Journal of Economic Geography (2023) **23**, 343–369 https://doi.org/10.1093/jeg/lbac016
Advance Access Publication 12 June 2022

Stars as catalysts: an event-study analysis of the impact of star-scientist recruitment on local research performance in a small open economy

John McHale *,†, *Jason Harold* *, *Jen-Chung Mei* **,
*Akhil Sasidharan** and *Anil Yadav**

*Whitaker Institute for Innovation and Societal Change, National University of Ireland Galway, Galway, Ireland
**Bennett Institute for Public Policy, University of Cambridge, Cambridge CB2 1TN, UK
†Correspondence to: *john.mchale@nuigalway.ie*

Abstract

There is increasing interest among policymakers in small open economies in the use of star-scientist recruitment policies to catalyse the development of local clusters in targeted research areas. We use Scopus to assemble a dataset on over 1.4 million publications and subsequent citations for Denmark, Ireland and New Zealand from 1990 to 2017. An event-study model is used to estimate the dynamic effects of a star arrival on quality-adjusted research output at both the department and matched individual incumbent levels. Star arrivals are associated with statistically significant increases in department output (excluding the output of the star) of between 12% and 25% after 4 years. At the incumbent level, star arrivals lead to an approximately 5% increase in individual output, with substantially larger increases for incumbents who co-author with the star.

Keywords: Star scientists, small open economies, innovation, scientist mobility
JEL classifications: O31, J24, J61, I23
Date submitted: 9 June 2021 **Editorial decision:** 24 May 2022 **Date Accepted:** 27 May 2022

1. Introduction

Policymakers in small open economies face particular challenges in justifying investments in scientific research. Smallness can act as a disincentive in two ways. First, given that investments add to the global stock of knowledge, smallness means that a large share of the benefits flows outside as spillovers to the global stock. And second, given the importance of scale economies for scientific productivity—and notably the importance of local knowledge-sharing and collaboration networks—a lack of scale could put small-country research institutions at a competitive disadvantage. Strategies such as targeting research with substantial local spillovers (especially more applied and use-inspired basic research), developing of centres of excellence to achieve necessary scale and integrating into international knowledge and collaboration networks[1] are pursued to overcome this size disadvantage. The targeted recruitment of star scientists is a potential additional element in the strategy mix for policymakers and research institutions. One objective behind such policies

1 For example, participation in EU-wide funding consortia for European Research Council funding.

is that 'stars'—disproportionately productive and connected scientists in their field—will catalyse the development of successful local clusters in targeted areas. In addition to their own direct outputs, it is hoped that stars will raise the productivity of their peers through such channels as increased access to knowledge networks, greater collaboration opportunities, mentoring and even changed norms relating to the conduct of scientific research. It is also hoped that they aid in the attraction and retention of other scientific talent, thereby supporting the growth of targeted high-performing clusters.

As examples of nationally supported star recruitment policies, the Danish National Research Foundation (DNRF) has implemented a series of programmes: the Niels Bohr Visiting Professorships, the DNRF Professorships, the Niels Bohr Professorships and the latest DNRF Professorships (2021). Although the design has shifted somewhat over time, the core purpose has been the 'enriching of Danish research communities with top-class researchers from abroad'. Attracting outstanding international research talent to Ireland is also the objective behind the Science Foundation Ireland (SFI) Research Professorship and Future Research Leadership Programmes. Since the initial launch in 2003, programme design has also evolved, but a consistent goal has been to support research institutions in their recruitment of world-class researchers.[2]

We hypothesise that the productivity of a scientist depends on their position in scientific networks and that an important component of these networks is local. The relevant networks could include, for example, those that support knowledge access or provide co-authorship opportunities. The importance of location is assumed to reflect the lower cost of forming network links when scientists are co-located. An arrival of a star at a scientist's department is hypothesised to strengthen that scientist's network in ways that support their productivity. Moreover, if other scientists are attracted by the presence of a star, and if already present scientists are less likely to leave with that presence, the arrival of the star could further catalyse the development of the department through improved recruitment and retention.[3]

We use a (panel) event-study framework to explore the dynamic impacts of star arrivals (excluding the direct publication impact of the star) on receiving university department and individual incumbent performance in three small countries—Denmark, Ireland and New Zealand. We assemble Scopus data on publications and citations to those publications for the period 1990–2017. Star scientists are defined as scientists above a particular percentile (the 95th percentile and above in our baseline specification) in the cumulative distribution of quality-adjusted output, where the relevant quality measure is based on observed citations to a publication to the end of the sample period. In turn, star arrivals are identified as long-term arrivals at institutions where they were not previously affiliated as evidenced by the affiliations recorded on their publications. We implement a standard event-study model with the assumption of homogenous arrival effects across arrival cohorts and also allow for heterogeneous arrival effects across cohorts using the method of Sun and Abraham (2021).

Notwithstanding the value of the event-study design to protect against a number of sources of endogeneity bias, there remains a concern that the decision of a department to hire a star, and/or the decision of a star to join a department, are not independent of

2 Conditional on funding availability, SFI's recent strategy for the years to 2025 targets the recruitment of '20 world-class researchers to Ireland annually' (Science Foundation Ireland, 2021, 9).

3 We develop a simple model of such catalytic effects of star arrivals in Online Appendix A.

foreseeable but uncontrolled for factors that affect departmental performance. A complementary focus on individual incumbent scientists at the time of the star's arrival allows us to make more credible causal inferences for one important component of the overall star-arrival effect on performance. There are a number of reasons. First, while a star's arrival may be influenced by expectations of future departmental performance (including future planned investments), the arrival is less likely to be influenced by expectations related to individual scientists. Second, we can implement a matching procedure at the individual level where each incumbent is paired with a highly similar scientist at a non-star recruiting department at another institution. As detailed in the next section, we implement this matching with a coarsened exact matching (CEM) procedure. And third, we can better explore particular network-related causal mechanisms through which a star arrival affects productivity. Specifically, we explore how co-authorship with the star affects post-arrival incumbent productivity. We therefore present results both for department-year and (matched) individual-incumbent-year units of analyses.

Our paper is related to a number of literatures, including those on stars, networks, mobility and agglomeration. The tendency for economic and social outcomes to follow fat-tailed and skewed distributions—with outsized outcomes for *stars*—has been observed in many settings.[4] In science, such fat-tailed distributions have long been observed for publications, citations and collaborations (Lotka, 1926; Price, 1963; Newman, 2001; Goyal et al., 2006). These distributions could be explained by a highly uneven underlying distribution of talent or by network effects that allow small initial differences to become magnified over time through a cumulative advantage process (Merton, 1968; Azoulay et al., 2014). In a market setting with economies of scale, Rosen (1981) shows how small differences in ability can lead to outsized differences in rewards for 'Superstars' through a 'winner-take-all' phenomenon. Adler (1985) provides an alternative account of the emergence of stars that depends on network effects and first-mover advantage—indeed he models the emergence of stars even where initial talents are identical.

The literature on economic and social effects of *networks* has highlighted the potential effects of network structure—captured by metrics such as density, centrality and clustering—on individual behaviour and outcomes (Jackson, 2008; Jackson et al., 2017). Moreover, the economic approach to network formation emphasises how network participants weigh the costs and benefits of forming connections (Goyal, 2007). Of relevance to our setting, where co-location lowers the cost of forming connections, network clustering will arise at the local level suggesting the potential for arriving stars to become embedded in local networks even while retaining more distant connections (Jackson and Rogers, 2005, 2007).[5]

For scientist networks, an important strand of the literature has used exogenous star departures or arrivals to identify the impact of a disruption to a peer network on the productivity of members of that network. Waldinger (2012) uses the dismissal of scientists in Nazi Germany as a source of exogenous variation in the quality and quantity of the peer

4 See, for example, Gabaix (1999) for city sizes, Gyourko et al. (2013) for city house prices, Malmendier and Tate (2009) for CEO pay and Autor et al. (2020) for firm profits.

5 A large literature has investigated the tendency of knowledge flow to be disproportionately localised (see Griliches, 1992, for an early survey). In an influential paper, Jaffe et al. (1993) focus on the geographic localisation of citations to patents and find that citing and cited patents are disproportionately co-located after controlling for the geographic distribution of activity in the relevant area. Breschi and Lissoni (2009) find that a significant fraction of knowledge flows across firms and locations is due to labour mobility that is geographically bounded.

group of remaining scientists, but does not find evidence of harm. However, Waldinger (2010) does find evidence of harm from these dismissals on PhD students. Azoulay et al. (2010) find that collaborators suffer a lasting reduction in quality-adjusted output upon the unexpected death of a star, with collaborators closer to the star in 'idea space' suffering a sharper declines. Also using unexpected star deaths, Oettl (2012) reports that the unexpected death of 'helpful star scientists' negatively impacts the quality of but not the quantity of their co-authors' output. In a recent paper, using data from pharmacology and pharmacy, Khanna (2021) finds an across-the-board decline in co-author productivity following a star death, but the effect is attenuated by the size of the scientist's collaboration network. In an event-study setting, Agrawal et al. (2017) report evidence that in the field of evolutionary biology star arrivals increase the productivity of incumbents working in related areas and also increase the quality of subsequent hires.[6]

The traditional migration literature has generally treated *mobility* as a one-time, all-or-nothing event, examining, for example, the impact of an immigration shock on local wages (e.g. Card, 1990; Borjas, 2003) or the impact of the 'brain drain' on sending destinations. In contrast, a more recent literature has emphasised the dynamic nature of migration (or 'brain circulation')—including, for example, anticipation effects relating to the prospect of future moves and the importance of return migration (Kapur and McHale, 2006; Docquier and Rapoport, 2012; Kerr, 2019). The nature of moves can also be more partial with, for example, members of diaspora networks retaining connections to networks in the country or region of origin even as they form connections in their new locations. Agrawal et al. (2006) find evidence of enduring links that support knowledge flow between departed inventors and their original locations. Using data for Indian inventors, Agrawal et al. (2011) examine the trade-off between increased access to knowledge from connections to the diaspora and the thinning of local knowledge networks that results from inventor out migration.

