Cover letter

I am a fourth-year Ph.D. Candidate in Management of Organizations at the Haas School of Business, University of California, Berkeley. In my dissertation, I am researching how new information digitized in the form of *data* becomes a key input into the innovation production function and changes the dynamics of novelty generation. My research explores how data differ from other information goods, and how firms and scientists employ them to generate knowledge that is useful for innovation and entrepreneurship. I am especially keen on understanding organizational and individual-level dynamics behind the use of large and complex data, and the possible trade-offs that data-driven innovation entails for companies.

Contact details:
Matteo Tranchero
Ph.D. Candidate, Management of Organizations
Haas School of Business, University of California Berkeley
US Phone: +1 (341) 400-3543; Italian Phone: +39 3495672454
m.tranchero@berkeley.edu

Dissertation Committee Members:
Lee Fleming, Professor, Coleman Fung Institute for Engineering Leadership, UC Berkeley (co-chair)
Abhishek Nagaraj, Assistant Professor, UC-Berkeley Haas (co-chair)
Toby Stuart, Professor, UC-Berkeley Haas
Enrico Moretti, Professor, Department of Economics, UC Berkeley

Letter writer:
Abhishek Nagaraj, Assistant Professor, UC-Berkeley Haas
+1 (510) 6447746
nagaraj@berkeley.edu

Matteo Tranchero

CONTACT

m.tranchero@berkeley.edu
+1(341)-400-3543

2220 Piedmont Ave, F533
Haas School of Business
Berkeley, CA

EDUCATION

University of California, Berkeley, Haas School of Business
PhD Candidate in Management of Organizations
Dissertation Committee: Lee Fleming, Abhishek Nagaraj, Toby Stuart, and Enrico Moretti
MS in Business Administration, 2020

Sant'Anna School of Advanced Studies, Pisa
Advanced Diploma in Economics, *cum Laude*, 2019
Diploma in Economics, *cum Laude*, 2017

University of Pisa
MS in Economics, *cum Laude & Distinction*, 2018
BS Economics & Management, *cum Laude*, 2016

FURTHER EDUCATION

Georgia Institute of Technology, School of Public Policy
Visiting Scholar, Fall 2017

École Normale Supérieure, Rue d'Ulm, Paris
Exchange Program, Spring 2016

WORKS IN PROGRESS

Is Correlational Knowledge useful for Corporate Innovation? Evidence from Genome-Wide Association Studies

Data-Driven Search and Innovation

How Do Data Shape Science? Evidence from U.S. Census' Federal Statistical Research Data Center Program
(with Abhishek Nagaraj)

Looking Where the Map is: The Streetlight Effect in Data-Driven Exploration
(with Johannes Hoelzemann, Gustavo Manso, and Abhishek Nagaraj)

PRE-DOCTORAL PUBLICATIONS

Alessandro Nuvolari, Valentina Tartari, and Matteo Tranchero. "Patterns of innovation during the Industrial Revolution: a reappraisal using a composite indicator of patent quality", *Explorations in Economic History*, 82, 101419, 2021 (Lead article)

Giovanni Dosi and Matteo Tranchero. "The role of comparative advantage, endowments and technology in structural transformation", in Alcorta et al. (eds), *New Perspectives on Structural Change: Causes and Consequences of Structural Change in the Global Economy*, 2021, Oxford: Oxford University Press

Valeria Cirillo, Arianna Martinelli, Alessandro Nuvolari, and Matteo Tranchero. "Only one way to skin a cat? Heterogeneity and equifinality in European national innovation systems", *Research Policy*, 48(4), 905-922, 2019

INVITED WORKSHOPS & PRESENTATIONS

2021: 9th BITSS Annual Meeting, UC Berkeley *(virtual)*
4th Annual Wharton Innovation Doctoral Symposium, UPenn *(virtual)*
NBER Economics of Digitization Tutorial *(virtual)*

