**By submitting an application, I agree to the EGR submission requirements**

|  |  |
| --- | --- |
| First Names (Please list all co-authors) | Nicolas |
| Surname/Last Name (Please list all co-authors) | Fernandez-Arias |
| Select Focus Area (Analytics, Economics, Or Relevance) | Economics |
| Proposed Title | Productivity Growth with Creative Destruction by Employee Spinouts |

|  |
| --- |
| A problem statement illustrating your research project |
| Productivity growth is of central concern to economists. A particularly salient example of productivity growth is that of the semiconductor industry in Silicon Valley. In explaining this rapid growth, the economics literature has pointed to high rates of knowledge diffusion through employee mobility, and in particular creative destruction by employee spinouts: that is, the formation of competing startups by former employees who benefit from knowledge or ideas developed at their former employer.  My project seeks to quantify the role of this process on productivity growth for a broader swath of the economy. By combining LinkedIn data with other datasets (Compustat, Crunchbase, NBER-USPTO patent database), I will construct a novel dataset of spinout entrepreneurship across the United States. I will then use this dataset to calibrate an economic model of productivity growth. This model features the following tradeoff: while encouraging spinouts may allow more productive and innovative firms to flourish, incumbent firms may reduce their investment in productivity-enhancing knowledge if they know it will eventually be used in spinouts that compete against them. Using the calibrated model, I will explore the implications of economic policies on productivity growth. Examples of such policies include R&D subsidies, subsidies for small business formation, and restrictions on the legality of non-competition clauses in employment contracts. |
| A brief description of your solution to the research problem |
| My solution to this problem involves using resume data from LinkedIn to explore the career history of founders of startups in the LinkedIn and/or Crunchbase dataset. Since LinkedIn resume data contain job descriptions, they are particularly suited to answering the question of which startups are founded by people who honed their relevant skills at firms they eventually compete against. This kind of analysis would not even be possible with adminstrative data from the United States federal government.  Next, I will match the resume data with data from Compustat, which has information on R&D expenditures, and data from the USPTO, which has information on patents and their citations. The match will be based on name, location, URL of the firms website, and potentially other identifying information if it helps. I will use these data to explore the causal relationship between R&D by incumbent firms and spinout formation, using changes to state-level R&D tax credits as an instrumental variable (based on the 2013 *Econometrica* article by Bloom et al., “Identifying technology spillovers and product market rivalry”).  Finally, I observe employment and fundraising at these startups (from LinkedIn company pages and/or Crunchbase data), and see how they compare to similar startups that are not spinouts. Using the USPTO, I can also observe any patents awarded to these startups and their future citations, providing an independent measure of their innovativeness, and how it compares to that of ordinary entrants and the firms their founders came from (building on recent work by Akcigit & Kerr, “Growth through Heterogeneous Innovations,” forthcoming in the *Journal of Political Economy*). All of these statistics, through the lens of my structural economic model, inform structural parameters governing the productivity advantage of spinout firms over ordinary firms. |
| A list of the data from LinkedIn that you believe you will need or want to access to conduct your proposed research |
| Individual profiles for founders of firms in LinkedIn Company database (and/or Crunchbase database). All that is essential is information on previous employment for the previous several years. E.g., the name of the employer, job description of the employee.  Company pages for these firms: description of firm, number of employees (over time would be ideal, snapshot would still be helpful). |
| Resources you plan to invest, e.g. how many people will join the research, what help you are looking for from LinkedIn side, and how much time you will need to complete the research |
| If granted access, this will be my main project for the entirety of next year. I can also hire an RA if I need to. I will need no more than one year to complete the research (I plan on going on the economics academic job market with this project in Winter 2019-2020). |
| Timeline, execution plan, key deliverables and milestones for the project |