Finally, our paper is related to the literature on *agglomeration* (Krugman, 1991; Black and Henderson, 1999; Duranton and Puga, 2004; Glaeser, 2008). Agglomeration models typically combine scale economies (including through network effects) and mobility (individuals and/or firms are attracted to highly performing regions). In a typical model, the spatial distribution of activity is affected by the balance between centripetal forces—pre-existing scale attracts newcomers leading to further scale—and centrifugal forces resulting from some form of congestion effect at the expanding location. Fujita et al. (2000) examine how the core-periphery pattern can emerge as the equilibrium outcome of such forces. Audretsch and Feldman (1996) report evidence on the importance of knowledge spillovers to the spatial clustering of activity in knowledge-intensive industries. One question is how a positive feedback loop that leads to a successful cluster of activity gets started. Of particular relevance to the importance of star scientists, Zucker et al. (1998) identify the pre-existing geographical distribution of star scientists as an important determinant of the geographical distribution of activity in the biotechnology industry.[7]

The remainder of the paper is structured as follows. Section 2 sets out the panel event-study framework that we use for the estimation of the dynamic effects of star arrivals

6 Lacetera et al. (2004) and Hess and Rothaermel (2011) examine the effect of star arrivals on firm-level innovation performance.

7 Also in the knowledge industry context, Agrawal and Cockburn (2003) find that the presence of a large R&D intensive firm—what they call an 'anchor tenant'—enhances the ability of the regional innovation system to absorb university research and stimulate knowledge-intensive output.

under the assumptions of both homogenous and heterogeneous cohort arrival effects. Section 3 describes our Scopus-derived data and outlines our star and star arrival identification strategies as well as our department/incumbent research output metrics and incumbent matching procedure. Section 4 reports our results for both the combined three-country sample and each of the three countries individually. Section 5 conducts a number of robustness tests on our baseline results. Section 6 concludes with a discussion of the relevance of our findings for policy and institution-level strategies.

2. Econometric methodology

2.1. Baseline distributed lead/lag specification

The gradual strengthening of network effects provides a mechanism through which the arrival of a star could impact departmental performance over time. To measure these dynamic effects, we begin with a general baseline two-way fixed effects (TWFE) specification that assumes the multiplicative infinite distributed lead/lag form:

$$\ln Y_{i,t} = \alpha + \sum_{j'=-\infty}^{\infty} \gamma_{j'} S_{i,t-j'} + X_{i,t}^{'}\lambda + \phi d_i + \varphi d_t + u_{i,t}, \tag{1}$$

where $Y_{i,t}$ is a measure of quality-adjusted output (e.g. citations-weighted publications) in department i in year t, $S_{i,t-j'}$ is a dummy variable indicating if a (first-arriving) star is present in the department in year $t-j'$, $X_{i,t}'$ is a vector of controls, ϕd_i is a department-fixed effect, φd_t is a year-fixed effect and $u_{i,t}$ is a zero mean error term. Depending on the specification, the unit of analysis can be departments in a given year or individual incumbents attached to a particular department in a given year.

The estimated dynamic multipliers, $\gamma_{j'}$, have the convenient interpretation of semi-elasticities. The cumulative dynamic multiplier j periods after a star arrives (where j can be negative if the star has yet to arrive) is[8]:

$$\beta_j = \sum_{j'=-\infty}^{j} \gamma_{j'} \text{ or } \beta_j = \beta_{j-1} + \gamma_{j'}. \tag{2}$$

We initially make the strong assumption of strict exogeneity so that the contemporaneous error is independent of past, present and future star arrivals. We assume that if a star arrives in a department that they remain present in the department from that time period forward so that the arrival of a star is an 'absorbing' event.[9] The multiplicative form allows for star arrival and time effects to be proportional to current department-level performance. These (proportional) star arrival effects in our baseline model are initially assumed to be homogeneous across time and departments/incumbents, but we then relax this assumption to allow for the possibility that star arrival effects are heterogeneous across arrival cohorts.

8 Therefore, j has the interpretation of 'event time'.
9 This assumption can be weakened to allow for star exits recognising that a star's arrival (and temporary presence) can have effects on the department even after they have left.

The presence of lead effects in our general specification allows in principle for anticipation effects of star arrivals on productivity performance. However, in our empirical analysis, we take a conservative approach and assume such anticipation effects are zero. Instead, we suppose that any observed pre-arrival effects—or, even more concerning, a positive trend in pre-arrival effects—is evidence of a failure of the parallel trends assumption. The absence of an observed pre-trend therefore helps mitigate concerns of a correlation between star recruitment and unobserved factors that independently affect department performance. However, even in the absence of a pre-trend, it is possible that contemporaneous (or subsequent) performance-enhancing developments in the department are associated with star recruitment. This could be because, for example, these developments attract the star (reverse causality) or possibly because the recruitment of the star is part of a package that includes other performance-enhancing changes in the department (omitted variables). Therefore, in the absence of a credible instrument for star recruitment, we will be cautious in inferring causality from any observed star effects on performance at the department level. However, the use of CEM allows us to make stronger causal inferences at the individual incumbent level.

2.2. Event-study specification with homogenous arrival effects across arrival cohorts and binning

The cumulative dynamic multipliers of interest (and their standard errors) can be estimated directly by reparametrising (2) as an event specification where the *event*[10] of a star arrival can take place in the past or in the future[11]:

$$\begin{aligned}\ln Y_{i,t} &= \alpha + \sum_{j=-\infty}^{\infty}\left(\left(\sum_{j'=-\infty}^{j}\gamma_{j'}\right)\Delta S_{i,\ t-j}\right) + X'_{i,t}\lambda + \phi d_i \ + \varphi d_t + u_{i,t}\\ &= \alpha + \sum_{j=-\infty}^{\infty}\beta_j \Delta S_{i,\ t-j} + X'_{i,t}\lambda + \phi d_i \ + \varphi d_t + u_{i,t}. \end{aligned} \tag{3}$$

To empirically implement this framework, we need to make assumptions about the effects of star arrivals when the arrivals take place in both the distant past and the distant future. We assume that the cumulative effect is constant at $\beta_{-\underline{j}}$ for all years up to $\underline{j}$ years before the arrival of the star and constant at $\beta_{\bar{j}}$ from $\bar{j}$ years after the arrival. With this 'binning' of effects, we can rewrite Equation (3) as

$$\ln Y_{i,t} = \alpha + \beta_{-\underline{j}}\sum_{j=-\infty}^{-\underline{j}}\Delta S_{i,t-j} + \sum_{j=-\underline{j}+1}^{\bar{j}-1}\beta_j \Delta S_{i,\ t-j} + \beta_{\bar{j}}\sum_{j=\bar{j}}^{\infty}\Delta S_{i,t-j} + X'_{i,t}\lambda + \phi d_i + \varphi d_t + u_{i,t}, \tag{4}$$

Moreover, using our assumptions that no department is initially treated and all departments are eventually treated and once treated stay treated, we can use the fact that after

10 There is a rapidly growing literature applying and developing panel event studies. Influential early studies include Jacobson (1993), Autor (2003) and Stevenson and Wolfers (2006).

11 See Schmidheiny and Siegloch (2020) for a discussion of relationship between an infinite distributed lead/lag model and an event-study specification.

cancelling terms in the first and third summations of changes, these summations can be written simply as

$$\sum_{j=-\infty}^{-\underline{j}} \Delta S_{i,t-j} = S_{i,t+\infty} - S_{i,t+\underline{j}-1} = 1 - S_{i,t+\underline{j}-1} \tag{5}$$

and

$$\sum_{j=\bar{j}}^{\infty} \Delta S_{i,t-j} = S_{i,t-\bar{j}} - S_{i,t-\infty-1} = S_{i,t-\bar{j}}. \tag{6}$$

This allows us to simplify Equation (4) as

$$\ln Y_{i,t} = \alpha + \beta_{-\underline{j}}\left(1 - S_{i,t+\underline{j}-1}\right) + \sum_{j=-\underline{j}+1}^{\bar{j}-1} \beta_j \Delta S_{i,\ t-j} + \beta_{\bar{j}} S_{i,t-\bar{j}} + X'_{i,t}\lambda + \phi d_i + \varphi d_t + u_{i,t}. \tag{7}$$

For our empirical implementation, we assume that binning occurs at five leads and five lags, so that $\underline{j} = \bar{j} = 5$. Normalising β_{-1} to zero, our estimation equation becomes[12]:

$$\begin{aligned}\ln Y_{i,t} = \alpha + \beta_{-5}(1 - S_{i,t+4}) + \sum_{j=-4}^{-2} \beta_j \Delta S_{i,\ t-j} \\ + \beta_0 \Delta S_{i,t} + \sum_{j=1}^{4} \beta_j \Delta S_{i,t-j} + \beta_5 S_{i,t-5} + X'_{i,t}\lambda + \phi d_i + \varphi d_t + u_{i,t}.\end{aligned} \tag{8}$$

We will present our results with event graphs that show the proportionate effect of the event of a star arrival from 4 years before the arrival to 4 years afterwards—that is a plot of $\hat{\beta}_{-4}$ to $\hat{\beta}_4$ with associated 95% confidence intervals. The binning variables play important roles as controls, but given their level form we are concerned they may be correlated with other excluded level variables and thus may not be strictly comparable to the leads and lags of the arrival dummies in the estimation of relevant dynamic effects. Statistical inference will be based on robust standard errors clustered at the department level or individual level dependent upon specification to allow for arbitrary forms of serial correlation and heteroscedasticity (Kripfganz, 2016).