Genealogies of Data Junior Scholars Workshop *(virtual)*
NBER Workshop on the Economics of AI *(virtual)*
WICK#9 - PhD workshop, Collegio Carlo Alberto, Turin *(virtual)*
RISE4 workshop, Max Planck Institute for Innovation and Competition *(virtual)*
2020: WICK#7 - PhD workshop, Collegio Carlo Alberto, Turin
2019: Economic History Association (EHA) Conference, Atlanta
2018: Workshop "New Perspectives on Structural Change", UNU-MERIT, Maastricht
INET-YSI European Regional Convening, Trento
2016: SPRU 50^{th} Anniversary Conference, University of Sussex, Brighton
INET-YSI Plenary, Central European University, Budapest

Teaching

University of California, Berkeley
Teaching Assistant to Abhishek Nagaraj, 2021, 2022
Entrepreneurial Strategy (EWMBA 295T)
Teaching Assistant to Sameer Srivastava, 2019, 2020, 2022
Power & Politics in Organizations (EWMBA 254)
Teaching Assistant to Jon Metzler, 2021
Strategy (MBA 299)

Awards And Grants

2022–24: Alfred P. Sloan Foundation grant ($398,970, with Abhishek Nagaraj)
2022–23: Strategy Research Foundation (SRF) Dissertation Scholar ($10,000)
2021–22: Panmure House Prize - Emergent Thought Award, Edinburgh ($25,000)
2021: Institute for Business Innovation grant, UC Berkeley ($10,000, with Lukas Leucht)
2019–20: Alessandro Pansa Scholarship, Feltrinelli Foundation, Milan (€20,000)
2019: Bernardo Nobile Prize for the best Master's thesis using patent data, Trieste
2019: The Ryoichi Sasakawa Young Leader Fellowship, UC Berkeley
2017: Summer School Program, Unicredit and Universities Foundation, Milan
2013: 5-years merit-based full scholarship, Sant'Anna School of Advanced Studies, Pisa

Service

First-generation, low-income student mentor through *Boost@BerkeleyHaas*, 2022–present
Mentor for *Undergraduate Research Apprentice Program*, UC Berkeley, 2020–present
Member of the DEI Taskforce in the Management of Organizations group, 2021–22
Tutor for Underrepresented Minorities in Economics, *Berkeley Economists for Equity*, 2021
Assistant to the Managing Editor, *Industrial and Corporate Change*, 2020–present
Referee for *Academy of Management*, *Economics Bulletin*, *Industrial and Corporate Change*

Further Experience

2021: Research Assistant for Prof. David Teece, BRG Institute
2020: Graduate Remote Instruction Innovation Fellowship, UC Berkeley
2018: Research Trainee, Italian Institute of Statistics, Rome
2017: Summer School in Microeconometrics, Barcelona Graduate School of Economics
2016: Summer School in Money and Banking (Grade: A+), London School of Economics
2015: Commercial Affairs Intern, Permanent Mission of Italy to the WTO, Geneva
2013: Personal Assistant, Cllr. Lorraine Kingston, Deputy Lord Mayor of Cork, Ireland

Languages

Italian (native), English (fluent), French (intermediate)