| Timeline:  Winter 2018 - Spring 2019:  1) Match founders in Crunchbase snapshot with their LinkedIn profiles  2) Extract previous employers and jobs data from these LinkedIn profiles  3) Parse information about jobs from resume entries  4) Match previous employers with firms in Compustat, Crunchbase and US Patent database  5) Classify startups as spinouts when they are founded by someone who recently worked at a competing firm, using data on:  (a) Previous employment of founders (firm names, dates, job descriptions)  (b) Industry / products of previous employers (line of business from Compustat, company descriptions on LinkedIn or Crunchbase)  (c) Industry / products of firm (company descriptions on LinkedIn or Crunchbase)  6) Match data on fundraising and employment from Crunchbase and LinkedIn (match by company name)  7) Estimate relationship between R&D spending by incumbents and employee spinout formation, using instrumental variable analysis  8) Deliver dataset prototype to LinkedIn  Spring - Summer 2019:  (1) Combine statistics from dataset constructed above with model to estimate structural parameters  (2) Conduct counterfactual exercises using structural model.  (3) Write up results in working paper.  Fall 2019:  (1) Deliver working paper to LinkedIn.  (2) Present work and incorporate feedback.  Winter 2019-2020:  (1) Finish paper and deliver.  Key deliverables: I will deliver a dataset (i.e. with name matching performed with Compustat and the NBER-USPTO) in summer 2019, a working paper in the fall of 2019, and a completed paper in Winter 2020 which I will submit for publication in an academic journal. |
| Names, affiliations, postal addresses, phone numbers, email address, and LinkedIn profile of the participants |
| Nicolas Fernandez-Arias, PhD Candidate in Economics at Princeton University  66 Linden Lane Apt 5, Princeton, NJ 08540  [nfernand@princeton.edu](mailto:nfernand@princeton.edu)  https://www.linkedin.com/in/nicolas-fernandez-arias-77705469/ |
| A short bio highlighting your background, expertise, and achievements, including prior relevant research |
| I am a 5th year PhD candidate in economics at Princeton University (expected graduation: June 2020). Previously I worked at Bridgewater Associates as an investment associate and Brookings Institution as a research assistant. For my undergraduate I did an AB in Mathematics, also at Princeton.  Within economics, my fields of specialization are macroeconomic and finance. My research explores the determinants of aggregate productivity growth.  My previous relevant research includes:  “Productivity Growth with an Aging Population: a Model with Vintage Human Capital”: This project proposes that an aging population could slow productivity growth through disincentivizing productivity enhancing investments that will only be adopted by the next generation. The model is calibrated using employer-employee matched data from the German social security administration. In doing this project, I have become well-acquainted with the many precautions that must be taken to ensure that sensitive personal data is used responsibly.  “Knowledge Diffusion in Economics: Advisor-advisee relationships” (preliminary): This project studies the spread of techniques and ideas in economics through advisor-advisee relationships and coauthor relationships. This project is relevant to the current work as it involves parsing text data (the text of research articles) and more specifically CV / resume data (information about advisors and coauthors). |
| Coding language (e.g. Python, R, Pig/Hive) skills and experiences |
| Proficient in R, Julia, Python, Matlab  Learning Spark  This fall I took a course in High Performance Computing for Economists (website: <https://www.sas.upenn.edu/~jesusfv/teaching.html>, “Lectures on Computational Methods”) |
| Additional details |
| References:  Esteban Rossi-Hansberg, Princeton University  Theodore A. Wells '29 Professor of Economics and International Affairs  erossi@princeton.edu  Gianluca Violante, Princeton University  Professor of Economics  violante@princeton.edu  Ezra Oberfield, Princeton University  Assistant Professor of Economics  edo@princeton.edu |

Please send your proposals in PDF format to [EconomicGraphResearchProposal@linkedin.com](mailto:EconomicGraphResearchProposal@linkedin.com) with the following subject line: “Analytics/Economics/Relevance: [Proposal title]” by December 1, 2018. By downloading and submitting a proposal, the individual submitting the proposal and each teammate agree they have read, understood and agree to be bound by the [proposal requirement](https://engineering.linkedin.com/data/economic-graph-research/economic-graph-details)s on the Economic Graph Research website. The individual submitting the proposal and each teammate agree that LinkedIn may conduct use, research, or develop software, products, or features that are the same or similar to your proposal and that you waive any claims with respect to any use, research, or developments by LinkedIn.