2.3. Heterogeneous star arrival effects by arrival cohort

Our baseline model has assumed star arrival effects that are homogenous across the timing of arrivals. An active recent literature has questioned the causal interpretation of estimates of coefficients such as the β_j coefficients in Equation (8) in TWFEs models (Borusyak and Jaravel, 2017; Athey et al., 2018; de Chaisemartin and D'Haultfoeuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021). A major concern is that in settings with variation in the timing of treatments—star arrivals in our case—and heterogeneous effects across different timings, the coefficients on a given

12 Freyaldenhoven et al. (2019) employ a similar specification for various applications but do derive it formally in the paper.

lead or lag can be contaminated by effects from other periods (see, e.g. Sun and Abraham, 2021). Using the approach of Sun and Abraham (2021), we therefore allow specifically for staggered treatments and our dynamic star arrival effects are derived as an appropriately weighted average of 'cohort average treatment effects on the treated' (CATT). This approach has the advantage of being relatively easy to implement in a regression framework and allows straightforward comparisons to our baseline results.

In allowing for staggered star arrival effects, we first define E_i as the year in which an arrival in department i occurs. The appropriate weighted average of cohort effects for a given time relative to the arrival event, j—or interaction-weighted (IW) estimator—is then obtained using a three-step procedure.

First, in place of Equation (8), we run a regression that allows for estimated arrival effects to vary based on the year that the arrival event occurs:

$$\begin{aligned}\ln Y_{i,t} = \sum_{e} \Bigg[& \delta_{e,-5}\Big((\mathbf{1}\{E_i = e\}(1 - S_{i,t+4})\Big) + \sum_{j=-4}^{-2} \delta_{e,j}(\mathbf{1}\{E_i = e\}\Delta S_{i,t-j}) \\ & + \sum_{j=0}^{4} \delta_{e,j}(\mathbf{1}\{E_i = e\}\Delta S_{i,t-j}) + \delta_{e,5}(\mathbf{1}\{E_i = e\}S_{i,t+5})\Bigg] + X_{i,t}^{'}\lambda + \phi d_i + \varphi d_t + u_{i,t},\end{aligned} \tag{9}$$

where $\mathbf{1}\{E_i = e\}$ is an indicator variable that takes the value 1 if the arrival to a department i takes place in year e and 0 otherwise, and $\delta_{e,j}$ is the star arrival effect j years after an arrival that occurs in year e. Our observation window for the analysis is from 1996 to 2017 and there are 19 treated cohorts in our sample after the first cohort (1996) is excluded from estimation since it is always treated across the observation window. Second, we estimate a set of weights $\Pr\{E_i = e | E_i \in [-j, T - j]\}$ that are equal to the sample shares of each cohort for the relevant periods j. And third, to obtain the IW estimator, we take a weighted average of the $\hat{\delta}_{e,j}$ (or $\text{CATT}_{e,j}$) estimates from the first step with the relevant weights from the second step:

$$\beta_j^* = \sum_{e} \left[\hat{\delta}_{e,j} \Pr\{E_i = e | E_i \epsilon [-j,\ T - j]\}\right]. \tag{10}$$

To allow comparison between the homogenous and heterogeneous effects models, we also exclude the 1996 arrival cohort from our homogenous effects model. This exclusion has minimal effects on the results.

3. Data

Our data consist of all publications (and citations to those publications at collection) recorded in Scopus in 27 subject fields for the period 1990–2017 where at least one author's affiliation is recorded as being in Denmark, Ireland or New Zealand. We chose Denmark and Ireland since both these countries are small open economies that have and are engaged in nationally supported formal programmes for the recruitment of star researchers. Then we selected New Zealand which is a small open economy with no record of any formal star recruitment programme. For each collected publication, we record the year, authors, affiliations, subject field[13], citation count to 2019, references, abstracts and acknowledgements, as well as, importantly, two unique identifiers—the EID Scopus publication identifier and the Scopus Author(s) ID. This procedure gives us approximately 1.43 million publications divided over 219,582 unique authors. For each author, we also access their catalogue of publications back to 1990, including publications that took place prior to joining the department in Demark, Ireland or New Zealand. We identify a 'department' with Scopus-defined subject field at a given institution so our departments need not coincide with the departmental structure at particular institutions. However, in some institutions, there was no activity in the subject area and these 'departments' were dropped from out dataset where there was not at least one cited publication in each sample year. Our data are spread across 457 remaining departments, comprising of 153 departments in Denmark, 141 in Ireland and 163 in New Zealand.[14]

3.1. Identification of stars

The first step is the identification of star scientists. We define a star scientist as a scientist who is at or above the 95th percentile of scientists in the cumulative distribution of citations received since 1990 for their subject field in any given year. To allow reasonable time for citations to publications to accumulate in our data, we only identify stars from 1996 onwards in our analysis. We allocate a star to a subject area (department) based on where they have the largest number of their publications. Where stars have an equal plurality of publications across multiple subject areas, then the star is assigned to multiple departments. It is also important to note that as our data only relate to scientists in Denmark, Ireland and New Zealand, our identification of stardom is relative to scientists in the subject field for these countries only. However, this procedure allows us to identify high-performing scientists for this reference group.

A natural concern is that this relative measure may not identify true *international* stars. As part of our robustness analysis, we therefore also explore the effects of augmenting our identification of star arrivals with an additional (variable) filter that requires that the star's Scopus Field-Weighted Citation Impact (FWCI) is above a threshold. The FWCI weights a scientist's citation score by the *international average* for the field. We compare the size of the star-arrival effect after 4 years based on different thresholds for this index. Although setting a higher threshold reduces the number of identified star arrivals and thus the precision of the estimated star arrival effects, an examination of how the size of the estimated effects varies with the threshold allows us to see how the effect size varies as we become more demanding in terms of this measure of international stardom.

3.2. Identification of star arrivals

We identify a star arrival at a department as a star that had not previously published with an affiliation to the department now recording an affiliation to that department. We record the year of the first such star arrival (if any) to each department. For our empirical analysis, we restrict identified star arrivals to those who publish with an affiliation to the

13 The 27 subject fields are: Agricultural and Biological Sciences, Arts and Humanities, Biochemistry, Business, Chemical Engineering, Chemistry, Computer Science, Decision Science, Dentistry, Earth and Planetary Sciences, Economics, Energy, Engineering, Environmental Science, Health Professions, Immunology and Microbiology, Multidisciplinary, Materials Science, Mathematics, Medicine, Neuroscience, Nursing, Pharmacology, Physics and Astronomy, Psychology, Social Sciences and Veterinary.

14 See Online Appendix Table B1 for additional detail.

department for at least the next 4 years, or for those that publish in 2015–2016, they must also be observed to publish at the same affiliation at the end of our observation window in 2017.[15] Overall, we identify 167 star arrivals over the period 1996–2017 in total using this procedure. As our empirical implementation uses only first star arrivals to a department, we record 81 such first star arrivals, excluding the 1996 cohort, across the three countries—38 in Denmark, 19 in Ireland and 24 in New Zealand.[16]

3.3. Output measures

Our output measures are at the department-year or individual incumbent level depending on the specification. We consider three output measures: (i) total count of publications in the department/incumbent year; (ii) field-normalised total citations (the sum across department/incumbent publications of citations divided by the average citations to a publication for that field in that year)[17] and (iii) a scaled Perry–Reny Euclidean Index (Perry and Reny, 2016) that places greater weight on highly cited publications than measure (ii).[18]

More formally, the three output measures are defined as follows:

$$\text{Publications}: \; Y^{P}_{i,t} = P_{i,t};$$

$$\text{Field Normalized Total Citations}: \; Y^{FNTC}_{i,t} = \sum_{p_{i,t}=1}^{P_{i,t}} \frac{c_{p_{i,t}}}{\bar{c}_{s,t}};$$

$$\text{Perry} - \text{Reny Euclidean Index}: \; Y^{E}_{i,t} = \left(\sum_{p_{i,t}=1}^{P_{i,t}} \left(\frac{c_{p_{i,t}}}{\bar{c}_{s,t}} \right)^{2} \right)^{\frac{1}{2}};$$

where $P_{i,t}$ is the total number of publications in department i or incumbent i in year t, $c_{p_{i,t}}$ is the total subsequent citations (or 'forward' citations recorded at 2019) to a publication $p_{i,t}$ that occurs in department i in year t and $\bar{c}_{s,t}$ is the average citations to a publication in the relevant subject field, s, for publications that occur in year t.