Last Updated: January 19^{th}, 2022

List of completed PhD courses

ARESEC 210: Probability and Statistics, grade: A (taught by Aprajit Mahajan)
ARESEC 212: Econometrics: Multiple Equation Estimation, grade: A (taught by Max Auffhammer)
ECON 244: Applied Econometrics, grade: A (taught by Pat Kline and Chris Walter)
ECON C222: Economics of Innovation, grade: A (taught by Lee Fleming)
ECON 250A: Labor Economics I, grade: A (taught by Pat Kline and Chris Walter)
ECON 250B: Labor Economics II, grade: A+ (taught by David Card)
INFO 206: Introduction to Programming for Information Management Application, grade: A+ (taught by Paul Laskowski)
INFO 251: Applied Machine Learning, grade: A (taught by Joshua Blumenstock)
PHDBA 297T: Doctoral Topics in Business Administration, grade: A+ (taught by Abhishek Nagaraj)
PHDBA 259B: Research in Macro-Organizational Behavior, grade: A (taught by Toby Stuart, Sameer Srivastava, Mathijs de Vaan)
PHDBA 259A: Research in Micro-Organizational Behavior, grade: A (taught by Don Moore, Jenny Chatman, Cameron Anderson)
SOCIOL 280D: Advanced Study in Substantive Sociological Fields: Organizations, grade: A+ (taught by Heather Haveman)
PHDBA 259S: Research Seminar in Organizational Behavior and Industrial Relations, grade: *Pass*
PHDBA 375: Teaching Business, grade: *Pass*

Overall GPA in doctoral coursework: 4.0
Qualifying Exam passed with Distinction.

Ongoing dissertation research

A few years ago, *The Economist* made a bold claim: the world's most valuable resource is no longer oil, but data. Just like the oil in the 18th century, the analogy goes, data is an untapped valuable asset that can become the fuel of the knowledge economy. Recent research has started investigating this idea, mostly looking at the impact of data as an input into production (Jones and Tonetti, 2020; Farboodi and Veldkamp, 2020), managerial decision-making (Brynjolfsson and McElheran, 2016), entrepreneurship (Bessen et al., 2021), innovation (Kao, 2021) and entrepreneurship (Nagaraj, 2021). Consistent with the view of data as oil, this strand of research tends to conceive it as a rather amorphous intermediate good into the production of some final good. However, this approach is quite narrow in scope and neglects the role that data might play in shaping the *direction* of innovative activities. In my work, I argue that data can affect novelty generation not just as direct input, but also if it leads to focus on alternative topics or to ask novel questions that fundamentally redirect experimental efforts in new directions.

My dissertation investigates how data are changing innovation and knowledge generation in a variety of corporate and academic settings. In particular, amidst the surge of big data, firms are increasingly making use of what I call *correlational knowledge*: namely, data-driven findings or predictive statements. Data-driven predictions fundamentally differ from the traditional conception of knowledge intended as an understanding of cause-effect relationships that can be directly employed in innovation. On the contrary, predictions are hardly causal in nature and thus prone to the risk of false positives. This is especially true in complex and dynamic environments, such as the search for innovations in the technological domain. Whether and under which conditions data-driven knowledge is useful for innovation is an open question that I tackle in my dissertation.

I suggest that correlational knowledge can impact corporate innovation by leading firms to explore promising but understudied areas of the technological landscape. In this perspective, predictive statements are important for firms not so much for their accuracy as for their effectiveness in generating useful technological exploration and escape from the trap of exploitative search. My current work investigates these hypotheses focusing on a novel setting in the context of pharmaceutical innovation. I empirically study how firms react to the exogenous arrival of innovation-relevant correlational knowledge on new genetic drug targets. My source of identification comes from a type of genomic research, called *genome-wide association studies* (GWAS), which are used to discover novel gene-disease associations to guide experimental research and drug development. GWAS scan the whole genome in search of significant associations with diseases, without targeting any specific gene ex-ante, providing credible variation at the level of individual genes. Crucially, GWAS do not provide cause-effect evidence, but just hints of potentially important unknown relations among previously known genes and diseases—in sum, correlational knowledge on promising targets for scientific investigations and pharmaceutical innovation.

In terms of measurement, I am assembling new data to go beyond existing studies that rely on patent-to-paper citations to track how firms use science in innovation. I evaluate the effect of data-driven

knowledge by looking at subsequent work that directly builds over the novel gene-disease association, rather than capturing this indirectly using citations to the paper that introduces the correlational finding. To do so, I employ state-of-the-art machine learning tools to extract all the genes mentioned in PubMed publications, USPTO patents, and clinical trials. This allows me to construct a novel, comprehensive gene-level database where I can investigate the role of correlational knowledge in a difference-in-differences framework. Preliminary results validate this novel measurement approach, confirming its potential to be applied to a variety of questions that go beyond my current dissertation topic.

