag{15}$$

Here, Y_{it} is a series of academic outcomes for researcher i in year t. These academic outcomes are divided into five different sets. First, Y_{it} includes indicators of whether researcher i has engaged with INPS data. Specifically, we use the yearly probability of producing at least one working paper in the dedicated VisitINPS Working Paper Series and the probability of having at least one manuscript that includes INPS data (regardless of the publication series). These variables serve as 'first-stage' outcomes, capturing whether selection into the VisitINPS program translates into tangible research outputs using INPS administrative data.

[1] The 5% missing PIs were individuals with minimal online presence, rendering data collection almost impossible. As a result, we have limited information about these missing PIs, raising the question of how our conclusions might change if data on them were available. We believe that minimal online presence may indicate that these PIs have left academia. If this is indeed the case, our estimated effects of accessing and using INPS administrative data through the VisitINPS program likely represent lower bounds of the true impact – in particular because the missing PIs did not submit proposals with scores sufficiently high to get access to administrative data.

We then assess whether $VisitINPS_{it}$ impacts conference participation. Given the nature of INPS data and the projects submitted to the VisitINPS program, we expect increased attendance primarily at labour economics conferences.[2] We track two specific participation indicators: attending either the EALE or the AIEL annual conference. EALE's general conference is widely regarded as Europe's most prestigious gathering for labour economists, while AIEL holds similar status in Italy.[3] This second set of outcomes allows us to capture the immediate and early signals of scholarly engagement and visibility resulting from administrative data access, setting the stage for the analysis of longer-term career and publication outcomes.

Next, we assess the effect of accessing administrative data on publication outcomes. Here, we focus on three primary variables. First, we look at the effect of $VisitINPS_{it}$ on the yearly probability of publishing at least one peer-reviewed journal article. Second, we consider the yearly probability of publishing in a high-rank journal, defined as a journal within the Top 30 Economics journal according to the aggregate RePec ranking (as of December 2024). Third, we explore whether $VisitINPS_{it}$ increases the probability of publishing an article in a top-field journal in Labour economics (i.e. in alphabetical order - Journal of Human Resources, Journal of Labor Economic and Labour economics). In the Online Appendix, we also analyse the likelihood of publishing in journals classified as Category 'A' in the official Italian academic journal rankings for Economics.

[2] We focus on the field of labour economics because the research produced by VisitINPS applicants predominantly aligns with topics in this field. As shown in Appendix Figure A2, a word cloud of paper keywords from early-career researchers highlights a strong emphasis on labour economics, as well as themes related to causality and policy evaluation – areas where administrative data are particularly valuable.

[3] We also considered participation to the conferences of the European Society for Population Economics (ESPE) and of the European Economic Association (EEA). However, more than 97% of the researchers in our estimation sample never attended any of these two conferences.

Fourth, we examine the impact of administrative data access on academic impact and professional networks. Beyond direct research outputs, an important aspect of academic success is the extent to which a scholar's work gains recognition and connections within the profession. We assess this by considering three key indicators. First, we analyse the effect of administrative data access on the annual number of Google Scholar citations. While citations primarily reflect the visibility and influence of published research, they can also serve as a proxy for a scholar's overall academic reach, including the extent to which working papers and conference presentations gain traction among peers. Given that the publication process in high-ranking journals is lengthy, citation counts may provide an early indication of research impact before formal publication occurs. Second, we investigate whether access to administrative data increases the likelihood of collaborating with highly ranked economists, defined as economists ranked in RePEc Top-100 world ranking and/or within labour economics over the last 10 years. This measure reflects the idea that access to administrative data may facilitate collaborations with leading researchers in the field by enabling more ambitious projects or providing networking opportunities. Finally, we consider a more localised dimension of professional connections by examining whether VisitINPS participants are more likely to co-author with highly ranked economists based in Italy (using again Top-100 economics based in Italy over the last 10 years according to RePec rankings). Given that INPS data pertains specifically to the Italian labour market, it is possible that researchers accessing these data develop stronger ties with domestic scholars.

The last family of academic outcomes we investigate refers to the career evolution. We start with the probability of receiving an academic promotion during the year. Since our estimation sample consists only of PhD students and Postdoctoral fellows at the time of their application to VisitINPS, PhD student advancing to at least a postdoctoral position or a postdoctoral fellow transitioning at least to an assistant professorship is considered to be a promotion. Note that the tenure status is near impossible to parse in most cases. Additionally, we consider the probabilities of working in an Economic department ranking in the Top-20 globally or in one of the five top ranked Economics department in Italy (again according to RePec ranking over the last 10 years).

The choice of all these various academic outcomes is motivated by two key reasons. First, they closely map the key components of our theoretical model (see Section 2). To recap briefly, in our model the perceived quality signal and hiring probability of applicant i depends on published work P_i and potential V_i. Publications in peer-reviewed journals are indicators of P_i while working papers based on INPS data represent future publications and contribute to V_i, as they reflect the applicant's ability to transform administrative data access into better signals (δ_P and δ_V). Conference participation also serves as a measure of V_i as it signals both ongoing research that meets conference selection standards and increased exposure to influential networks. Citation counts have a hybrid status: while citations originate from published work (in peer-reviewed journals or not), they may also signal the researcher's potential to recruiters. Connections to highly ranked economists follow a similar rationale. Finally, our measures of career development directly capture the hiring process, which is central to the theoretical model in Section 2.

Second, our set of outcomes allows us to examine whether access to administrative data from a specific country (in our case, Italy) facilitates the development of an international academic career or one more closely tied to the country providing the data. We capture both career trajectories through distinct outcomes. Indicators such as participation in EALE, publication in top-field economics journals, citation counts, and connections to highly ranked economists serve as proxies for the development of an international career. Conversely, outcomes such as participation in AIEL, publication in highly ranked journals according to the Italian classification, and a higher probability of securing a position in a top Italian economics department reflect a more domestically oriented career trajectory. While these two career paths are not mutually exclusive, comparing how each set of outcomes responds to administrative data access provides insight into which trajectory is more strongly influenced.

Turning now to the independent variables, the terms α_i and λ_t in Equation (11) are individual and year fixed-effects, respectively. $VisitINPS_{it}$ is a dummy variable equal to one when researcher i enrolls in the VisitINPS program and begins accessing INPS administrative data in year t. Notably, $VisitINPS_{it}$ is an absorbing state, implying that once a researcher is enrolled in the program in year t, she remains enrolled throughout the observation period. The coefficient β captures the effect of interest. Standard errors are clustered at the individual level.

Given that our dataset encompasses all young researchers who submitted proposals to the VisitINPS program, our sample comprises only two treatment statuses: the never-treated (individuals who applied but were never accepted into the program) and those who eventually received treatment (applicants who were accepted into the program). Note that there are no always-treated individuals in our sample due to our panel starting in 2010, which predates the start of the VisitINPS program in 2016.

While the absence of always-treated individuals simplifies our identification strategy to some extent, we estimate Equation (11) using the Callaway and Sant'Anna (2021) estimator. This estimator is particularly suited for our setting, where treatment timing varies across researchers who join the VisitINPS program in different years. In our robustness checks, we demonstrate that our conclusions are not sensitive to the choice of the estimation technique. Similar results emerge when using standard OLS, as well as the de Chaisemartin and D'Haultfoeuille (2020) estimator. Results for these alternative specifications are reported in Online Appendix Table A1.

We use the *academic biographies* collected for VisitINPS candidates, as detailed in Section 3, to estimate Equation (11). Our main estimation sample comprises researchers at the outset of their careers, specifically PhD students or Postdoctoral fellows, at the time of their VisitINPS program application. Exploiting the panel structure of our dataset, we incorporate yearly outcomes for each researcher from six years prior to submitting a proposal to six years afterward. Given the possibility of multiple applications (across different sub-programs and years) by the same researcher, we implement a series of adjustments to ensure the robustness of the analysis. For instance, for those who applied multiple times without success, we focus on the six years surrounding the closest instance to the selection threshold. Conversely, for applicants who were eventually selected, we concentrate on the years surrounding their successful application. Furthermore, for respondents who applied and were selected multiple times, we retain observations only for the years surrounding their initial success. Robustness checks exclude respondents who attempted to access the data multiple times, demonstrating consistent results.

This process yields a total estimation sample of 2,087 observations for 203 applicants, with each applicant observed between nine and twelve times. Descriptive statistics at the observation level are detailed in the first column of Table 1. Approximately 41% of the observations are women, while 20% are postdoctoral fellows. The probability of publishing at least one VisitINPS working paper or any type of manuscript based on INPS data stands at 2% and 4%, respectively. The probability of publishing a peer-reviewed article is 22%, while the probabilities of publishing in a Top-30 Economics journal and a top-field journal in labour economics are 2% and 0.6%, respectively. Conference participation rates for EALE and AIEL are 3.7% and 6.4%, respectively, while the likelihood of receiving a promotion during the year equals 9%. The annual average citation count per researcher is 11.6.

Columns (2) and (3) of the same Table present analogous descriptive statistics before and after application to the VisitINPS program. Although our sample is not perfectly balanced, the proportion of women remains constant over time. However, all the other variables significantly increased post-application to VisitINPS. The systematic differences observed in column (4) reflect the evolving careers of young economists, characterised by greater publication output, increased conference participation and citation count, and eventual promotions. The only exception is the drop in the probability of working in high-ranked Economics department (both Italian and international), which is consistent with the idea that securing a stable position in a high-rank department is difficult. The estimation of Equation (11) allows to assess the extent to which the VisitINPS program contributed to the development of researchers' careers observed in column (4).

The use of the VisitINPS applicant pool offers distinct advantages in mitigating unobserved heterogeneity between those who accessed the administrative data and enrolled in the program and those who did not. Given that all applicants were at the beginning of their careers when applying to the program, possessed awareness of its existence, were capable of crafting research proposals and shared an interest in labour economics, there is a relatively high degree of homogeneity among applicants (as opposed to more experienced researchers or economists with diverse interests). However, eligibility for the program and subsequent access to INPS administrative data were not random. Only applicants with the highest proposal scores were granted access, meaning that those selected may possess higher technical skills or research aptitude. This selection concern is illustrated in the two panels of Figure 1. Panel A displays the distribution of scores for the 203 INPS proposals included in our sample, showing a roughly normal distribution ranging from 20 to 95, with a mean around 65. Panel B separates scores between selected and non-selected applicants, revealing that non-selected applicants have lower average scores. This confirms the concern that 'treated' applicants (i.e. those who accessed the administrative data) may be positively selected.

To address this selection issue, Equation (11) incorporates individual fixed effects to control for time-invariant heterogeneity across applicants, such as innate differences in research ability or inherent quality. In addition, the inclusion of time fixed effects controls for temporal shocks and common environmental factors, like changes in the academic job market during the COVID-19 pandemic. Nevertheless, we employ two additional strategies to mitigate the potential bias from non-random selection. First, Table 2 presents balance tests comparing pre-application variables, measured one year prior to application, between non-selected and selected applicants. Reassuringly, 14 out of 15 differences in means between the two groups are not statistically significant. However, our sample size is relatively small, and we may lack statistical power to detect small differences between groups. However, irrespective of statistical power limitations, the inclusion of individual fixed effects in Equation (11) allows to partial out the effect of any systematic differences between applicants. Second, we draw inspiration from the Regression-Discontinuity Design (RDD) approach and conduct a TWFE-RDD analysis to address any remaining concerns related to non-random selection. This approach compares applicants with similar scores, with some scoring just above the threshold to gain access to the VisitINPS program. By leveraging this variation, $VisitINPS_{it}$ likely captures exogenous differences in INPS reviewers' scale-use rather than inherent differences in applicant quality. However, due to the already limited size of our estimation sample, the TWFE-RDD approach is kept for robustness checks.