These output measures provide the dependent variables for our regressions. Our estimation period is selected as 1996–2017, which also allows for the inclusion of the chosen number of leads and lags ($\underline{j} = \bar{j} = 5$) for all post-1996 star arrivals. Descriptive statistics for the department-level output measures are shown in Table 1 (Columns 1–3). Table 1 Columns 4–7 split departments into 'never-treated' and 'ever-treated' subgroups where the treatment refers to the first arrival of a star. The two groups are found to be different with the treated departments having larger outcome values on all dimensions. In addition to controlling for time-invariant department characteristics using department-fixed effects, we also control for department-specific linear time trends and a time-varying university-level control (excluding

15 In terms of the same first star arrival at multiple departments, Columns 5 and 6 in Online Appendix Table B2 present such occurrences. Specifically, there are six instances of the same individual star being allocated to two separate subject fields in this analysis. Additionally, there are three instances where more than one star first arrived at the same department in a given year (see Column 4, Online Appendix Table B2); however, it is worth highlighting that our estimation results are robust to both their inclusion and exclusion as treated departments.

16 Online Appendix Figure B1 shows the distribution across time of first star arrivals in aggregate and for each of the three countries. The year with the highest number of departments with a first star arrival is 1998 with nine stars first arriving across the three countries.

17 See Radicchi et al. (2008).

18 See Perry and Reny (2016).

Table 1. Descriptive statistics

	Total sample			Never-treated and ever-treated			
	1996–2017			1997–2017			
	Mean (1)	SD (2)	*N* (3)	Never-treated (4)	Ever-treated (5)	*p*-value (6)	% Treated (7)
Unit of analysis: department-year							
Publications (log)	3.603	1.346	10,054	3.346	4.794	<0.001	17.7
Field normalised (log)	3.406	1.582	10,040	3.108	4.787	<0.001	17.7
Euclidean (log)	3.487	1.764	10,041	3.160	5.000	<0.001	17.7
Incumbents' Publications (log)	1.258	1.264	6781	0.962	2.387	<0.001	20.7
Incumbents' Field normalised (log)	0.879	1.683	6720	0.522	2.226	<0.001	20.9
Incumbents' Euclidean (log)	0.957	1.872	6742	0.581	2.382	<0.001	20.9
University output (log)	7.079	1.137	10,054	6.968	7.588	<0.001	17.7

Notes: The percentage of observations for 'ever-treated' groups is reported in Column (7) (in % of total observations 1996–2017). Column (6) presents the *p*-value for the null hypothesis that there is no significant difference between the two groups.

the focal department). Pre-arrival trends are examined to check for violations of the parallel trends assumption in our estimated specifications.

3.4. CEM Procedure for Incumbent Sample

For the matched individual-level analysis, we use the following CEM procedure to identify appropriate matches. First, we identify all incumbents in the department in the year, t, that the star arrives. (Incumbents are required to be present 4 years before the star's arrival and 4 years after the arrival.) For each identified incumbent in a star-receiving department, we identify an appropriate match based on covariates measured in $t-1$. The following covariates are used to identify the match: scientist career age (i.e. years since first publication); citations received by the scientist in $t-1$; cumulative citations received by the scientist up to $t-1$; subject field; country and year. We utilise relatively coarse bins for career age (13 bins) and for cumulative citations (26 bins), but require an exact match for citations received in $t-1$, subject field, country and year. Our matching procedure successfully matches 16,730 incumbent scientists. The resulting panel is unbalanced as all scientists will not be present in the department over the entire 22-year sample period. As a robustness check, we repeat the analysis for a balanced panel of 4696 incumbents whom were present at a star-receiving department over the entire sample period. Table 2 provides summary statistics for the matching process as applied to both the unbalanced and balanced panels. Overall, the application of CEM is successful in identifying good matches for our sample of incumbents.

4. Results

4.1. Star arrival effects on department output

Our star arrival results for under both the homogenous and heterogeneous arrival effects by cohort are shown in Table 3. The department output variable is field-normalised total

Table 2. Summary statistics: control and treated subfields at baseline (k-to-k matched)

Variable	Control	Treated	Diff. in mean	*p*-value
Panel A: Unbalanced panel on 16,730 matched authors, with 8365 in each group				
Citation received	34.742	35.221	−0.478	0.6988
Career age (Bins)	2.209	2.209	0.000	1.000
Cumulated citation received (Bins)	3.455	3.455	0.000	1.000
Career age (Num.)	8.907	8.876	0 .031	0.7807
Cumulated citation received (Num.)	350.902	345.197	5.704	0.7149
Panel B: Balanced panel on 4696 matched authors, 2348 in each group				
Citation received	52.029	53.689	−1.659	0.5997
Career age (Bins)	2.892	2.892	0.000	1.000
Cumulated citation received (Bins)	5.621	5.621	0.000	1.000
Career age (Num.)	12.451	12.408	0.042	0.8537
Cumulated citation received (Num.)	670.287	675.770	−5.482	0.8916

Notes: We report the *t*-test for mean differences between the treated and control groups at 1 year before the star arrival for these time-varying covariates. For career age, we create 13 bins: less than 5 years; between 5 and 15 years; between 15 and 20 years; between 20 and 25 years; between 25 and 30 years; between 30 and 45 years; between 45 and 50 years; between 50 and 55 years; between 55 and 60 years and over 60 years of career age. Similarly, we coarsen the distribution of cumulated citations at baseline 26 mutually exclusive bins; between 1 and 100 citations; between 200 and 300 citations; between 300 and 400 citations; between 400 and 500 citations; . . . ; between 5000 and 10,000 citations and larger than 10,000 citations.

citations to the output published in year *t*. (We compare the results with other output measures as part of robustness tests in Section 5.) We present two specifications to allow the reader to compare the homogenous and heterogeneous models with and without the inclusion of control variables. The control variables are a set of department-specific time trends and university-level output excluding the focal department.

Column 1 of Table 3 shows the results for the homogenous model without controls. The star arrival effects are also shown in the top left panel of Figure 1. While the results show some visual evidence of a pre-trend, we see evidence of economically and statistically significant star arrival effects, with a star arrival being associated with a 12.2 log points (12.9%) increase in output after 4 years. Column 2 of Table 3 and the top right panel of Figure 1 show the results with the controls included. With the inclusion of the controls, we do not see evidence of a pre-trend and the estimated arrival effects are larger, reaching 18.9 log points (20.8%) after 4 years. We treat this specification as our baseline in conducting our robustness tests below. These results are also robust to the separate inclusion of the set of department-specific time trends and the university-level controls.

As our parallel trends assumption is critical to the identification of a casual effect of star arrivals, it is important to validate the assumption for our data. A common method to assess the credibility of the parallel trends assumption is to test for significant pre-trends in the event study setting; however, a recent body of literature has raised important concerns about this test for pre-trends. In general, it warns that the test may have low power to detect meaningful violations of parallel trends (Kahn-Lang and Lang, 2018; Freyaldenhoven et al., 2019; Bilinski and Hatfield, 2020; Roth, 2021). Using an approach proposed by Roth (2021), we present diagnostics to evaluate whether the limitations of pre-trends testing are an important concern for our analysis.

Table 3. Department level—dynamic star arrival effects for field normalised total citations

	Homogenous arrival effects		Heterogeneous arrival effects	
	No controls (1)	With controls (2)	No controls (3)	With controls (4)
$1 - S_{i,t+4}$ (binned lead)	−0.176	−0.128	−0.142	−0.118
	(0.118)	(0.106)	(0.110)	(0.108)
$\Delta S_{i,t+4}$	−0.0718	−0.00404	−0.047	−0.092
	(0.0808)	(0.0881)	(0.085)	(0.090)
$\Delta S_{i,t+3}$	−0.0739	−0.0426	−0.055	−0.082
	(0.0761)	(0.0795)	(0.072)	(0.095)
$\Delta S_{i,t+2}$	−0.0736	−0.0295	−0.046	−0.042
	(0.0644)	(0.0648)	(0.055)	(0.069)
$\Delta S_{i,t}$	0.0507	0.0634	0.044	0.067
	(0.0619)	(0.0641)	(0.054)	(0.058)
$\Delta S_{i,t-1}$	0.115*	0.138**	0.101**	0.163***
	(0.0589)	(0.0637)	(0.050)	(0.062)
$\Delta S_{i,t-2}$	0.111*	0.153**	0.104*	0.197***
	(0.0638)	(0.0679)	(0.054)	(0.068)
$\Delta S_{i,t-3}$	0.113	0.166**	0.104*	0.181**
	(0.0722)	(0.0777)	(0.060)	(0.071)
$\Delta S_{i,t-4}$	0.122	0.189**	0.112*	0.220***
	(0.0743)	(0.0823)	(0.067)	(0.076)
$S_{i,t-5}$ (binned lag)	−0.0360	0.226**	−0.013	0.241**
	(0.0934)	(0.0965)	(0.089)	(0.098)
Log university output		0.857***		0.838***
(excl. dept.)		(0.0701)		(0.071)
R-squared	0.4337	0.6561	–	–
Pre-test against hypothesised trend (Roth, 2021)				
Power	0.50	0.50	–	–
Hypothesised trend	0.05	0.05	–	–
Bayes factor	0.56	0.56	–	–
Likelihood ratio	0.40	0.09	–	–
Department FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Department time trend	No	Yes	No	Yes
Observations	10,040	10,040	10,040	10,040

Notes: Event window: 1997–2017 (1996 cohort dropped). Estimation window: 1996–2017. Dependent variable is field-normalised total citations. This table reports the estimates based on the model specification in Equation (8) in Columns 1 and 2. The dependent variable always excludes the star's own output. Robust standard errors are clustered at department-level and reported in parentheses. Columns 3 and 4 are the estimates based on the Sun and Abraham (2021) method (Equation 10). *, ** and *** represent significance levels at the 10%, 5% and 1%, respectively.