Future research agenda

I plan to keep investigating a broader agenda centered on the role that data have in innovation. Together with Johannes Hoelzemann, Gustavo Manso, and Abhishek Nagaraj, we are studying how data affects decision-making under uncertainty. While data is typically seen as reducing uncertainty and improving welfare, we present a simple theoretical framework where data provision can decrease group and individual payoffs. We test our theory in an online lab experiment where agents are presented with multiple risky projects and must learn about their payoffs through costly exploration, finding that data provision on the true value of one project can hurt individual payoffs by as much as 11% and reduce the group's likelihood of discovering the optimal outcome by 52%. These results speak to the often-heard complaints that increased data availability could lead researchers to focus on uninteresting questions or crowd out new theory generation. Indeed, in a related project with Abhishek Nagaraj, we are testing these predictions empirically by exploiting differential access to confidential data for academic researchers in economics. To identify the causal impact of data access, we study the gradual expansion of the Federal Statistical Research Data Center (FSRDC) program launched by the U.S. Census Bureau in 1994.

Why I want to attend the Boot Camp

There are many reasons why I am applying to the NBER Entrepreneurship Research Boot Camp 2022. One is certainly the stellar faculty that will lead the lectures, and another is the possibility of broadening my horizon by learning more about the research topics that I love so much. However, of all the good reasons to apply, I confess that what thrills me the most is the perspective of engaging with peers and researchers from institutions all over the U.S. and the world during the boot camp. During the pandemic, I have especially struggled to be part of an intellectual community of students that share my interests. I am alone in my cohort at Haas, and there are no students in the two years above me, so I find it difficult to get support and guidance from peers with the same background and passion for entrepreneurship research that I have. I hope that this program will create a learning community that can help me not only research-wise but also emotionally during the final stretch of the PhD. My dream would be traveling to Cambridge alone, but coming back as part of a community of PhD students (and future colleagues) that share the same passion for research on innovation and entrepreneurship. Therefore, I am deeply convinced that meeting fellow students, as well as participants to the NBER Summer Institute (which I never had a chance to attend), will be in and of itself an invaluable part of the boot camp. This experience will immensely benefit my academic and personal growth.

Berkeley
UNIVERSITY OF CALIFORNIA

Abhishek Nagaraj
Assistant Professor
Haas School of Business

2220 Piedmont Ave, F565
Berkeley, CA 94720
nagaraj@berkeley.edu
http://www.abhishekn.com

Dear Colleague,

I'm writing this letter to strongly recommend Matteo Tranchero for the NBER PhD bootcamp in Entrepreneurship (ERBC) for summer 2022. I have no doubt that Matteo is going to be one of the most impactful scholars in the economics of innovation and entrepreneurship in the next decade. In terms of maturity, he will be a rising 5th year this summer and will have 3-4 incredibly strong papers in his pipeline (more on that below). And he is incredibly dedicated, hard-working and committed to the profession — and a fountain of useful spillovers for those around him. As someone quite familiar with the PhD bootcamps at the NBER, I cannot think of a more ideal candidate — I feel like he would benefit a lot from participating, and the program itself would be much richer with him in there. I recommend him without hesitation and despite the significant capacity constraints you face, urge you to seriously consider him for participation.

Matteo is interested in examining how new sources of data are shaping the innovation process. In particular, his dissertation project (which will be his job-market paper) examines the impact of Genome Wide Association Surveys (GWASes) on pharmaceutical and scientific innovation in the biomedical sciences. GWASes are an interesting phenomenon because they sequence the genome of a large number of individuals with a specific disease (say diabetes) and by comparing their genome with a control group, they point out likely locations of genetic mutations that could underlie the onset of the disease among humans. GWASes are not "science" in the traditional sense — they are rough "maps," pointing out hot-spots in the genome that scientists should explore in order to better understand the disease and develop treatments.