5. Empirical Analysis

5.1. The Average Effect of Administrative Data on Academic Outcomes

Table 3 reports the impact of administrative data access through the VisitINPS program on a range of academic outcomes. Column (1) presents the Callaway and Sant'Anna (2021) aggregate TWFE estimates, while column (2) reports the p-value for the pre-treatment effect – which indicates whether pre-trends pose an identification issue. Notably, all p-values in columns (2) are greater than 0.1, providing support for the validity of the parallel trend assumption. Detailed event-study graphs in the Online Appendix (Figures A3 to A15) further confirm this, as none of them reveal systematic and significant pre-trend differences. Columns (3) and (4) respectively show the number of years after accessing the administrative data when significant effects start to appear and disappear.

As anticipated, participation in the VisitINPS program and access to administrative data lead to an increase in the publication of working papers in the program's series and manuscripts based on INPS data. These initial outcomes validate the idea that administrative data access can generate academic research and enhance the potential of young researchers. Significant effects for these two outcomes emerge two years after joining the program - as shown in column (3). Such timing is typical of academic projects, where data access precedes manuscript drafting by a plausible two-year gap.

Next, we examine whether access to administrative data affects labour economics conference participation. The third and fourth rows of Table 3 indicate that accessing INPS data significantly increases the probability of participating in the Italian AIEL conference, with positive effects discernible as early as one year post-program enrolment. In contrast, we find no effect for participation to EALE conferences. This may indicate that researchers may prefer presenting preliminary findings at more regionally-focused conferences. Taken together, the results presented so far confirm that access to administrative data enables young researchers to improve the signal they send to recruiters by demonstrating greater potential.

We then evaluate the impact of VisitINPS program participation on publication records. Row five of Table 3 reports the treatment effect on the annual probability of publishing at least one article in any peer-reviewed journal. The estimates indicate no significant effect on this outcome, either on average or at any specific time. A similar result emerges when examining the probability of publishing in Top-30 economics journals (as ranked by RePec), with no discernible impact from data access. This is perhaps unsurprising, given that Top-30 economics journals are typically general-interest outlets, whereas INPS administrative data specifically focus on the Italian labour market, which may not align well with the broader scope of these journals. Given that INPS data are heavily centered on labour market outcomes, we test whether they are more likely to affect the probability of publishing in top-field labour economics journals. Row seven of Table 3 reveals a positive and statistically significant effect at the 1% level, supporting this hypothesis. These findings suggest that administrative data may not affect publication volume *per se* but they ultimately enhance the quality of publication records over a long horizon.[4]

We now turn to our various measures of academic impact and professional networks. The pattern observed in citation counts closely mirrors that of publications in top field labour economics journals, with a statistically significant increase emerging three years after accessing the data.

[4] Online Appendix Table A2 provides a detailed analysis of the effect of accessing administrative data through the VisitINPS program on peer-reviewed publications. Journals ranked as 'A' in the official Italian classification (for Area 13) are divided into four subcategories: 'Economics,' 'Business,' 'Economic History,' and 'Statistics and Mathematical Methods for Decision-Making.' It is important to note that these categories are not mutually exclusive; in practice, substantial overlap exists, with most journals assigned to at least two subcategories. In column (1) of Online Appendix Table A2, we first show that accessing administrative data increases the probability to publish in journals ranked as 'A'. We then investigate whether this increase is driven more by one subcategory than another. The results indicate that while the effect of data access appears slightly stronger for 'Economics' and 'Economic History' journals, the differences across subcategories are not statistically significant. This finding is consistent with the extensive overlap between categories, which dilutes any clear distinction.

However, working with administrative data through the VisitINPS program does not appear to increase the likelihood of collaborating with highly ranked economists, whether based in Italy or abroad.

Finally, we investigate whether access to administrative data impact researchers' career trajectories. The eleventh row of Table 3 shows that accessing INPS administrative data increases the probability of receiving a promotion. The rise in promotion becomes statistically significant two years after program enrolment and data access. This result has several implications. First, it suggests that administrative data does indeed strengthen the signal young researchers send to potential employers, ultimately making them more likely to secure a new position – thus justifying a posteriori the decision to work with administrative data. Second, the timing of these effects suggests that the career advancement observed is unlikely to be driven by enhanced publication records, which manifest later. Instead, program participants, armed with working papers or drafts based on INPS data and conference participation, may leverage these outputs during recruitment seminars to emit a stronger signal through potential. While not yet published, employers may perceive their work as likely to result in eventual publication in prestigious journals.

The last two rows of Table 3 complete this academic trajectory by highlighting a geographic shift in career advancement. Access to Italian administrative data reduces the probability of securing a position in a world top-ranked economics department but increases the probability of employment in a top-ranked Italian economics department by an equivalent magnitude. Columns 3 and 4 suggest that these career shifts take place a few years after the initial promotion, consistent with a typical academic career path: young researchers first secure short-term contracts that elevate them to a higher academic rank (e.g., from PhD to postdoc, or from postdoc to assistant professor). Over time, as their publications in top labour economics journals begin to materialise, their profiles strengthen, signalling their ability to convert research potential into high-impact publications. This ultimately makes them competitive candidates for positions in Italy's leading economics departments.

Overall, the estimated effects of administrative data access are not only statistically significant but also economically large. Drawing on the average effects reported in Table 3 and Appendix Table A1 across different estimation techniques, we find that access to administrative data increases the probability of publishing in a top field labour economics journal by 2.7 to 6.3 percentage points. Similarly, it raises yearly citation counts by approximately 10 citations per year and boosts the probability of academic promotion by 7.2 to 11.3 percentage points. These results provide a tangible measure of the advantage conferred by administrative data access, reinforcing the idea that such access enhances research productivity and visibility, and eventually offers a comparative advantage on the academic job market.

The findings in Table 3 also provide insight into the types of academic careers that administrative data access fosters. Specifically, access to Italian administrative data appears to facilitate prestigious career advancement within Italy rather than internationally. Several factors could explain this pattern, operating on both the demand and supply sides of the job market. On the supply side, a majority of VisitINPS applicants are Italian nationals, many of whom may actively seek academic positions within their home country. On the demand side, Italian research institutions may be more likely to hire candidates who not only demonstrate strong research potential but also have a higher likelihood of remaining in Italy over time. Consequently, researchers with experience working with INPS data (and hence greater potential) may be particularly attractive hires for Italian institutions, reinforcing the observed career path toward leading economics departments in Italy. This effect is likely further amplified by policies such as

Italy's "Rientro dei Cervelli" (Brain Gain) program, which provides substantial tax rebates to highly educated Italian researchers returning from abroad (Bassetto and Ippedico, 2024).

5.2. Robustness Checks

Table 4 presents a series of robustness checks to evaluate the reliability of our findings. Column (1) reports the baseline estimates from Table 3, serving as a benchmark for the rest of this subsection.

A first legitimate concern is that VisitINPS applicants who were initially unsuccessful may have re-applied to the program, potentially learning from prior applications and thereby biasing our results. We address this issue in column (2) by excluding PIs who submitted multiple applications, which removes 21 individuals (192 observations in total) from our sample. Despite this reduction, our conclusions remain robust, with all point estimates retaining their magnitude.

In the remaining columns, we tackle the issue that eligibility for the program – and subsequent access to INPS administrative data – is not random within our sample, as only applicants with the highest proposal scores were granted access. Although we use individual fixed effects and documented a reasonable balance of covariates between selected and non-selected applicants in Table 2, we further address this selection concern by leveraging the scores received on VisitINPS applications.

First, we remind that the acceptance threshold for the VisitINPS program varied over time and across programs, influenced by both demand-side and supply-side factors (e.g., available workspace for INPS data users, total number of applicants). As a result, a proposal with a score of, say, 70, might be accepted in one year and rejected in another. In column (3), we re-estimate

Equation (11) using only applicants whose scores fell between the lowest and highest acceptance thresholds across all VisitINPS programs from 2016 to 2022. This sample construction implies that non-selected applicants might have been accepted in a different year, while selected applicants could have been denied access had they applied at another time. In essence, we here only leverage variation in access to administrative data stemming from reasons that are orthogonal to the quality of the project. By doing so, we reduce our sample by more than a third and, as a by-product, we also exclude proposals at the top and the bottom of the score distribution. Column (3) of Table 4 reveals that all our conclusions remain unchanged with this sample.

In the last three columns of Table 4, we adopt the TWFE-RDD approach we mentioned in Section 4, keeping only respondents whose proposals received scores close to the selection threshold. As in any RDD, the narrower the bandwidth, the more similar the treated and non-treated units become, with differences in access to administrative data presumably stemming from exogenous differences in scoring style among INPS reviewers. However, narrower bandwidths reduce statistical power. Therefore, we use three distinct bandwidths: 7.5 points, 10 points, and 12.5 points below and above the selection threshold.

Overall, the TWFE-RDD estimates corroborate our main findings, indicating that applicants selected for the VisitINPS program and granted access to administrative data are more likely to produce manuscripts based on INPS data, participate in conferences, and eventually secure promotions. The magnitudes of significant point estimates from our baseline specification generally remain stable, although standard errors increase due to the reduced sample size in each bandwidth. Nevertheless, statistical significance is preserved in most cases. Taken together, these robustness checks indicate that selection issues related to program eligibility and subsequent access to INPS administrative data are unlikely to confound our conclusions.

A final concern relates to the possibility that some junior researchers in our estimation sample held the role of 'Collaborator' on other accepted VisitINPS projects, thereby granting them access to administrative data without serving as the designated PI. This scenario could introduce a bias in our estimates if, for instance, Collaborator status carries its own distinct career advantages or systematically differs across treatment and control groups. It might also confound the effect of data access through PI status if the same individuals who appear in our sample of young PIs were also frequent collaborators on more established researchers' projects, in which case the observed impact on career outcomes would partly reflect the benefits of such collaborations rather than data access alone. Fortunately, INPS provided us with information on all Collaborators in accepted projects, enabling an explicit assessment of these concerns. In Appendix Table A3, we first reproduce our baseline specification and then add a dummy in Column (2) for whether a researcher also became a Collaborator on another project, thus holding constant any potential time-varying confounding influence. In Column (3), we adopt a more conservative approach by excluding all individuals who served as Collaborators. In both cases, our results remain quantitatively and qualitatively consistent with the baseline, indicating that any additional benefits of Collaborator status do not drive the estimated gains from direct data access through the VisitINPS program.

5.3. Administrative Data: Reinforcing Hierarchies or Levelling the Field?

The econometric approach adopted so far allowed us addressing the first research question of this paper about whether accessing administrative data materialise into tangible academic progress for early-stage career researchers. In line with the simplest theoretical model described in Section 2, we have assumed that the parameters α, β, δ_P and δ_V were constant. However, it is plausible that α and β, i.e., the recruiter's valuation of the publication records and potential of a candidate, may vary with applicant characteristics (e.g. statistical discrimination). Similarly, applicants' abilities to convert access and use of administrative data into stronger signals (δ_P and δ_V) may differ across individuals. Consequently, it might be more appropriate to express the improvement in signal quality as $\alpha_r \delta_{P,r} + \beta_r \delta_{V,r}$ (i.e. Equation (14) in Section 2).

Given the limited set of sociodemographic characteristics available, we focus on two types of researchers using (inferred) gender and the rank of the department where the applicant completed their PhD.[5]

5.3.1. Differences across gender

A growing body of research highlights persistent gender disparities in academic hiring and promotion, suggesting that the returns to research productivity may differ across men and women. In our framework, these differences can be captured through gender-specific variation in recruiter perceptions of research signals (i.e. $\alpha_{female} \neq \alpha_{male}$ and/or $\beta_{female} \neq \beta_{male}$). Prior work shows that women in academia face higher performance thresholds for career progression, even when controlling for research productivity (Ginther and Kahn, 2004, 2014). This implies that recruiters may place lower weight on women's publication records and perceived potential, leading to systematically lower α and β for women. If administrative data access enhances research productivity but fails to close this gendered evaluation gap, we expect women to experience

[5] We inferred the gender of applicants using a two-step procedure. First, we checked whether applicants explicitly stated their pronouns on online platforms (e.g., X) and assigned gender accordingly. Notably, none of the applicants in our sample reported pronouns suggesting non-binary identification. For applicants without self-reported pronouns, we used generative AI to infer gender based on their name and surname, adopting a binary classification framework.

weaker career benefits from administrative data than men. Similarly, given the importance of networks in hiring and promotion, and the documented underrepresentation of women in elite research collaborations (Ductor *et al.*, 2021), it is possible that recruiters perceive the research signals of female candidates as less credible or impactful, further dampening their hiring and promotion prospects relative to equally productive male counterparts.