Specifically, we implement a power calculations analysis to construct a hypothesised linear violation of parallel trends and then compare the likelihood of the estimated pre-trend coefficients under both the hypothesised trend and parallel trends. The results from this analysis for a relevant violation of parallel trends are also reported in Table 3. For our analysis, the slope of the hypothesised trend is constructed so that a pre-trends test based on the statistical significance of the pre-trend coefficients would detect a

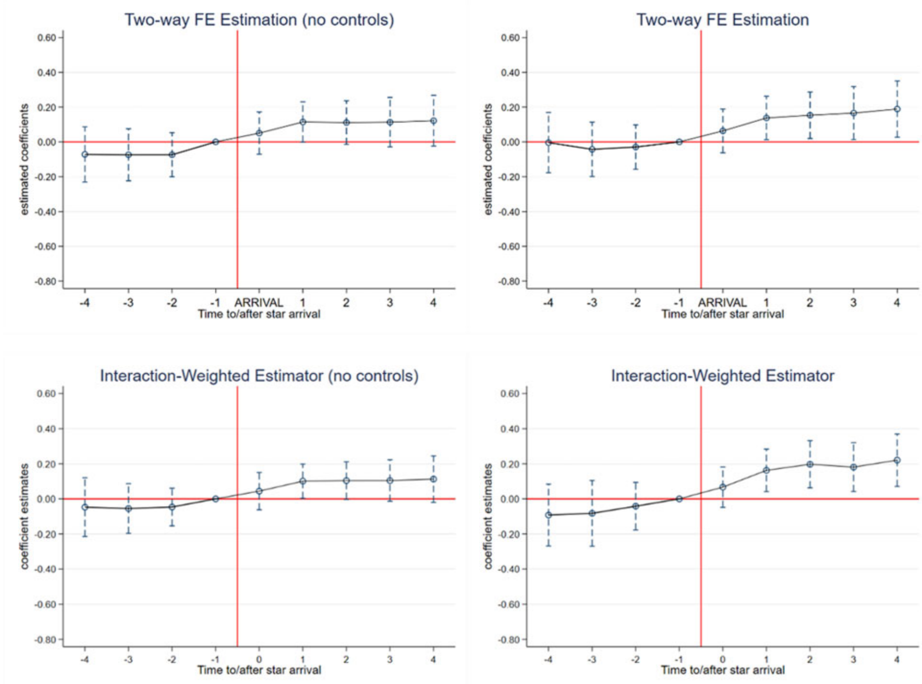

Figure 1. Event study model with the homogeneous (top) versus heterogeneous (bottom) arrival effects across cohorts (department level).

Notes: This figure plots the dynamic effect of a first star arrival with 95% confidence intervals. The event and observation window is from 1996 to 2017 (*the 1996 cohort is omitted*). The dependent variable is the field-normalised total citations and excludes the output of the star arrival. First star arrivals are restricted to stars staying in the department for 4 years or more. The omitted category is 1 year before the first star arrival. University-level controls (excluding the focal department) and department time-specific trend are included in the model specification on which the right-hand plot is based.

significant pre-trend 50% of the time. This slope estimate is 0.05 both with and without the inclusion of controls. The likelihood ratio which reports the ratio of the likelihood of the observed coefficients under the hypothesised trend relative to under parallel trends suggests that, for both specifications, we are much more likely to observe our estimated coefficients under parallel trends rather than the hypothesised trend with 50% power in detecting a pre-trend, which gives us confidence in the parallel trends assumption. It is also worth highlighting that the likelihood ratio for the model that identifies a star arrival based on the baseline criteria decreases from 0.40 to 0.09 when the University level controls and the department time-specific trend are included in the specification, underlining the importance of including these controls in our preferred specification.

Column 3 of Table 3 (and the bottom left panel of Figure 1) records the estimated star arrival effects using the Sun and Abraham methodology without the controls. We now see little evidence of a pre-trend and the estimated star arrival effects are quantitatively similar to the homogenous effects model results, with an 11.2 log point (11.8%) increase in department output 4 years after the arrival of the star. As noted above, we provide for comparison purposes the results of the heterogeneous effects model with controls in Column 4 (and bottom right panel of Figure 1). This specification again shows little evidence of a pre-trend and the largest star arrival effects of the specification shown, with a star-arrival effect after 4 years of 22.0 log points (24.6%). Overall, the results appear reasonably robust to the choice of specification with star arrival effects 4 years after the arrival of the star of between approximately 11 and 22 log points (or 11.8% and 24.6%).

One possible concern with our method of star identification is that it applies a relative standard—that is where the scientist fall in the distribution of total (eventual) citations to their published work in their respective fields across the three small open economies. It is reasonable to argue that star recruitment policies aim to recruit true international star, where the relevant distribution would be the full international distribution in that field. As an additional robustness test we therefore add an additional threshold (or filter) to our identification method based on the Scopus SciVal reported FWCI. For each scientist, Scopus measures their citations relative to the international average for the field, so that a scientist with a FWCI of 1 would exactly match the international average.

Figure 2(a) shows the evolution of the star arrival effect measured 4 years post arrival based on different thresholds for the FWCI. The figure shows that the number of identified stars does indeed fall as we move to higher thresholds, indicating that a number of our relative stars appear less outstanding from the truly international perspective. While care must be taken in interpreting the results because of the falling number of identified stars as we adopt a stricter threshold, it is notable that the estimated star arrival effect is smaller at the higher thresholds, with an apparent step drop in the estimated effect at a FWCI of $\geq$3. These results suggest that the largest star arrival effects may not occur for the most accomplished international stars, possibly because such stars embed less deeply in local networks. This result would also appear consistent with the result reported below that star arrivals effects are generally larger for earlier career stars (career age $\leq$20), whom are also likely to rank lower in the international distribution. In terms of policy design, this underlines the importance of identifying stars that have a strong incentive to embed in local networks, which may depend, *inter alia*, on their career horizons and the strength of their (possibly competing) existing international networks.

For our star arrivals, one interesting split is between stars that arrive relatively early in their careers and those that arrive later. We measure career age as the number of years since the scientist recorded their first publication and split arriving stars into ‘earlier’ career age stars (career age $\leq$20) and ‘later’ career age stars (career age >20). Due to space limitations, the results are shown in Online Appendix C. Online Appendix Figure C1 indicates that the star arrival effect on department productivity is large for the earlier career arrivals. As discussed further in Section 5, this may reflect a greater willingness for scientists with longer career horizons at their new institution to embed into local scientist networks leading to more pronounced catalytic effects on department performance.

One concern is that the results for the aggregate of the three countries may hide different dynamic patterns at the country level; we therefore present country-specific results in Online Appendix D. Online Appendix Figure D1 shows the results of our baseline regression with inclusion of the controls for the three countries separately. The broad pattern for Denmark mirrors those found for the aggregate results. However, the patterns for Ireland and New Zealand are notably different with no evidence of star arrival effects for Ireland and a notable positive pre-trend visible and at best limited evidence of post-arrival star effects for New Zealand. While care must be taken in over interpreting the country-specific results due to the relatively small number of treatments, the differences across countries suggest the importance of country- and institution-specific policy design.

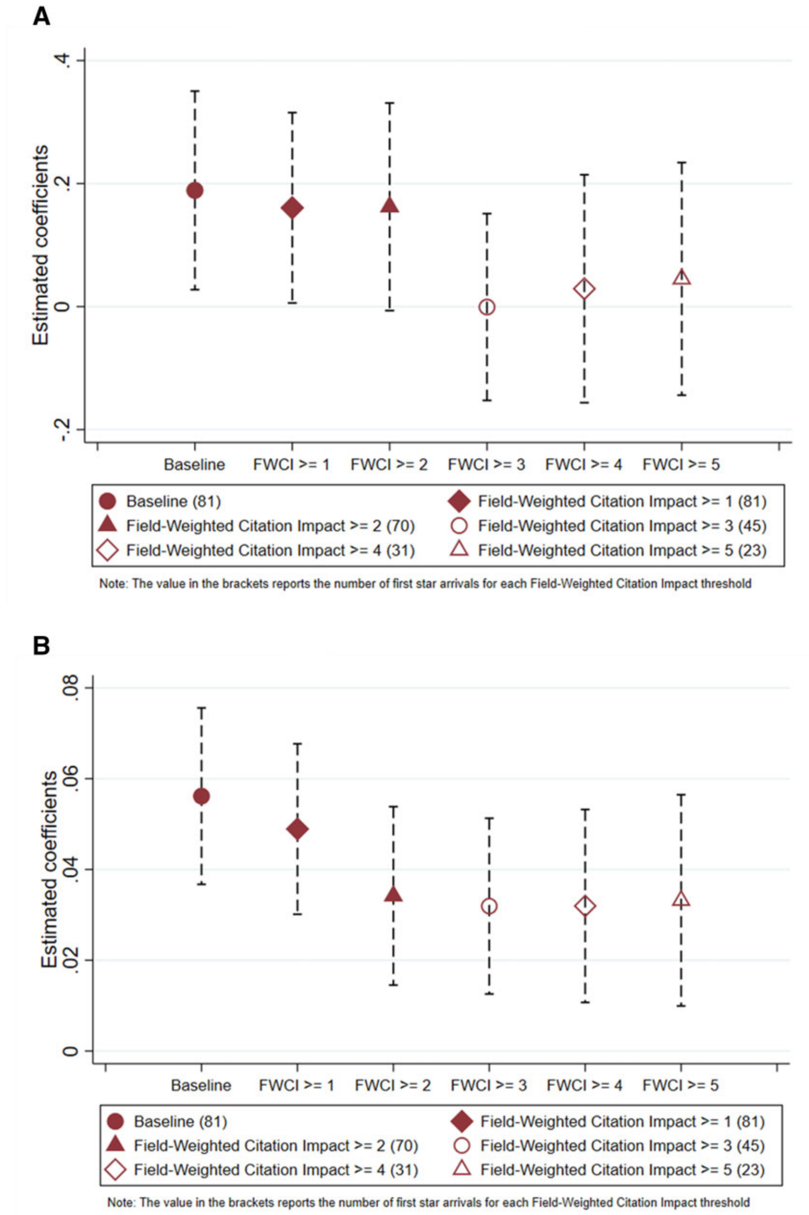

Figure 2. Size of estimated star arrival effect 4 years post star arrival based on alternative thresholds for the FWCI. (**a**) Departmental level. (**b**) Individual level.