Matteo's research so far shows that GWASes are tremendously useful for follow-on innovation. Even though he is studying academic and firm research and development in areas with large economic markets (e.g. diabetes) —he finds that GWASes can significantly alter the course of follow-on innovation. The key mechanism is that GWASes highlight how the research enterprise can get "stuck" in local exploitation, and how data like GWAS maps can open up uncharted territories (areas of the genome) that scientists have hereto ignored. This is best illustrated by an amazing case study Matteo has developed showing how a GWAS for irritable bowel diseases (IBDs) like Chron's disease and ulcerative colitis identified a novel genetic marker (IL23R), which had previously been completely ignored in the literature. The publication of the GWAS led to both basic research to understand the pathways through which IL23R could lead to IBD, as well as the development of new therapeutics to help develop treatments for the disease.

So far, Matteo has done a tremendous amount of work understanding the phenomenon and collecting an incredible dataset linking the original studies, follow-scientific publications and related patents to help paint a complete picture of how correlational datasets like GWASes shape biomedical innovation. Of particular note is a completely novel method to link GWASes to patent data. Since patents rarely cite GWASes directly, Matteo has worked with an industry collaborator to use ML techniques to extract "entities" from patents, including the specific genes that might be mentioned in patent claims. This allows him to complete a gene-level matching exercise between a patent and the results of a given GWAS, along with the particular disease a patent might be targeting. This approach and resulting dataset is a goldmine — as it will allow Matteo to examine firm and strategic responses to GWAS publications, understand where follow-on innovation comes from and examine the nature of follow-on innovation in terms of its direction. So far, Matteo has already completed one descriptive

paper on GWASes using publication data and is working on a more ambitious study to further exploit the patent data as explained above.

If the GWAS papers were all that Matteo had in his pipeline - he would already be at the near top of his cohort in terms of academic progress. However, working with me (and a few other collaborators) Matteo has promising results for two additional sets of projects.

In one project, we (along with Gustavo Manso and Johannes Haushofer) have developed an experimental framework to understand how data provision affects risky exploration in a "bandits" setup. We are modeling the possibility that data might causes innovators to "look where the light is" — and theoretically and experimentally show when this might be the case. Our headline finding is that data (such as GWASes or satellite maps) are useful when they reveals "duds" (i.e. bad projects) or "winners" (i.e. global optima) — but when they reveal "medium" projects (local optima) they can cause under exploration and lead to suboptimal outcomes at both the firm and social level.

In the second project, Matteo and I are exploring how data access affects follow-on innovation in the context of data from the US Census Bureau (through their statistical data center program) on economics research. Here we are exploring the gradual expansion of physical statistical data centers across the US, matched with a complete dataset of academic economics publications to understand how access to census micro-data shapes academic research in economics. We are finding that data access is associated with a significant increase in the likelihood of "top 5" publications for affected economists, but this effect holds only for empiricists rather than theorists. Our results point to the role of data access in shaping the empirical turn in economics, and also speak to epistemological debates about the role of observations and theory in scientific progress.

As you can tell by this long letter, I am energized and excited about the potential for Matteo to develop as an academic scholar. My one regret is that Matteo's exposure to the academic innovation and entrepreneurship community has largely been through the PhD class I teach — rather than through direct interactions with the researchers he is building on and citing. As someone who got his PhD in Cambridge and frequently attends NBER gatherings, I know how valuable it can be to physical interact with the normally invisible college, and so I know that participating in the bootcamp would help Matteo better frame his ideas, deepen his theoretical and empirical training and introduce him to a network that will further propel his academic journey.

In sum, I recommend Matteo Tranchero in the highest terms — please do not hesitate to reach out in case of any questions or clarifications.

Thank you.

Abhishek Nagaraj
Assistant Professor, UC Berkeley Haas