Beyond recruiter perceptions, gender differences may also arise from variation in applicants' abilities to leverage administrative data effectively (i.e. $\delta_{P,female} < \delta_{P,male}$ and/or $\delta_{V,female} < \delta_{V,female}$). Several factors could shape these differences. First, women in economics are underrepresented in empirical, policy-oriented fields such as labor economics, where administrative data are particularly relevant (Chari and Goldsmith-Pinkham, 2017). If prior exposure to empirical work influences the ability to capitalise on data access, we might expect women to have a lower δ_P and δ_V. Second, given the importance of research networks in knowledge transmission, and evidence that women are less likely to collaborate with top-ranked economists (Ductor *et al.*, 2021), differences in access to mentorship and co-authorship opportunities could limit their ability to convert data access into high-impact research. Finally, time constraints and work-life balance considerations – previously shown to contribute to gender disparities in academic productivity (Antecol *et al.*, 2018) – could further constrain women's ability to invest in the early-stage learning required to maximise the benefits of administrative data. If administrative data access requires substantial upfront investment in terms of data management and empirical skills, these constraints could widen the gender gap in the ability to translate data access into career gains. Taken together, these mechanisms suggest that the effectiveness of administrative data as a career-enhancing tool may be gendered, reinforcing rather than reducing pre-existing disparities.

In line with Chari and Goldsmith-Pinkham (2017), columns (1) and (2) of Table 2 show that more men than women apply for access to INPS administrative data. However, column (3) of the same Table indicates that, conditional on applying, there is no evidence of a female penalty in the probability of being granted access. The first three columns of Online Appendix Table A4 complement this analysis by reporting mean differences in various academic outcomes between female and male applicants one year before applying to the VisitINPS program. We observe only one significant difference: male applicants are more likely to come from a high-ranked economics department. For all other pre-application academic outcomes, there is no systematic female disadvantage.

In the first two columns of Table 5, we re-estimate the effects of administrative data access through the VisitINPS program on all outcomes from Table 3, separately for female and male applicants. Our findings indicate that administrative data access increases the likelihood of producing manuscripts based on INPS data, attending conferences, and publishing in high-ranked peer-reviewed journals for both groups. However, these effects are consistently larger for women, with most differences being statistically significant. Female applicants may, if anything, have a slightly higher ability to convert administrative data access into improved publication records and stronger research potential (i.e. $\delta_{P,female} > \delta_{P,male}$ and/or $\delta_{V,female} > \delta_{V,female}$). One might expect this advantage to translate into higher career advancement probabilities for women. However, the estimated treatment effect on career progression for female applicants (0.128) is not statistically different from that for male applicants (0.094).

Under the assumption that our set of academic outcomes provides a reasonably accurate representation of an applicant's research profile, the fact that hiring probabilities increase equally for men and women – despite women achieving greater research improvements – is consistent with the presence of a double standard in the hiring process. That is, female applicants may need to produce stronger research output to achieve the same hiring outcomes as their male counterparts, implying that recruiters place lower weight on their research records and perceived potential (i.e. $\alpha_{female} < \alpha_{male}$ and/or $\beta_{female} < \beta_{male}$).

5.3.2. Differences across PhD-programs

Scholars trained at top-ranked PhD programs are expected to be positively selected and to benefit from superior mentorship, stronger research networks, and greater visibility in the job market. If these initial advantages compound over time, administrative data access may serve as yet another mechanism that amplifies pre-existing disparities, consistent with the Matthew Effect (Merton, 1968). Conversely, if administrative data access provides a research advantage that is particularly valuable to those with fewer initial resources, it could help mitigate these disparities by allowing scholars from lower-ranked PhD programs to compete more effectively with their peers from elite institutions. The empirical analysis that follows tests which of these mechanisms dominates.

First, we define 'top PhD programs' as those awarded by economics departments ranked in the Top-20 in RePEc rankings. The last three columns of Online Appendix Table A4 document differences between applicants from top and non-top PhD programs one year prior to their application to the VisitINPS program. The first notable difference is that researchers trained at top-ranked economics departments are significantly more likely to apply for administrative data access while still completing their PhD. This is shown in column (5), where the proportion of applicants already holding a postdoctoral position one year before applying is higher among those from non-top PhD programs (second row). Hence, applicants from top PhD programs apply earlier in their careers, potentially benefiting from institutional encouragement to engage with administrative data at an earlier stage.

Despite being less experienced in terms of career stage, applicants from top PhD programs exhibit a higher probability of working with a highly ranked economist and having recently published in a Top-30 economics journal. This is consistent with a stronger research environment and greater exposure to top-tier networks typically associated with top PhD programs. In contrast, applicants from non-top PhD programs, who have often spent more time in postdoctoral positions, are more likely to have published in any peer-reviewed journal, suggesting a focus on accumulating publications as part of their academic progression. These differences highlight a key distinction: while researchers from top PhD programs tend to place their work in higher-impact outlets, those from non-top PhD programs appear to prioritise building a broader publication record before applying for administrative data access. Second, while the difference is not statistically significant at conventional levels due to our limited sample size, we observe that applicants from top PhD programs are more likely to be granted administrative data access than those from non-top PhD programs (30% versus 20%). This advantage in securing access to administrative data may stem stronger proposal-writing skills or better institutional support.

The last two columns of Table 5 examine the effects of administrative data access on academic outcomes and career progression separately for researchers from top and non-top PhD programs. The former exhibit stronger treatment effects in the production of manuscripts based on INPS data and are also the only group we identified so far to experience a significant increase in participation at EALE. They also show a slightly higher probability of attending AIEL. These patterns suggest that researchers from top PhD programs are particularly effective at leveraging administrative data to strengthen their academic presence.

The findings on publication records, however, present a more nuanced picture. While administrative data access does not significantly increase the probability of publishing in any peer-reviewed journal, it raises the likelihood of publishing in Top-30 economics journals and top-field labour economics journals for researchers from non-top PhD programs. Notably, the probability of publishing in a top-field labour economics journal also increases for researchers from top PhD programs, but the effect is not statistically significant, likely due to the small sample size.

Ultimately, both groups benefit from administrative data access in terms of career advancement. However, the effect is slightly larger for researchers from top PhD programs, reinforcing the idea that they may be better positioned to capitalise on administrative data. Moreover, while we showed in Table 3 that administrative data access reduces the likelihood of securing a position in a top world-ranked economics department while increasing the probability of employment in a top Italian economics department on average, Table 5 reveals that this pattern is entirely driven by researchers from top PhD programs.

Whether these results are driven by a greater ability to translate data access into stronger research output or by recruiter preferences favouring researchers from top PhD programs is difficult to disentangle. However, the overall pattern suggests that administrative data access plays an active role in academic mobility. While it provides valuable research advantages to all applicants, the extent to which it translates into broader career gains may depend on existing institutional affiliations. In particular, with the exception of increased Top-30 journal publications among non-top PhD applicants, the results are largely consistent with the Matthew Effect, whereby those already positioned at the top extract greater benefits from research resources.[6]

6. Conclusion

This paper examines the impact of access to administrative data on the academic trajectories of early-career economists, focusing on conference participation, publication outcomes, and career advancement. We first developed a simple theoretical model that frames administrative data access as a signal of potential and publication quality, guiding our empirical analysis of its impact on hiring and promotion. Using longitudinal academic biographies and employing a TWFE approach, supplemented by RDDs in robustness checks to address selection issues, we find that access to INPS administrative data – granted through the VisitINPS program – improves early markers of research productivity—such as working papers and conference participation—and increases the likelihood of publication in high-ranking journals over time. Importantly, we observe that career advancement often precedes high-quality publications, suggesting that data access strengthens perceived research potential in hiring and promotion decisions.

Our findings also show that administrative data access influences geographic mobility: researchers are more likely to secure positions in top Italian economics departments and less likely to be hired at globally ranked institutions. This reflects both supply-side preferences for domestic careers—

[6] One potential concern is that the differences we observe between researchers from top and non-top PhD programs may not stem from their institutional background per se, but rather from differences in career stage at the time of application. The top PhD group contains a larger proportion of younger researchers, whereas the non-top PhD group has a higher share of postdoctoral fellows. If administrative data access is more beneficial when acquired earlier in one's career, what we may be capturing is a form of dynamic complementarity rather than an effect of PhD program prestige. To address this, in Online Appendix Table A5, we re-estimate the effects of administrative data access separately for applicants who were PhD students and those who were postdocs at the time of application. We do not find systematically larger effects for PhD students, except for a higher probability of working in a top Italian economics department after accessing administrative data. However, this effect disappears once we exclude applicants from top PhD programs, suggesting that the observed differences are indeed linked to institutional background rather than early-career access alone.

many VisitINPS applicants are Italian nationals who may prefer to remain within the domestic academic market—and demand-side factors—Italian institutions may favor researchers with strong research potential and long-term retention prospects. This effect is further reinforced by Italy's Brain Gain programs, which provide tax incentives for Italian researchers returning from abroad, making domestic career even more attractive.

Crucially, our results reveal that the benefits of administrative data access are not equitably distributed. Scholars from top-ranked PhD programs derive greater visibility and recognition, amplifying existing institutional hierarchies rather than leveling the playing field. Additionally, our findings reveal persistent gender disparities: although women demonstrate stronger research improvements from administrative data access, they do not experience comparable career gains. This pattern suggests the presence of a double standard in academic evaluation, where women face higher performance thresholds to achieve similar career outcomes as men.

These findings highlight the interplay between resource access and structural inequalities in academia, raising important questions about whether administrative data democratises research opportunities or primarily amplifies existing advantages. While administrative data access undoubtedly enhances academic trajectories, its full potential to foster equity in research careers remains contingent on broader institutional and market dynamics.

Our study carries broader implications for the academic community. As administrative data grow in importance for high-impact empirical research and scholarly success, the rules governing access take on critical importance. If access is disproportionately available to those already in elite positions—or if only some scholars can effectively leverage it—then administrative data risk becoming yet another mechanism through which advantage accumulates. While our analysis demonstrates that administrative data access enhances research output and career progression, the extent to which it translates into broader academic mobility remains an open question.

Our findings suggest that realising the equalising potential of administrative data requires intentional policy design. Public data access programs must be structured to broaden—not narrow—opportunities, particularly for underrepresented researchers and institutions. Otherwise, the growing centrality of administrative data may challenge the ideals of meritocracy and equal opportunity in research.

References:

Ajzenman, N., Ferman, B., & Sant'Anna, P. (2023). Discrimination in the Formation of Academic Networks: A Field Experiment on #EconTwitter. *IZA Discussion Paper Series*, No.15878.

Anderson, K. A., & Richards-Shubik, S. (2022). Collaborative production in science: An empirical analysis of coauthorships in economics. *Review of Economics and Statistics*, *104*, 1241-1255.

Angrist, J., P. Azoulay, G. Ellison, R. Hill, & S.F. Lu (2020): Inside job or deep impact? Extramural citations and the influence of economic scholarship. *Journal of Economic Literature*, 58, 3–52.

Antecol, H., Bedard, K., & Stearns, J. (2018). Equal but inequitable: Who benefits from gender-neutral tenure clock stopping policies?. *American Economic Review*, *108*, 2420-2441.

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*, 959-975.

Azoulay, P., Graff Zivin, J. S., & Wang, J. (2010). Superstar extinction. *Quarterly Journal of Economics*, *125*, 549-589.

Azoulay, P., Fons-Rosen, C., & Zivin, J. S. G. (2019). Does science advance one funeral at a time?. *American Economic Review*, *109*, 2889-2920.

Bassetto, J., & Ippedico, G. (2024). Tax Incentives and Return Migration. *IZA Discussion Paper Series*, No. 17224.