Notes: The graph shows the star arrival effects on the departmental level (Panel A) and individual level (panel B) for the FWCI, which compares the entity's number of citation received for a publication compared with the world average. The FWCI is calculated as an average 4 years prior to the star arrival year (1996 onwards, given Scival's limitation) from Scival metrics. While the number of star arrivals are the same between the baseline and the group FWCI ≥ 1 (i.e. 81 star arrivals), the groups are marginally different since the threshold restriction (FWCI ≥ 1) omits a small number of first star arrivals identified under the baseline criteria, yet includes a number of star arrivals otherwise not identified under the baseline.

However, as we report below, at the individual incumbent level, there is a more consistent pattern of star arrival effects across the three countries.

4.2. Star arrival effects on incumbent output

Overall, our department-level results point to a significant effect of star arrivals on subsequent department productivity performance. Moreover, the event-study results appear robust to a misspecification of the pre-trend and allowance for heterogeneity in the arrival effect across arrival cohorts. However, there remains a concern that observed arrivals are not independent of productivity developments that occur contemporaneous with or subsequent to the star arrival. For the reasons noted earlier, we therefore augment our department-level analysis with an individual-level analysis of the effect of star arrivals on matched incumbents. We require incumbents to be present in the department 4 years before the arrival of the star and also present in the 4 years following the arrival. The use of matching allows us to give firmer causal interpretations to the observed star arrival effects.

Table 4 records the results for the individual-level incumbent regressions. Columns (1) and (2) show the results for the balanced and unbalanced panels under the assumption of homogenous effects by arrival cohorts. In discussing the results, we concentrate on the unbalanced panel as the long-stay incumbents in the balanced panel are a selected subgroup that are more likely to be older and individuals who have never moved institutions.

Figure 3 shows the event-study graph for a sample of 16,730 matched incumbents in the unbalanced panel, where the dependent variable is an individual's field-normalised total citations to their publications in a given year. Visual inspection and tests for significance of the pre-trend coefficients reveal no evidence of a pre-trend and star arrival effects are statistically significant after 1 year and grow over time. The star arrival effect is 5.6 log points (5.8%) after 4 years.

Paralleling our approach to the department-level analysis, we also apply the Sun and Abraham (2021) methodology to allow for heterogeneous effects by arrival cohort. Column (4) of Table 4 shows the results for the unbalanced panel. For ease of visual inspection, Figure 3 (on the right) graphs the results for the unbalanced panel under the assumption of heterogeneous effects by arrival cohort. We again find our results are robust to the allowance for such effects.

Table 4 also reports the Roth (2021) test for the sensitivity to the violation of the parallel trends assumption. For the unbalanced panel with homogenous effects, the slope of the hypothesised trend required for finding a significant pre-trend 50% of the time is 0.005. Not surprisingly given the absence of visual evidence of a pre-trend, the likelihood ratio of the observed pre-trend coefficients under the hypothesised trend relative to under parallel trends is low at 0.04. This gives us confidence that the parallel trends assumption is supported in our individual-level data after matching.

As we did for the department-level results, Figure 2(b) shows the implications of applying an increasingly strict filter to our identification of a star based on the SciVal FWCI measure. The results across different thresholds are similar to what we observe at the departmental level, with evidence of smaller star arrival effects as we apply an increasingly stringent threshold for being identified as a true international star.

Online Appendix Figure C2 also confirms the qualitative picture from the department-level results that shows that the star-arrival effect is larger for stars arriving earlier in the careers, which again is suggestive of early-career arriving stars to embed in local networks when their horizons at the institution are longer.

Table 4. Individual level—dynamic star arrival effects for field normalised total citations

	Homogenous arrival effects		Heterogeneous arrival effects	
	Balanced panel (1)	Unbalanced panel (2)	Balanced panel (3)	Unbalanced panel (4)
$1 - S_{i,t+4}$ (binned lead)	0.003	0.007	−0.020	−0.008
	(0.020)	(0.009)	(0.027)	(0.011)
$\Delta S_{i,t+4}$	−0.007	0.008	−0.009	0.006
	(0.019)	(0.009)	(0.021)	(0.010)
$\Delta S_{i,t+3}$	−0.006	0.008	−0.006	0.009
	(0.018)	(0.008)	(0.019)	(0.009)
$\Delta S_{i,t+2}$	0.021	0.008	−0.014	0.005
	(0.016)	(0.007)	(0.017)	(0.008)
$\Delta S_{i,t}$	0.043***	0.014**	0.043***	0.015**
	(0.015)	(0.007)	(0.015)	(0.008)
$\Delta S_{i,t-1}$	0.051***	0.029***	0.053***	0.035***
	(0.016)	(0.007)	(0.017)	(0.008)
$\Delta S_{i,t-2}$	0.056***	0.042***	0.065***	0.050***
	(0.017)	(0.008)	(0.018)	(0.009)
$\Delta S_{i,t-3}$	0.064***	0.045***	0.073***	0.052***
	(0.018)	(0.009)	(0.018)	(0.009)
$\Delta S_{i,t-4}$	0.066***	0.056***	0.075***	0.063***
	(0.018)	(0.009)	(0.019)	(0.010)
$S_{i,t-5}$ (binned lag)	0.108***	0.070***	0.119***	0.080***
	(0.018)	(0.009)	(0.019)	(0.011)
R-squared	0.0473	0.0347	–	–
Pre-test against hypothesised trend (Roth, 2021)				
Power	0.50	0.50	–	–
Hypothesised trend	0.01	0.005	–	–
Bayes factor	0.58	0.57	–	–
Likelihood ratio	0.57	0.04	–	–
Department FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Department time trend	No	No	No	No
Observations	98,978	260,629	98,978	260,629

Notes: Event window: 1997–2017 (*1996 cohort dropped*). Estimation window: 1996–2017. Dependent variable is field-normalised total citations. This table reports the estimates based on the model specification in Equation (8) in Columns 1 and 2. The dependent variable always excludes the star's own output. Robust standard errors are clustered at individual-level and reported in parentheses. Columns 3 and 4 are the estimates based on the Sun and Abraham (2021) method (Equation 10). *, ** and *** represent significance levels at the 10%, 5% and 1%, respectively.

Turing to the country-specific results, Online Appendix Figure D2 shows a more consistent star arrival effect on incumbents than we found for the department level analysis. While standard errors are of course larger than for our three-country sample, we find evidence of a statistically significant star arrival after 3 years for each country and broadly similar evolutions over time. The size of the effect is also roughly similar across the three countries (≈5%) and consistent with the combined sample.

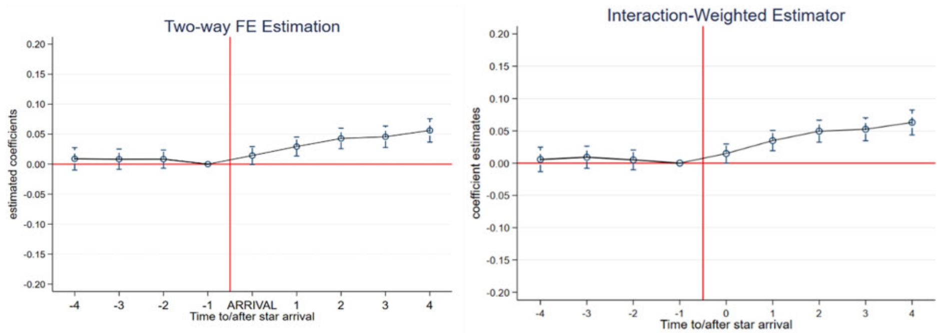

Figure 3. Event study model for matched incumbents (unbalanced panel).
Notes: This figure plots the dynamic effect of a first star arrival with 95% confidence intervals based on the matched sample, with 16,351 (379 treated author in 1996 cohort removed) authors. It consists of 260,629 observations. The event and observation window is from 1996 to 2017. The top panel shows the homogenous effects results; the bottom panel shows the heterogeneous effects results based on Sun and Abraham (2021). The dependent variable is the field-normalised total citations in logarithm (with 1 added across all observations). The omitted category is 1 year before the first star arrival. Robust standard errors are clustered by author.