Bosquet, C., Combes, P. P., & García-Peñalosa, C. (2019). Gender and promotions: Evidence from academic economists in France. *The Scandinavian Journal of Economics*, 121, 1020-1053.

Bransch, F., & Kvasnicka, M. (2022). Male gatekeepers: gender bias in the publishing process?. *Journal of Economic Behavior & Organization*, *202*, 714-732.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics, 225*, 200-230.

Card, D., Chetty, R., Feldstein, M. S., & Saez, E. (2010). Expanding access to administrative data for research in the United States. *American economic association, ten years and beyond: Economists answer NSF's call for long-term research agendas*.

Carrell, S., Figlio, D., & Lusher, L. (2024). Clubs and networks in economics reviewing. *Journal of Political Economy*, *132*, 2999-3024.

Chari, A., & Goldsmith-Pinkham, P. S. (2017). Gender representation in economics across topics and time: evidence from the NBER. *NBER Working Paper Series*, No. 23953.

Cole, S., I. Dhaliwal, A. Sautmann, & L. Vilhuber (2020). *Handbook on Using Administrative Data for Research and Evidence-Based Policy*, JPAL and MIT Press.

Colussi, T. (2018). Social ties in academia: A friend is a treasure. *Review of Economics and Statistics*, *100*, 45-50.

Connelly, R., Playford, C. J., Gayle, V., & Dibben, C. (2016). The role of administrative data in the big data revolution in social science research. *Social Science Research*, *59*, 1-12.

Currie, J., Kleven, H., & Zwiers, E. (2020). Technology and big data are changing economics: Mining text to track methods. *American Economic Association Papers and Proceedings*, 110, 42-48.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review, 110*, 2964-2996.

Ductor, L., Goyal, S., & Prummer, A. (2023). Gender and collaboration. *Review of Economics and Statistics*, *105*, 1366-1378.

Ductor, L., & Visser, B. (2022). When a coauthor joins an editorial board. *Journal of Economic Behavior & Organization*, *200*, 576-595.

Eberhardt, M., Facchini, G., & Rueda, V. (2023). Gender differences in reference letters: Evidence from the economics job market. *The Economic Journal*, 133, 2676-2708.

Einav, L., & Levin, J. (2014). Economics in the age of big data. *Science*, 346, 1243089.

Figlio, D., Karbownik, K., & Salvanes, K. (2017). The promise of administrative data in education research. *Education Finance and Policy*, 12, 129-136.

Ginther, D. K., & Kahn, S. (2004). Women in economics: moving up or falling off the academic career ladder?. *Journal of Economic Perspectives*, *18*, 193-214.

Ginther, D. K., & Kahn, S. (2014). Academic women's careers in the social sciences. In *The Economics of Economists: Institutional Setting, Individual Incentives, and Future Prospects* (pp. 285–315). Cambridge University Press.

Gualavisi, M., Kleemans, M., & Thornton, R. (2024). Moving in Academia: Who Moves and What Happens After?. *American Economic Association Papers and Proceedings, 114*, 232-237.

Hale, G., Regev, T., & Rubinstein, Y. (2023). Do looks matter for an academic career in economics?. *Journal of Economic Behavior & Organization*, 215, 406-420.

Heckman, J. J., & Moktan, S. (2020). Publishing and promotion in economics: The tyranny of the top five. *Journal of Economic Literature*, 58, 419-470.

Jacob, B. A., & Lefgren, L. (2011). The impact of research grant funding on scientific productivity. *Journal of public economics*, 95(9-10), 1168-1177.

Jones, T. & Sloan, A. (2024). The Academic Origins of Economics Faculty. *The Journal of Economic Education, 55*, 434–454.

Li, W., Aste, T., Caccioli, F., & Livan, G. (2019). Early coauthorship with top scientists predicts success in academic careers. *Nature Communications*, 10, 5170.

Lundberg, S., & Stearns, J. (2019). Women in economics: Stalled progress. *Journal of Economic Perspectives*, 33, 3-22.

Nagaraj, A., E. Shears, & M. de Vaan (2020). Improving data access democratizes and diversifies science. *Proceedings of the National Academy of Sciences*, 117, 23490–23498.

Nagaraj, A., & Tranchero, M. (2023). How does data access shape science? Evidence from the impact of US census's research data centers on economics research. *NBER Working Paper Series*, No. 31372.

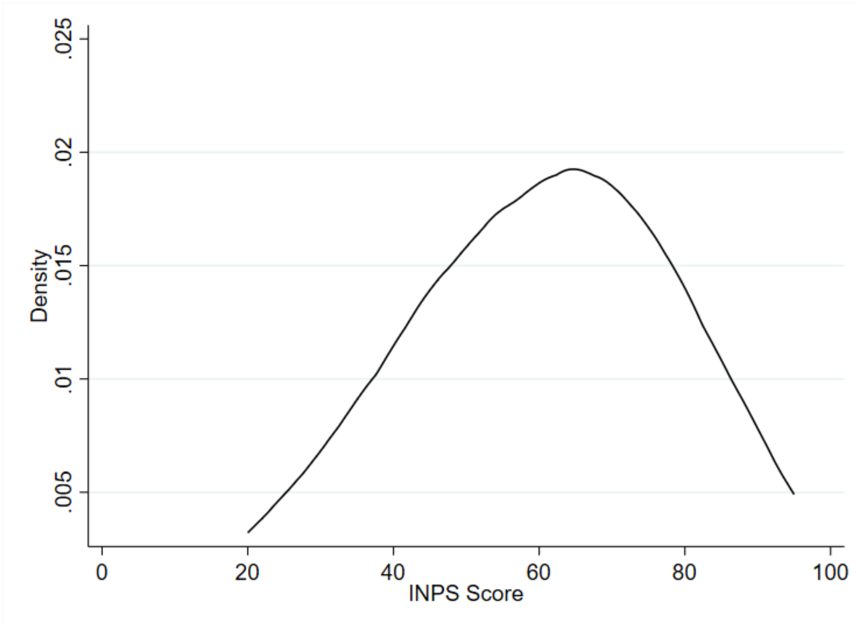
Tabakovic, H., & Wollmann, T. G. (2019). The impact of money on science: Evidence from unexpected NCAA football outcomes. *Journal of Public Economics*, 178, 104066.

Waldinger, F. (2010). Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany. *Journal of Political Economy*, 118, 787-831.

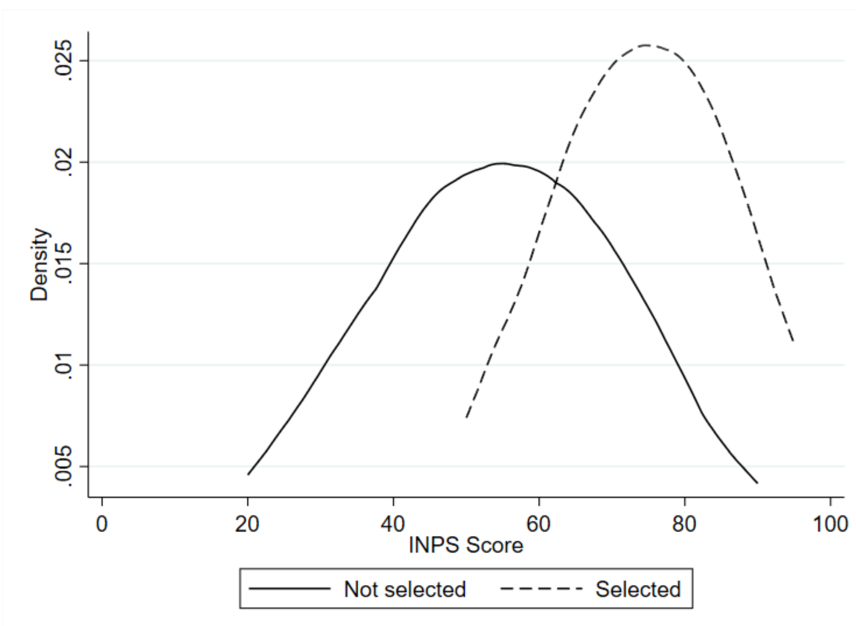
Tables and Figures:

Figure 1: Distributions of scores assigned to proposals

Panel A: All applicants



Panel B: Selected vs not selected applicants



Note: These figures refer to the 203 applicants in our estimation sample.

Table 1: Descriptive statistics – All sample and over time

	All sample (1)	Prior Application (2)	Post Application (3)	Difference (4)
Applicants' characteristics:				
Female	0.414	0.414	0.414	0.000
	[0.493]	[0.493]	[0.493]	(0.022)
Postdoc	0.209	0.104	0.356	0.251***
	[0.407]	[0.306]	[0.479]	(0.017)
Academic outcomes:				
At least one VisitINPS WP	0.018	0.001	0.043	0.042***
	[0.134]	[0.029]	[0.202]	(0.006)
At least one manuscript with INPS data	0.036	0.002	0.084	0.082***
	[0.186]	[0.041]	[0.278]	(0.008)
Presence at EALE conference	0.037	0.018	0.063	0.045***
	[0.189]	[0.133]	[0.244]	(0.008)
Presence at AIEL conference	0.064	0.017	0.129	0.112***
	[0.244]	[0.130]	[0.335]	(0.011)
At least one peer-reviewed publication	0.224	0.109	0.384	0.275***
	[0.417]	[0.312]	[0.487]	(0.018)
At least one peer-reviewed publication in Top-30	0.020	0.003	0.044	0.040***
Economics journals	[0.140]	[0.057]	[0.205]	(0.006)
At least one peer-reviewed publication in a high-ranked	0.006	0.000	0.015	0.015***
Labour economics journal	[0.079]	[0.000]	[0.121]	(0.003)
Yearly citation count	11.568	2.858	23.777	20.919***
	[31.011]	[10.724]	[43.523]	(1.299)
Work with a high-ranked economist	0.023	0.011	0.040	0.029***
	[0.151]	[0.107]	[0.197]	(0.007)
Work with a high-ranked Italian economist	0.045	0.022	0.076	0.054***
	[0.206]	[0.147]	[0.265]	(0.009)
Recent academic promotion	0.092	0.031	0.176	0.145***
	[0.288]	[0.174]	[0.381]	(0.012)
Work in a high-ranked Economic department	0.215	0.235	0.186	-0.048***
	[0.411]	[0.424]	[0.390]	(0.018)
Work in a high-ranked Italian Economic department	0.130	0.158	0.091	-0.067***
	[0.336]	[0.365]	[0.288]	(0.015)
Observations	2087	1218	869	
Individuals	203	140	63	

Notes: Standard deviations are in square brackets and standard errors are in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively.

Table 2: Balance test – A year prior submitting a VisitINPS proposal

	Not selected (1)	Selected (2)	Difference (3)
Applicants' characteristics:			
Female	0.421	0.397	-0.025
	[0.496]	[0.493]	(0.075)
Postdoc	0.207	0.238	0.031
	[0.407]	[0.429]	(0.063)
Academic output and achievements:			
At least one VisitINPS WP	0.000	0.000	0.000
	[0.000]	[0.000]	(0.000)
At least one manuscript with INPS data	0.000	0.016	0.016
	[0.000]	[0.126]	(0.011)
Presence at EALE conference	0.036	0.063	0.028
	[0.186]	[0.246]	(0.031)
Presence at AIEL conference	0.029	0.063	0.035
	[0.167]	[0.246]	(0.030)
At least one peer-reviewed publication	0.200	0.190	-0.010
	[0.401]	[0.396]	(0.061)
At least one peer-reviewed publication in Top-30	0.007	0.016	0.009
Economics journals	[0.085]	[0.126]	(0.015)
At least one peer-reviewed publication in a high-ranked	0.000	0.000	0.000
Labour economics journal	[0.000]	[0.000]	(0.000)
Yearly citation count	5.521	12.905	7.383***
	[13.273]	[28.657]	(2.938)
Work with a high-ranked economist	0.036	0.063	0.028
	[0.186]	[0.246]	(0.031)
Work with a high-ranked Italian economist	0.007	0.032	0.025
	[0.085]	[0.177]	(0.018)
Recent academic promotion	0.050	0.063	0.013
	[0.219]	[0.246]	(0.035)
Work in a high-ranked Economic department	0.214	0.286	0.071
	[0.412]	[0.455]	(0.065)
Work in a high-ranked Italian Economic department	0.057	0.111	0.054
	[0.233]	[0.317]	(0.040)
Individuals	140	63	

Notes: Standard deviations are in square brackets and standard errors are in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively.

Table 3: Effect of administrative data access on academic outcomes – Two-Way Fixed-Effects approach

	TWFE (1)	Pre-treatment p-value (2)	First year significant (3)	Last year significant (4)
Manuscript with INPS data				
At least one VisitINPS WP	0.188***	0.291	2	6
	(0.030)			
At least one manuscript with INPS data	0.180***	0.438	2	6
	(0.029)			
Presence at Labour economics conference				
Presence at EALE conference	0.018	0.160	NA	NA
	(0.048)			
Presence at AIEL conference	0.175***	0.792	1	6
	(0.041)			
Publication records				
At least one peer-reviewed publication	0.053	0.672	NA	NA
	(0.052)			
At least one peer-reviewed publication in Top-30 Economics journals	0.023	0.951	NA	NA
	(0.030)			
At least one peer-reviewed publication in a high-ranked Labour economics journal	0.063***	0.333	3	6
	(0.023)			
Academic impact and network				
Yearly citation count	9.318**	0.172	3	6
	(4.143)			
Work with a high-ranked economist	0.002	0.457	NA	NA
	(0.032)			
Work with a high-ranked Italian economist	0.018	0.397	NA	NA
	(0.033)			
Career evolution				
Recent academic promotion	0.106***	0.855	2	4
	(0.023)			
Work in a high-ranked Economic department	-0.089*	0.207	4	6
	(0.046)			
Work in a high-ranked Italian Economic department	0.099**	0.147	5	6
	(0.050)			

Notes: Coefficients show the effect of accessing administrative data via the VisitINPS program. The sample size is equal to 2087 (N=203). Outcomes are measured yearly. Standard errors in parentheses are clustered at the individual level. All regressions include individual and time fixed-effects. The TWFE estimates in column (1) comes from the Callaway and Sant'Anna (2021) estimator. Column (2) reports the p-value attracted by the average outcome for treated units in the pre-treatment periods. Column (3) and (4) respectively indicate, when applicable, the first and last year when the effect of the treatment is statistically significantly different from zero. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table 4: Effect of administrative data access on academic outcomes – Robustness Checks

	Full Sample (1)	First application only (2)	Between the lowest and highest acceptance threshold (3)	Up to 12.5 points below or above acceptance threshold (4)	Up to 10 points below or above acceptance threshold (5)	Up to 7.5 points below or above acceptance threshold (6)
Manuscript with INPS data						
At least one VisitINPS WP	0.188***	0.183***	0.238***	0.215***	0.223***	0.224***
	(0.030)	(0.031)	(0.035)	(0.037)	(0.038)	(0.040)
At least one manuscript with INPS data	0.180***	0.185***	0.178***	0.151***	0.183***	0.198***
	(0.029)	(0.029)	(0.030)	(0.032)	(0.035)	(0.040)
Presence at Labour economics conference						
Presence at EALE conference	0.018	-0.005	-0.004	0.002	0.017	0.001
	(0.048)	(0.051)	(0.050)	(0.057)	(0.056)	(0.062)
Presence at AIEL conference	0.175***	0.175***	0.175***	0.170***	0.155***	0.176***
	(0.041)	(0.043)	(0.042)	(0.048)	(0.048)	(0.052)
Publication records						
At least one peer-reviewed publication	0.053	0.059	0.054	0.064	0.093	0.062
	(0.052)	(0.056)	(0.053)	(0.058)	(0.058)	(0.062)
At least one peer-reviewed publication	0.023	0.015	0.024	0.056**	0.066**	0.055
in Top-30 Economics journal	(0.030)	(0.031)	(0.033)	(0.025)	(0.028)	(0.036)
At least one peer-reviewed publication in a	0.063***	0.052**	0.058**	0.059**	0.083**	0.075*
high-ranked Labour economics journal	(0.023)	(0.023)	(0.024)	(0.025)	(0.038)	(0.042)
Academic impact and network						
Yearly citation count	9.318**	8.609**	8.640*	7.907	7.394	6.595
	(4.143)	(4.019)	(4.577)	(5.100)	(5.500)	(6.229)
Work with a high-ranked economist	0.018	-0.003	0.019	0.032	0.038	0.043
	(0.033)	(0.026)	(0.034)	(0.028)	(0.030)	(0.037)
Work with a high-ranked Italian economist	0.002	0.007	-0.015	-0.006	-0.004	-0.023
	(0.032)	(0.034)	(0.035)	0(.032)	(0.035)	(0.035)
Career evolution						
Recent academic promotion	0.106***	0.102***	0.109***	0.099***	0.083***	0.080**
	(0.023)	(0.025)	(0.026)	(0.029)	(0.030)	(0.033)
Work in a high-ranked Economic department	-0.089*	-0.085*	-0.101**	-0.061	-0.052	-0.025
	(0.046)	(0.047)	(0.052)	(0.054)	(0.057)	(0.063)
Work in a high-ranked Italian Economic	0.099**	0.087*	0.134**	0.116*	0.117*	0.138*
department	(0.050)	(0.053)	(0.058)	(0.064)	(0.065)	(0.073)
Observations	2087	1895	1378	1011	918	788

Notes: Coefficients show the effect of accessing administrative data via the VisitINPS program using the Callaway and Sant'Anna (2021) estimator. Outcomes are measured yearly. Standard errors in parentheses are clustered at the individual level. All regressions include individual and time fixed-effects. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

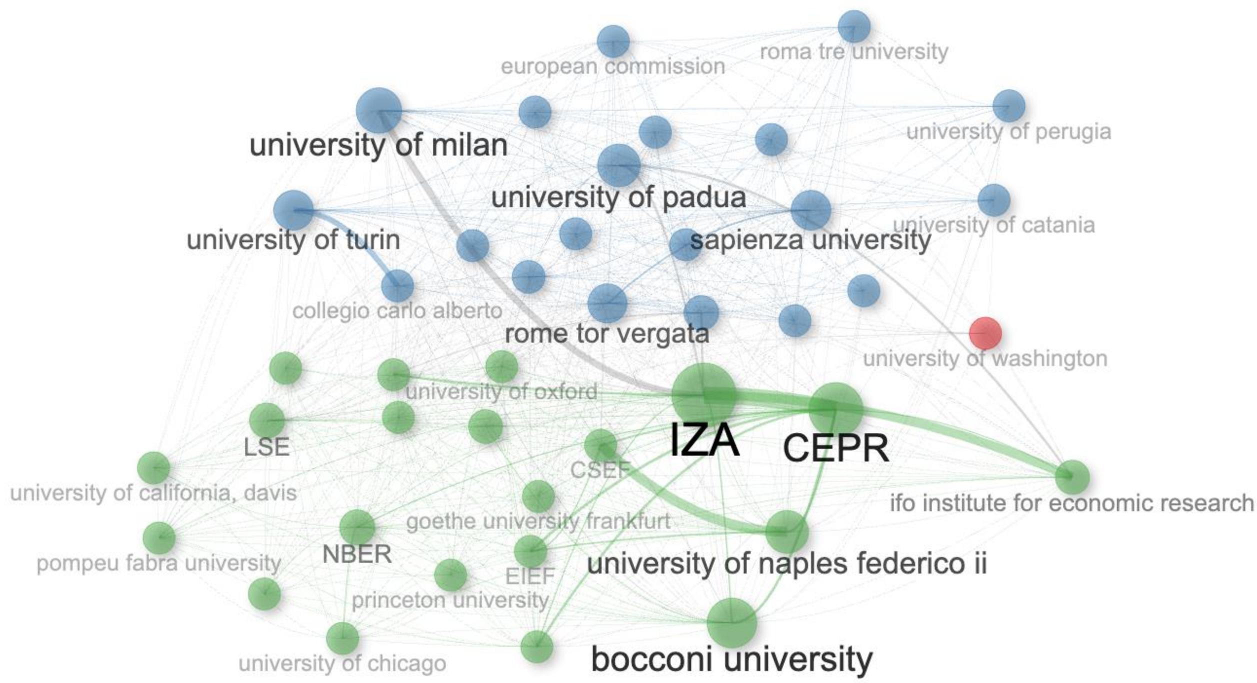
Table 5: Effect of administrative data access on academic outcomes – Two-Way Fixed-Effects Approach by characteristics of applicants

	Gender of applicant		From Top PhD program	
	Female	Male	Yes	No
	(1)	(2)	(3)	(4)
Manuscript with INPS data				
At least one VisitINPS WP	0.249***	0.242***	0.463***	0.170***
	(0.064)	(0.041)	(0.046)	(0.043)
At least one manuscript with INPS data	0.271***	0.183***	0.439***	0.152***
	(0.061)	(0.037)	(0.066)	(0.036)
Presence at Labour economics conference				
Presence at EALE conference	-0.052	0.056	0.181**	-0.004
	(0.125)	(0.046)	(0.080)	(0.067)
Presence at AIEL conference	0.252***	0.127***	0.223***	0.178***
	(0.070)	(0.048)	(0.071)	(0.053)
Publication records				
At least one peer-reviewed publication	0.018	0.074	0.040	0.052
	(0.085)	(0.065)	(0.096)	(0.064)
At least one peer-reviewed publication	0.063*	0.006	-0.014	0.037**
in Top-30 Economics journal	(0.036)	(0.053)	(0.103)	(0.018)
At least one peer-reviewed publication in a	0.214**	0.069**	0.113	0.070**
high-ranked Labour economics journal	(0.100)	(0.034)	(0.090)	(0.027)
Academic impact and network				
Yearly citation count	8.739*	9.402	9.658	8.673*
	(5.094)	(5.954)	(8.807)	(4.613)
Work with a high-ranked economist	0.024	-0.010	-0.090	0.024
	(0.057)	(0.040)	(0.083)	(0.033)
Work with a high-ranked Italian economist	0.037	0.010	0.032	0.001
	(0.036)	(0.051)	(0.119)	(0.004)
Career evolution				
Recent academic promotion	0.128***	0.094***	0.164***	0.090***
	(0.040)	(0.029)	(0.030)	(0.031)
Work in a high-ranked Economic department	0.021	-0.160**	-0.379***	0.018
	(0.066)	(0.062)	(0.122)	(0.039)
Work in a high-ranked Italian Economic	0.097	0.101	0.229*	0.092
department	(0.079)	(0.066)	(0.122)	(0.060)
	864	1223	485	1602

Notes: Coefficients show the effect of accessing administrative data via the VisitINPS program using the Callaway and Sant'Anna (2021) estimator. Outcomes are measured yearly. Standard errors in parentheses are clustered at the individual level. All regressions include individual and time fixed-effects. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Online Appendix:

Figure A1: Institutional affiliation and network of VisitINPS applicants

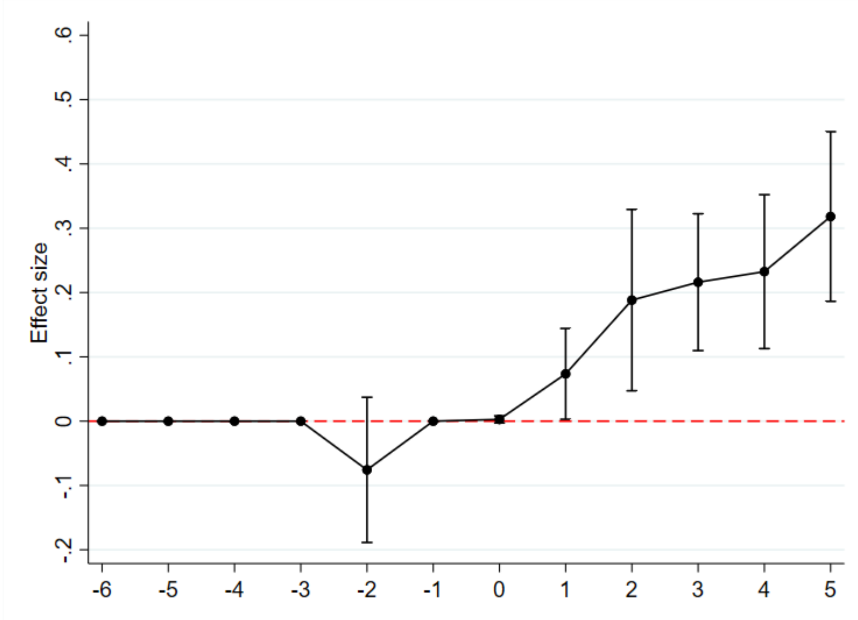


Notes: This map has been produced using Bibliometrix R package (Parameters: Top 50 institutions for number of articles associated to VisitINPS applicants; Community detection through "walktrap algorithm"). See Aria and Cuccurullo (2017) for technical details.