We can further increase confidence that we are identifying a causal effect of star arrivals if there is a clear mechanism through which arrivals are affecting individual productivity. One plausible mechanism is through opportunities to co-author with the star. A co-authorship relationship should indicate that the work of the incumbent and star are related, and also collaboration with the star should directly boost incumbent productivity. As expected, Figure 4 shows a large additional effect on incumbent productivity if the incumbent develops a co-authorship relationship with the star.

To strengthen our confidence in a causal interpretation of the star-arrival effect, a further hypothesis is that incumbent scientists who are working in areas related to the star should experience larger beneficial productivity impacts from the star's arrival. Although there are many possible indicators of relatedness, an obvious possibility is to identify incumbents that have cited the star at some point in the sample. Using this indicator of 'relatedness', Figure 5 compares the results for unrelated and related incumbents. The hypothesis that related incumbents benefit more from an arriving star is clearly supported, with related incumbents increasing their productivity after 4 years by approximately 40%.

5. Robustness

In this section, we conduct several robustness tests on our baseline model. We limit our reporting to the department-level analysis due to space constraints, although we also find our incumbent-level analysis is robust to these alternative specifications. Specifically, we compare our results to the estimates from specifications with alternative output measures and alternative star identification methods. Further to this, we conduct two robustness tests in the Online Appendix that accompanies this analysis: we explore a restricted specification

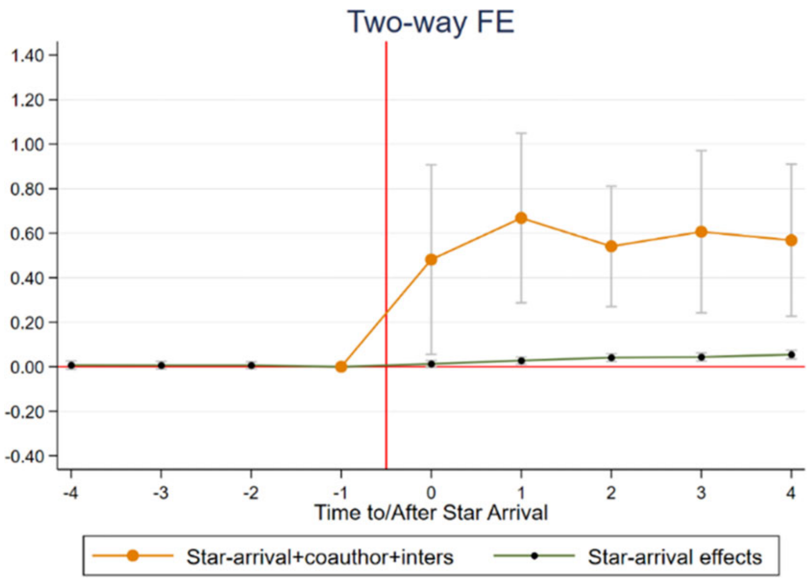

Figure 4. Event study model for matched incumbents with star co-authoring effects (binary). *Notes*: This figure plots the dynamic effect of a first star arrival with 95% confidence intervals based on the matched sample, with 16,351 (379 treated author in 1996 cohort removed) authors. It consists of 260,629 observations. The event and observation window is from 1996 to 2017. The dependent variable is the field-normalised total citations in logarithm (with 1 added across all observations). The omitted category is 1 year before the first star arrival. Robust standard errors are clustered by author.

that imposes a piecewise linear structure on the dynamic effects[19] and we test for miss-specified dynamics[20]. Overall, the comparisons outlined below and in the Appendix indicate that our results are generally robust across the different measures.

5.1. Robustness to alternative outcome measures

Our baseline regression uses department-level field-normalised total citations as the dependent variable. This measure can be viewed as a quality-adjusted publications measure of department output, where a publication with the average number of citations for the field gets a weight of 1, a publication with twice the average gets a weight of 2, etc. As a first test of the robustness of our results, we examine the sensitivity of these results for the baseline specification of the model to alternative output measures: raw publication counts and the scaled Euclidean Index due to Perry and Reny (2016) that puts greater weight on highly cited publications relative to the field-normalised total citations measure. These measures therefore provide reasonable bounds in terms of the relative weighting of quantity versus quality.

19 Under specific conditions, the efficiency of our estimates can be improved by imposing appropriate restrictions on Equation (8)—see Deryugina (2017). Online Appendix E describes our restricted specification and presents the related results.

20 Online Appendix F describes our robustness test for miss-specified dynamics in Equation (8) and reports the results.

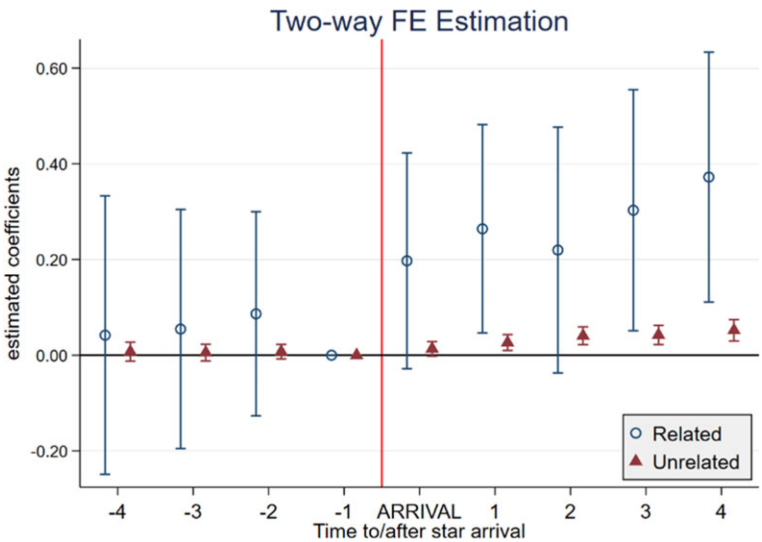

Figure 5. Related versus unrelated.
Notes: This figure plots the dynamic effect of a first star arrival with 95% confidence intervals based on the matched sample. The panel consists of 32 authors related to the star arrival (estimated 125,698 observations) and other unrelated authors with 259,916 estimated observations. The dependent variable is the field-normalised total citations in logarithm (with 1 added across all observations). The omitted category is 1 year before the first star arrival. Robust standard errors are clustered by author.

The results for the alternative measures are shown in Figure 6. The middle panel (b) repeats our baseline results (with controls) from Figure 1 for ease of comparison. The top panel (a) shows the results with raw publication counts as the dependent variable. The results are broadly similar to the baseline but with a steadier rise post arrival and a smaller positive impact on department output in the arrival year and at each lag. The bottom panel (c) shows the results for the Euclidean Index measure. Again, the results are broadly similar to the baseline with a similar dynamic pattern and somewhat larger post-arrival effects. At the fourth lag, the star arrival effect is 14.4 log points (15.4%) for raw publications, 18.9 log points (20.8%) for field-normalised total citations and 20.9 log points (23.2%) for the Euclidean Index. Thus, our results are broadly robust to the choice of output measure but with evidence of greater star-arrival impacts the more weight that is put on quality (as measured by subsequent citations to a publication) compared with publication quantity.

5.2. Robustness to alternative star identification methods

For our baseline results, we identified a star arrival as the arrival of a scientist in the top 5% of the relevant cumulative distribution of total citations to their work measured at the time of arrival. Figure 7 provides a comparison to our baseline results (middle panel) to a more permissive definition of a star arrival (top 10%) and a stricter definition of a star arrival (top 1%). Of course, the number of star arrivals will fall as we adopt a progressively stricter definition, which impacts the standard errors.

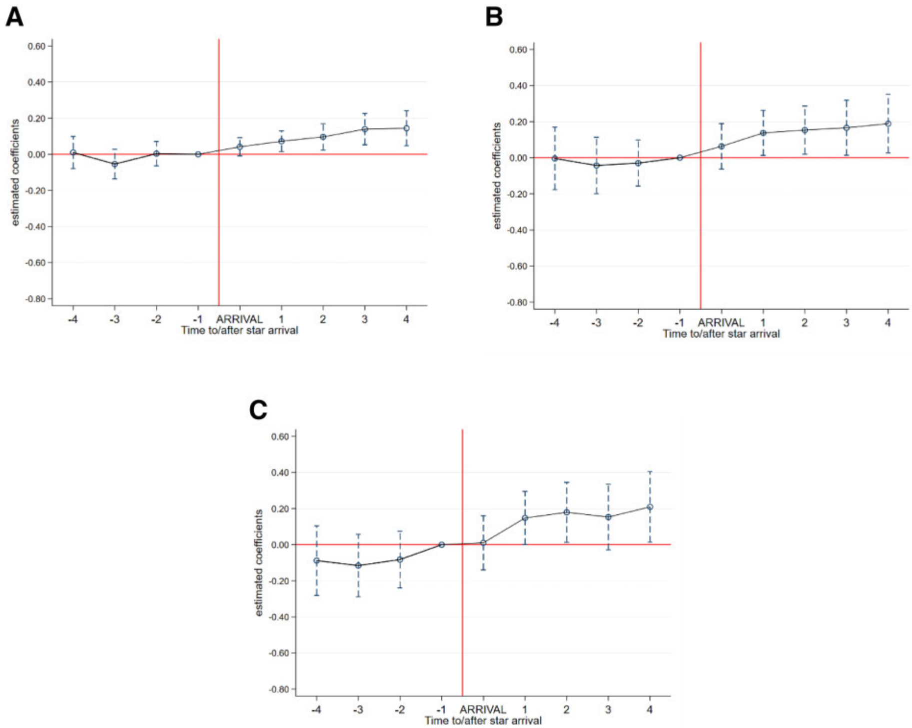

Figure 6. Event study model with homogenous arrival effects across cohorts for alternative outcome measures. (**a**) Publications. (**b**) Field normalised total citations. (**c**) Euclidean index.
Notes: This figure plots the dynamic effect of a first star arrival with 95% confidence intervals and assuming homogenous arrival effects across cohorts based on the specification in Equation (8) for alternative outcome measures. The event and observation window is from 1996 to 2017 (*the 1996 cohort is omitted*). Each alternative outcome measure excludes the output of the star arrival. First star arrivals are restricted to stars staying in the department for 4 years or more. The omitted category is 1 year before the first star arrival. University-level controls (excluding the focal department) and department time-specific trend are included in the model specification. There are 10,054 observations in Publications measure and 1782 are treated; 10,040 observations in field normalised total citations measure and 1782 are treated and 10,041 observations in Euclidean Index measure and 1782 are treated.