Figure A2: Word cloud – Manuscripts of early-stage career VisitINPS applicants

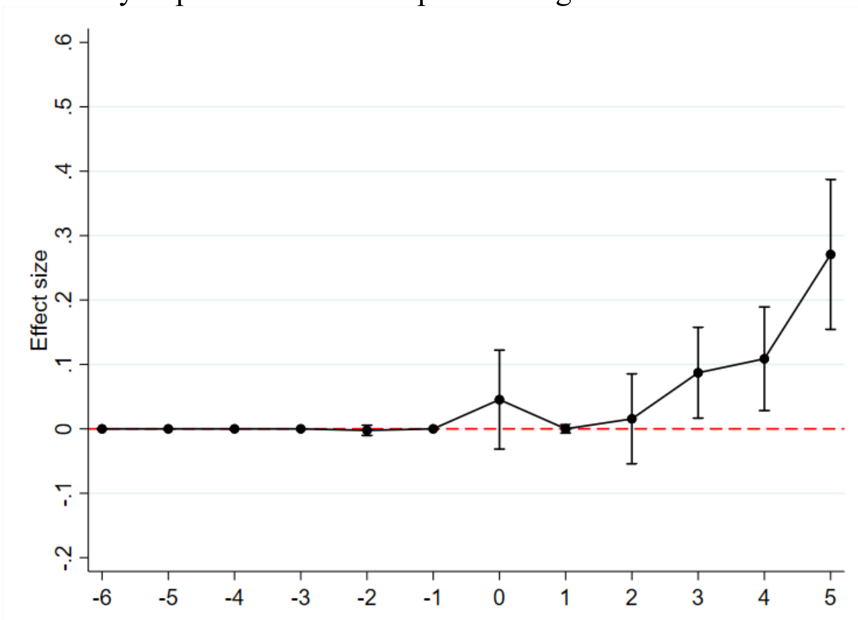
Note: This word cloud displays the most used keywords in the manuscripts published by the early-stage career VisitINPS applicants. This word cloud has been produced using Bibliometrix R package.

Figure A3: Probability to publish a working paper in VisitINPS WP Series – Event-study approach



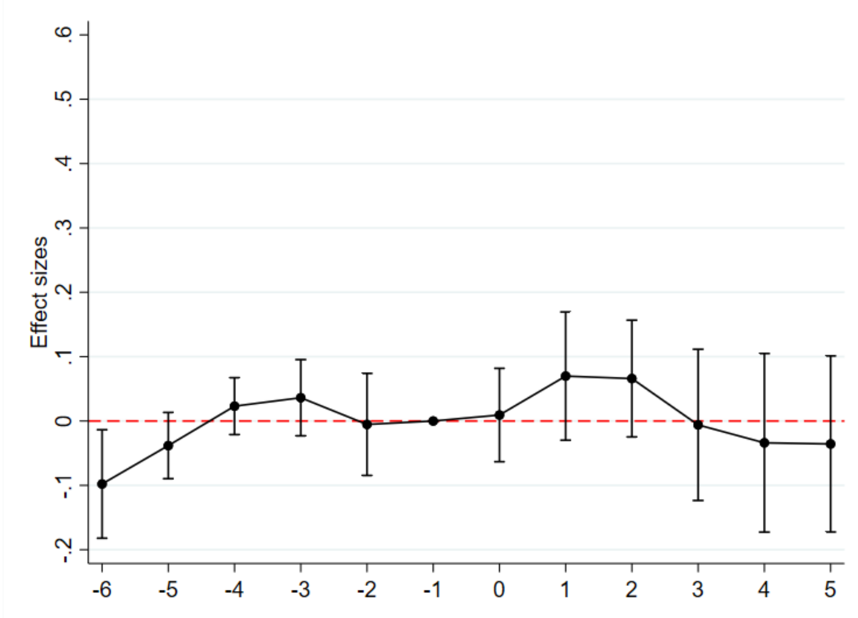
Notes: These are the event-study coefficients estimated using the Callaway and Sant'Anna (2021) estimator. Standard errors are clustered at the individual level and 90% confidence intervals are depicted.

Figure A4: Probability to publish a manuscript including INPS data – Event-study approach



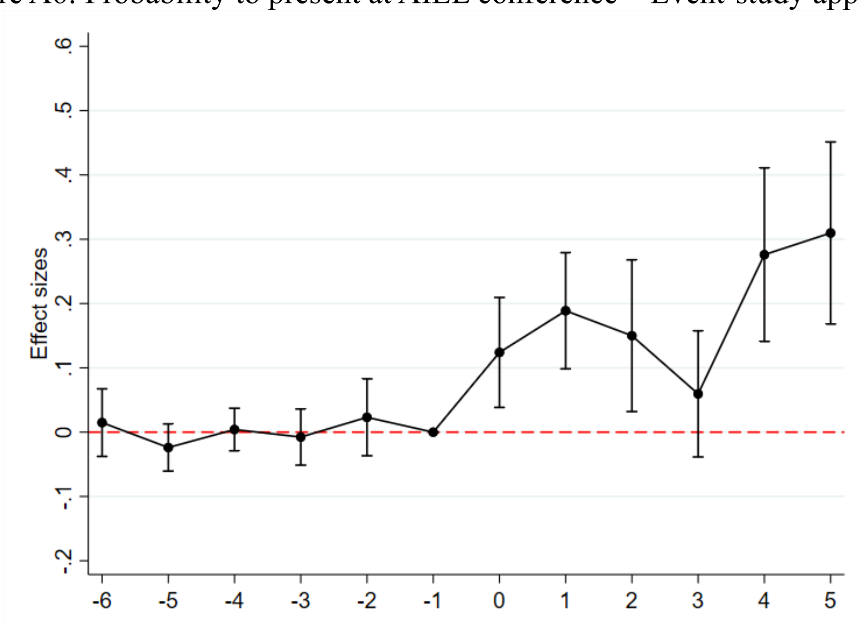
Notes: These are the event-study coefficients estimated using the Callaway and Sant'Anna (2021) estimator. Standard errors are clustered at the individual level and 90% confidence intervals are depicted.

Figure A5: Probability to present at EALE conference – Event-study approach



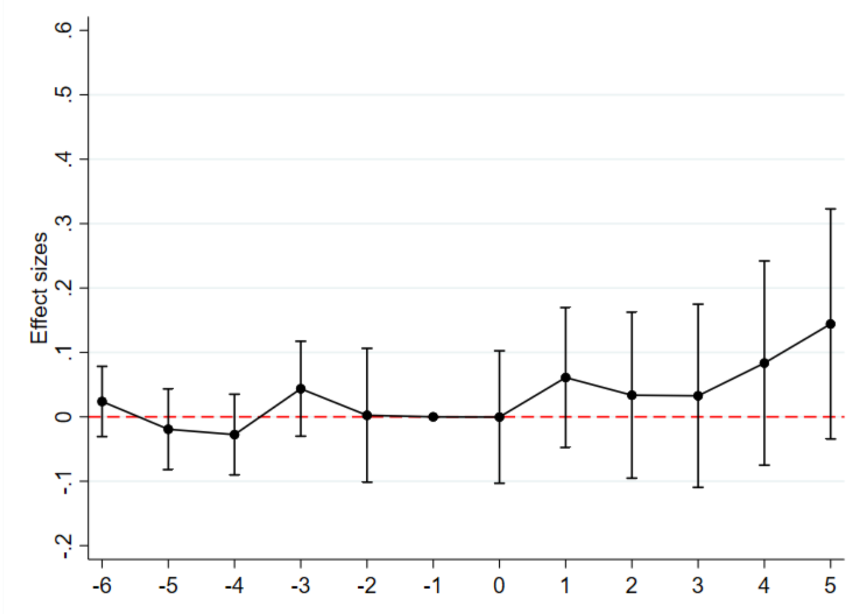
Notes: These are the event-study coefficients estimated using the Callaway and Sant'Anna (2021) estimator. Standard errors are clustered at the individual level and 90% confidence intervals are depicted.

Figure A6: Probability to present at AIEL conference – Event-study approach



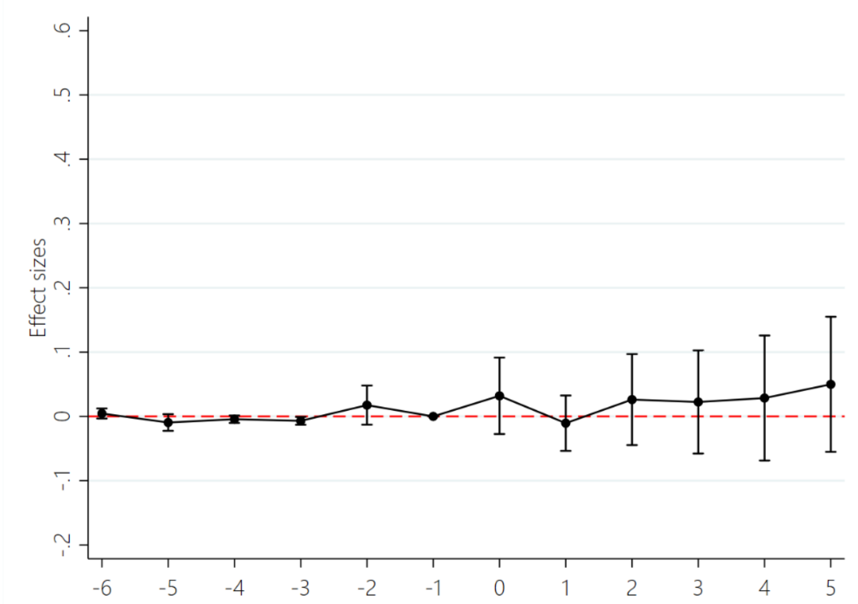
Notes: These are the event-study coefficients estimated using the Callaway and Sant'Anna (2021) estimator. Standard errors are clustered at the individual level and 90% confidence intervals are depicted.

Figure A7: Probability to publish at least one peer-reviewed article – Event-study approach



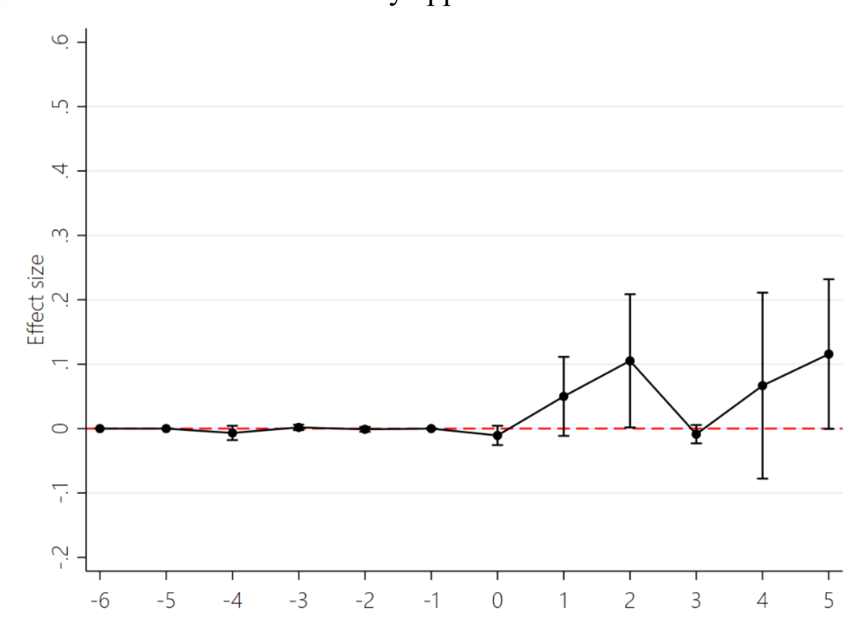
Notes: These are the event-study coefficients estimated using the Callaway and Sant'Anna (2021) estimator. Standard errors are clustered at the individual level and 90% confidence intervals are depicted.

Figure A8: Probability to publish an article in high-rank journals – Event-study approach



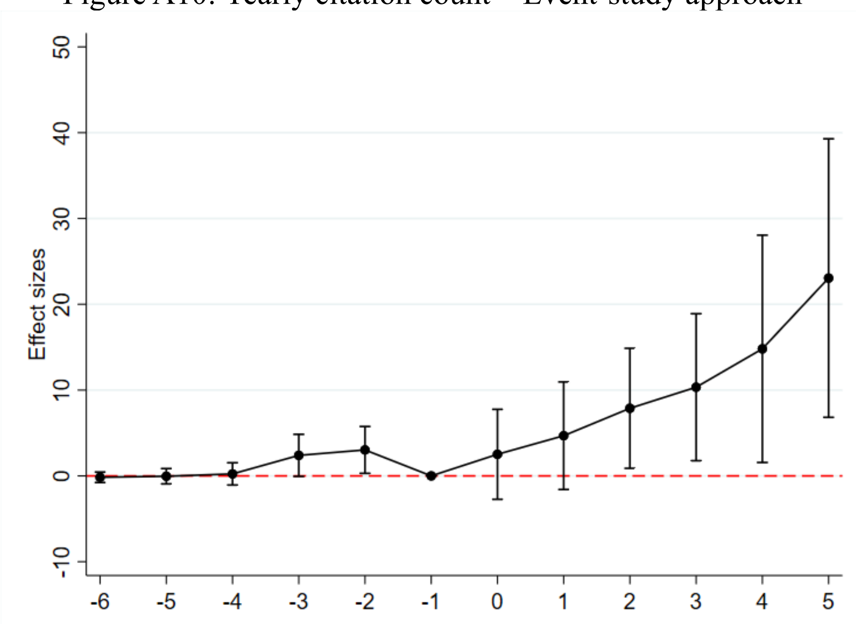
Notes: These are the event-study coefficients estimated using the Callaway and Sant'Anna (2021) estimator. Standard errors are clustered at the individual level and 90% confidence intervals are depicted.

Figure A9: Probability to publish an article in high-rank Labour economics journals – Event-study approach



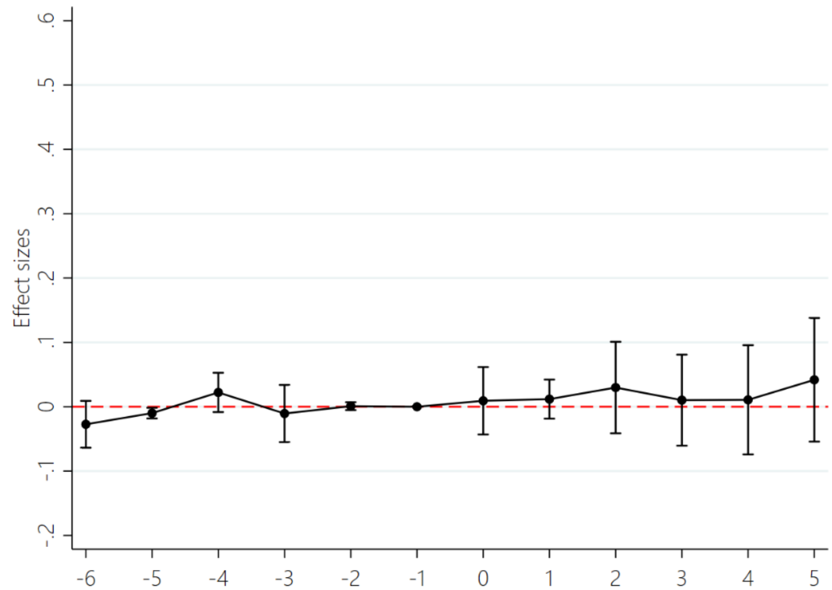
Notes: These are the event-study coefficients estimated using the Callaway and Sant'Anna (2021) estimator. Standard errors are clustered at the individual level and 90% confidence intervals are depicted.

Figure A10: Yearly citation count – Event-study approach



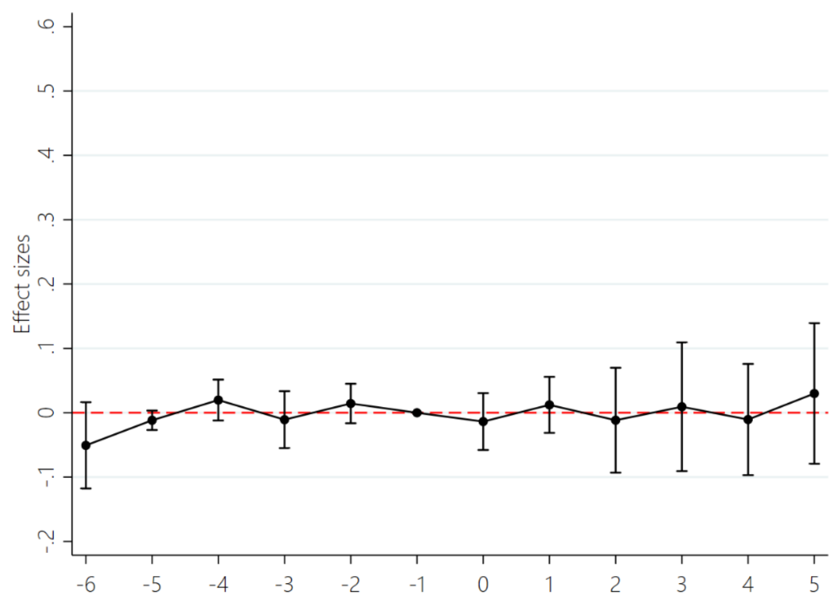
Notes: These are the event-study coefficients estimated using the Callaway and Sant'Anna (2021) estimator. Standard errors are clustered at the individual level and 90% confidence intervals are depicted.

Figure A11: Working with high-ranked economists – Event-study approach



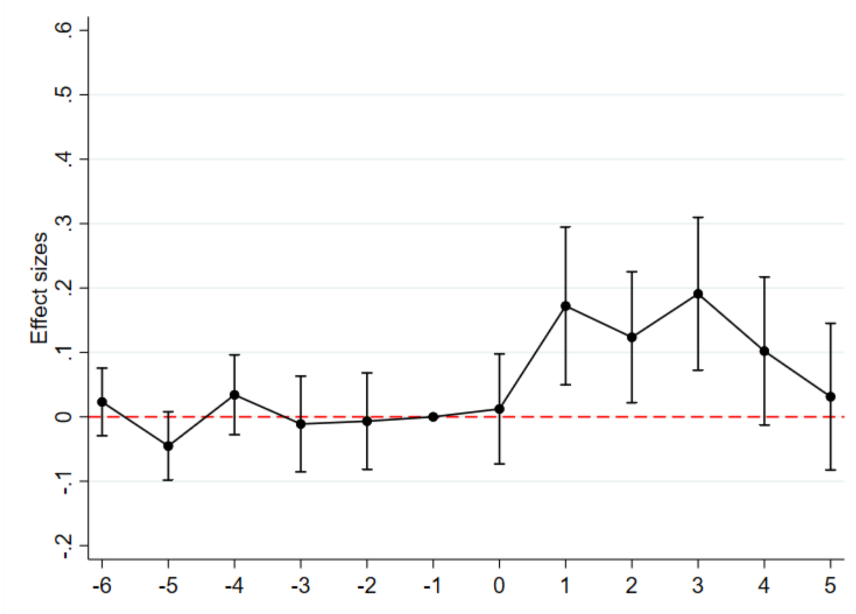
Notes: These are the event-study coefficients estimated using the Callaway and Sant'Anna (2021) estimator. Standard errors are clustered at the individual level and 90% confidence intervals are depicted.

Figure A12: Working with high-ranked Italian economists – Event-study approach



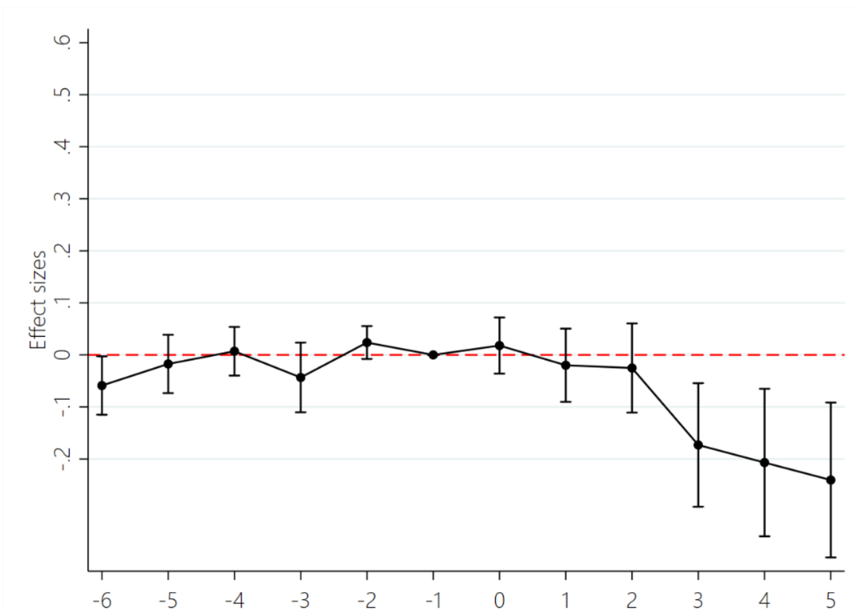
Notes: These are the event-study coefficients estimated using the Callaway and Sant'Anna (2021) estimator. Standard errors are clustered at the individual level and 90% confidence intervals are depicted.

Figure A13: Probability of recently receiving an academic promotion – Event-study approach



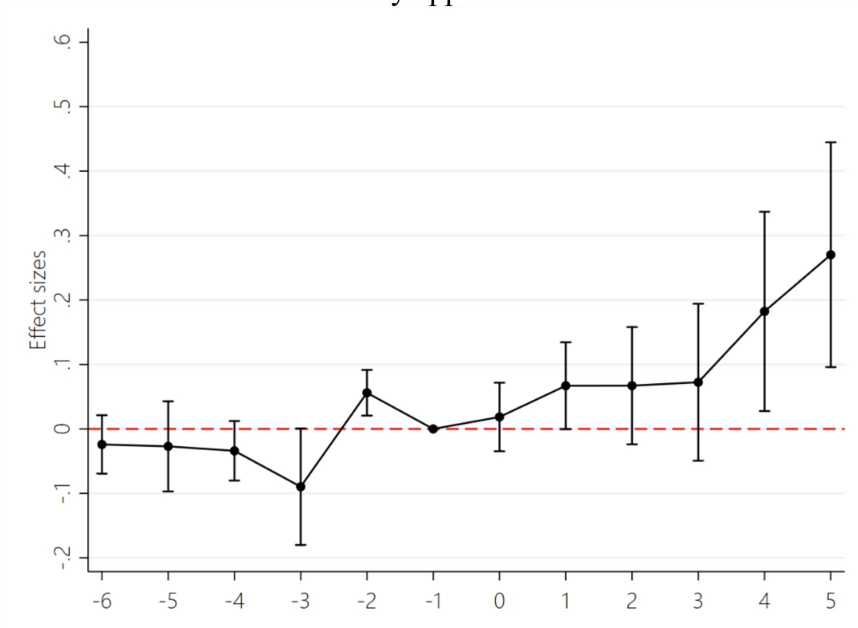
Notes: These are the event-study coefficients estimated using the Callaway and Sant'Anna (2021) estimator. Standard errors are clustered at the individual level and 90% confidence intervals are depicted.

Figure A14: Probability of working in a high-ranked Economics department – Event-study approach



Notes: These are the event-study coefficients estimated using the Callaway and Sant'Anna (2021) estimator. Standard errors are clustered at the individual level and 90% confidence intervals are depicted.

Figure A15: Probability of working in an Italian high-ranked Economics department – Event-study approach



Notes: These are the event-study coefficients estimated using the Callaway and Sant'Anna (2021) estimator. Standard errors are clustered at the individual level and 90% confidence intervals are depicted.

Table A1: Effect of administrative data access on academic outcomes – Alternative Two-Way Fixed-Effects approaches

	CS TWFE (1)	Naïve TWFE (2)	CD TWFE (3)
Manuscript with INPS data			
At least one VisitINPS WP	0.188***	0.105***	0.128***
	(0.030)	(0.018)	(0.017)
At least one manuscript with INPS data	0.180***	0.154***	0.171***
	(0.029)	(0.025)	(0.023)
Presence at Labour economics conference			
Presence at EALE conference	0.018	0.066***	0.026
	(0.048)	(0.024)	(0.045)
Presence at AIEL conference	0.175***	0.168***	0.172***
	(0.041)	(0.035)	(0.040)
Publication records			
At least one peer-reviewed publication	0.053	0.048	0.055
	(0.052)	(0.038)	(0.053)
At least one peer-reviewed publication in Top30	0.023	0.038*	0.031
Economics journals	(0.030)	(0.020)	(0.030)
At least one peer-reviewed publication in a high-ranked	0.063***	0.027***	0.033***
Labour economics journal	(0.023)	(0.010)	(0.011)
Academic impact and network			
Yearly citation count	9.318**	10.416**	9.678**
	(4.143)	(5.184)	(4.181)
Work with a high-ranked economist	0.018	0.034	0.041
	(0.033)	(0.028)	(0.037)
Work with a high-ranked Italian economist	0.002	0.003	0.002
	(0.032)	(0.022)	(0.031)
Career evolution			
Recent academic promotion	0.106***	0.072***	0.113***
	(0.023)	(0.024)	(0.023)
Work in a high-ranked Economic department	-0.089*	-0.086*	-0.100**
	(0.046)	(0.045)	(0.047)
Work in a highly-ranked Economic department	0.099**	0.025	0.101**
	(0.050)	(0.048)	(0.048)

Notes: Coefficients show the effect of accessing administrative data via the VisitINPS program. The sample size is equal to 2087 (N=203). Outcomes are measured yearly. Standard errors in parentheses are clustered at the individual level. All regressions include individual and time fixed-effects. The TWFE estimates in column (1) comes from the Callaway and Sant'Anna (2021) estimator. Estimates in column (2) come from a naïve TWFE model and those in column (3) from the de Chaisemartin and D'Haultfoeuille (2020) estimator. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table A2: Effect of administrative data access on publication records using the Italian official classification – Two-Way Fixed-Effects Approach

	All sample	Gender of applicant		Top PhD program	
		Female	Male	Yes	No
	(1)	(2)	(3)	(4)	(5)
About peer-reviewed publications:					
In a category-A journal	0.088*	0.096	0.083	0.077	0.096
	(0.053)	(0.079)	(0.071)	(0.109)	(0.063)
In an ‘Economics’ journal (sub-category A)	0.109**	0.113	0.106	0.063	0.128***
	(0.051)	(0.078)	(0.066)	(0.107)	(0.059)
In a ‘Business’ journal (sub-category B)	0.083	0.138**	0.052	0.058	0.092
	(0.054)	(0.070)	(0.074)	(0.111)	(0.062)
In an ‘Economic history’ journal (sub-category C)	0.113**	0.118	0.109*	0.063	0.133**
	(0.051)	(0.078)	(0.065)	(0.107)	(0.059)
In a ‘Statistics and mathematical methods for decision-making’ journal (sub-category D)	0.085	0.105	0.073	0.058	0.097
	(0.057)	(0.088)	(0.075)	(0.111)	(0.067)
Observations	2087	864	1233	485	1602

Notes: Coefficients show the effect of accessing administrative data via the VisitINPS program. Outcomes are measured yearly. The categories of journals refer to the official Italian classification (Area 13). All regressions include individual and time fixed-effects. The TWFE estimates in column (1) comes from the Callaway and Sant’Anna (2021) estimator. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table A3: Effect of administrative data access on academic outcomes – Accounting for the 'Collaborator' Status

	Full Sample (1)	Controlling for Collaborator Status (2)	Excluding Collaborators (3)
Manuscript with INPS data			
At least one VisitINPS WP	0.188***	0.189***	0.217***
	(0.030)	(0.030)	(0.040)
At least one manuscript with INPS data	0.180***	0.181***	0.172***
	(0.029)	(0.029)	(0.032)
Presence at Labour economics conference			
Presence at EALE conference	0.018	0.019	0.010
	(0.048)	(0.048)	(0.050)
Presence at AIEL conference	0.175***	0.177***	0.149***
	(0.041)	(0.041)	(0.043)
Publication records			
At least one peer-reviewed publication	0.053	0.053	0.069
	(0.052)	(0.052)	(0.057)
At least one peer-reviewed publication in Top-30	0.023	0.023	0.024
Economics journals	(0.030)	(0.030)	(0.036)
At least one peer-reviewed publication in a high-ranked	0.063***	0.063***	0.044*
Labour economics journal	(0.023)	(0.023)	(0.025)
Academic impact and network			
Yearly citation count	9.318**	9.297**	6.988
	(4.143)	(4.150)	(4.433)
Work with a high-ranked economist	0.002	0.002	0.012
	(0.032)	(0.032)	(0.036)
Work with a high-ranked Italian economist	0.018	0.018	0.017
	(0.033)	(0.033)	(0.039)
Career evolution			
Recent academic promotion	0.106***	0.105***	0.108***
	(0.023)	(0.024)	(0.027)
Work in a high-ranked Economic department	-0.089*	-0.088*	-0.051
	(0.046)	(0.046)	(0.045)
Work in a high-ranked Italian Economic department	0.099**	0.099**	0.105**
	(0.050)	(0.050)	(0.048)
Observations	2087	2087	1902

Notes: Coefficients show the effect of accessing administrative data via the VisitINPS program using the Callaway and Sant'Anna (2021) estimator. Column (1) reports the baseline effects. Column (2) is the baseline model augmented with a dummy equal to one when an individual becomes Collaborator on an accepted VisitINPS proposal. Column (3) excludes all the individuals with the status of Collaborator. Outcomes are measured yearly. Standard errors in parentheses are clustered at the individual level. All regressions include individual and time fixed-effects. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.

Table A4: Balance test – A year prior submitting a VisitINPS proposal by characteristics of applicants

	Gender of applicant		Difference	Top PhD program		Difference
	Female	Male		Yes	No	
	(1)	(2)	(3)	(4)	(5)	(6)
Female	1.000	0.000	.	0.319	0.442	-0.123
	[0.000]	[0.000]	(.)	[0.471]	[0.498]	(0.082)
Postdoc	0.202	0.227	-0.025	0.064	0.263	-0.199***
	[0.404]	[0.421]	(0.059)	[0.247]	[0.442]	(0.067)
At least one VisitINPS WP	0.000	0.000	0.000	0.000	0.000	0.000
	[0.000]	[0.000]	(0.000)	[0.000]	[0.000]	(0.000)
At least one manuscript with INPS data	0.000	0.008	-0.008	0.000	0.006	-0.006
	[0.000]	[0.092]	(0.010)	[0.000]	[0.080]	(0.012)
Presence at EALE conference	0.060	0.034	0.026	0.021	0.051	-0.030
	[0.238]	[0.181]	(0.029)	[0.146]	[0.221]	(0.034)
Presence at AIEL conference	0.036	0.042	-0.006	0.000	0.051	-0.051
	[0.187]	[0.201]	(0.028)	[0.000]	[0.221]	(0.032)
At least one peer-reviewed publication	0.214	0.185	0.029	0.064	0.237	-0.173***
	[0.413]	[0.390]	(0.057)	[0.247]	[0.427]	(0.065)
At least one peer-reviewed publication in a Top-30	0.000	0.017	-0.017	0.043	0.000	0.043***
Economics journal	[0.000]	[0.129]	(0.014)	[0.204]	[0.000]	(0.016)
At least one peer-reviewed publication in a high-ranked	0.000	0.000	0.000	0.000	0.000	0.000
Labour economics journal	[0.000]	[0.000]	(0.000)	[0.000]	[0.000]	(0.000)
Yearly citation count	6.798	8.529	-1.732	9.468	7.314	2.154
	[15.611]	[22.055]	(2.800)	[29.192]	[15.741]	(3.269)
Work with a high-ranked economist	0.012	0.017	-0.005	0.043	0.006	0.036*
	[0.109]	[0.129]	(0.017)	[0.204]	[0.080]	(0.020)
Work with a high-ranked Italian economist	0.048	0.042	0.006	0.021	0.051	-0.030
	[0.214]	[0.201]	(0.029)	[0.146]	[0.221]	(0.034)
Recent academic promotion	0.048	0.059	-0.011	0.000	0.071	-0.071*
	[0.214]	[0.236]	(0.032)	[0.000]	[0.257]	(0.038)
Work in a high-ranked Economic department	0.167	0.286	-0.119**	0.936	0.026	0.911***
	[0.375]	[0.454]	(0.060)	[0.247]	[0.159]	(0.030)
Work in a high-ranked Italian Economic department	0.071	0.076	-0.004	0.128	0.058	0.070
	[0.259]	[0.266]	(0.037)	[0.337]	[0.234]	(0.043)
Observations	84	119		47	156	

Notes: Standard deviations are in square brackets and standard errors are in parentheses. *, ** and *** indicate significance at the 10%, 5% and 1% levels respectively.

Table A5: Effect of administrative data access on academic outcomes – Two-Way Fixed-Effects Approach by characteristics of applicants

	PhD student (1)	Postdoctoral researcher (2)
Manuscript with INPS data		
At least one VisitINPS WP	0.233***	0.429***
	(0.038)	(0.142)
At least one manuscript with INPS data	0.253***	0.184***
	(0.040)	(0.060)
Presence at Labour economics conference		
Presence at EALE conference	-0.001	0.092
	(0.060)	(0.108)
Presence at AIEL conference	0.177***	0.184**
	(0.050)	(0.076)
Publication records		
At least one peer-reviewed publication	0.056	0.094
	(0.056)	(0.122)
At least one peer-reviewed publication in Top-30 Economics journal	0.007	0.094*
	(0.040)	(0.055)
At least one peer-reviewed publication in a high-ranked Labour economics journal	0.112**	0.054
	(0.043)	(0.046)
Academic impact and network		
Yearly citation count	4.575	21.142***
	(4.193)	(8.900)
Work with a high-ranked economist	-0.007	0.107
	(0.035)	(0.097)
Work with a high-ranked Italian economist	0.005	-0.016
	(0.037)	(0.066)
Career evolution		
Recent academic promotion	0.131***	0.088
	(0.021)	(0.073)
Work in a high-ranked Economic department	-0.065	-0.203
	(0.049)	(0.123)
Work in a high-ranked Italian Economic department	0.166***	-0.023
	(0.058)	(0.127)
	1443	635

Notes: Coefficients show the effect of accessing administrative data via the VisitINPS program using the Callaway and Sant'Anna (2021) estimator. Outcomes are measured yearly. Standard errors in parentheses are clustered at the individual level. All regressions include individual and time fixed-effects. *, ** and *** indicate significance at the 10%, 5% and 1% levels, respectively.