The results are again broadly similar across the alternatives, although the estimated star arrival effects are somewhat smaller for both the top 10% definition and the top 1% definition compared with our baseline. For example, 4 years after arrival, the impact on department output is 18.7 log points (10% definition), 18.8 log points (5% definition) and 16.4 log points (1% definition). The wider confidence intervals reflecting the smaller number of arrivals is also apparent in the bottom panel.

6. Conclusions

Our various event analyses—where the event is the arrival of a star—are supportive of the hypothesis that star recruitments have positive and sustained effect on productivity. Although formal tests support the parallel-trends assumption at the department level, we

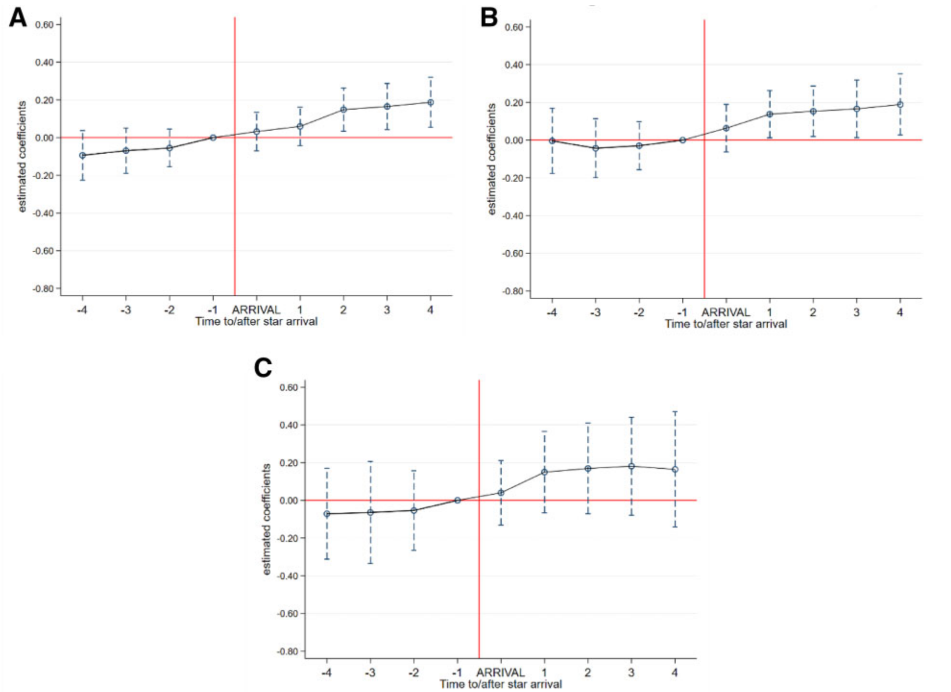

Figure 7. Event study model with homogenous arrival effects across cohorts for alternative star definitions. (**a**) Stars = Top 10% of cumulative distribution (139 arrivals). (**b**) Stars = Top 5% of cumulative distribution (81 arrivals). (**c**) Stars = Top 1% of cumulative distribution (29 arrivals).
Notes: This figure plots the dynamic effect of a first star arrival with 95% confidence intervals and assuming homogenous arrival effects across cohorts based on the specification in Equation (8) for alternative star definitions. The event and observation window is from 1996 to 2017 (*the 1996 cohort is omitted*). The dependent variable is the field-normalised total citations and excludes the output of the star arrival. First star arrivals are restricted to stars staying in the department for 4 years or more. The omitted category is 1 year before the first star arrival. University-level controls (excluding the focal department) and department time-specific trend are included in the model specification.

have been cautious in inferring causality at that unit of analysis. However, the employment of matching at the individual incumbent unit of analysis, combined with strong support for the parallel trends assumption, gives us confidence in offering a causal interpretation of the observed star arrival effects. This interpretation is further strengthened by the finding of the expected larger effects when certain intermediating effects—such as the formation of local collaborations with the star—are present. Moreover, both sets of results are robust to explicitly allowing for heterogeneous effects by arrival cohort.

These findings have potentially important implications for the adoption and design of star recruitment policies. In general, we hypothesise that the beneficial effects of star arrivals depend on stars embedding in local networks. The results indicate that it is not necessarily the most accomplished stars that have the largest local impacts. In particular, we find that stars arriving relatively early in their careers have a larger positive effect on both departments and individual incumbents, possibly reflecting the effects of longer horizons and less developed existing networks on the willingness to make the investments necessary to embed in local networks.

While relatively small numbers of arrivals at the individual-country level mean that country-specific results need to be interpreted with care, we did find notable differences in the size of the star arrival effect across our three countries. At the departmental level, a clear star-arrival effect is only present for Denmark, although there is more consistent evidence of positive effects for individual incumbents. The differences across countries do point to the importance of the design details of recruitment strategies and post-arrival supports. The importance of these design questions is heightened by the large national-level investments made to support star recruitments in some countries. Although our study does not specifically focus on the relatively small number of stars recruited through these national programmes, in ongoing related work we have interviewed stars recruited through these programmes in Denmark and Ireland as well as relevant institutional stakeholders. The early results from this work point to stars successfully embedding in local networks. However, the findings also point to various challenges. Several categories of challenges were evident in both programmes. As examples, a number of recruits in Ireland identified challenges such as poorly planned post-arrival supports and overtly bureaucratic institutional support processes; whereas some recruits in Denmark identified cultural challenges in internationalising the research focus and challenges posed by their institution's recruitment policies. With cluster-forming star-recruitment policies likely to become an increasingly important component of the science-policy mix, it is essential to better understand the policy and institutional strategy designs that maximise the effectiveness of these investments.

Supplementary material

Supplementary data for this paper are available at *Journal of Economic Geography* online.

Acknowledgements

For helpful discussions, we are indebted to Ajay Agrawal, Eoin Cullina, Jefferson Galetti, Alex Oettl, Stephen O'Neill and participants in an SFI workshop on science policy research. We also thank the editor Bill Kerr and two anonymous referees for their excellent advice and feedback. Of course, we alone are responsible for any remaining errors and omissions.

Funding

We gratefully acknowledge funding from Science Foundation Ireland under the SFI Science Policy Research Programme, Grant Number 17/SPR/5329.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

Adler, M. (1985) Stardom and talent. *The American Economic Review*, 75: 208–212.

Agrawal, A. and Cockburn, I. (2003) The anchor tenant hypothesis: exploring the role of large, local, R&D-intensive firms in regional innovation systems. *International journal of industrial organization*, 21: 1227–1253.

Agrawal, A., Cockburn, I., McHale, J. (2006) Gone but not forgotten: knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography*, 6: 571–591.

Agrawal, A., Kapur, D., McHale, J., Oettl, A. (2011) Brain drain or brain bank? The impact of skilled emigration on poor-country innovation. *Journal of Urban Economics*, 69: 43–55.

Agrawal, A., McHale, J., Oettl, A. (2017) How stars matter: recruiting and peer effects in evolutionary biology. *Research Policy*, 46: 853–867.

Athey, S., Eckles, D., Imbens, G. W. (2018) Exact *p*-values for network interference. *Journal of the American Statistical Association*, 113: 230–240.

Audretsch, D. B., Feldman, M. P. (1996) R&D spillovers and the geography of innovation and production. *American Economic Review*, 86: 630–640.

Autor, D. H. (2003) Outsourcing at will: the contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics*, 21: 1–42.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., Van Reenen, J. (2020) The fall of the labor share and the rise of superstar firms. *Quarterly Journal of Economics*, 135: 645–709.

Azoulay, P., Stuart, T., Wang, Y. (2014) Matthew: effect or fable? *Management Science*, 60:92–109.

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

Bilinski, A., Hatfield, L. A. (2020) Nothing to see here? Non-inferiority approaches to parallel trends and other model assumptions. arXiv:1805.03273 [stat].

Black, D., Henderson, V. (1999) A theory of urban growth. *Journal of Political Economy*, 107: 252–284.

Borjas, G. J. (2003) The labor demand curve is downward sloping: reexamining the impact of immigration on the labor market. *Quarterly Journal of Economics*, 118: 1335–1374.

Borusyak, K., Jaravel, X. (2017) Revisiting event study designs. Available at SSRN 2826228.

Breschi, S., Lissoni, F. (2009) Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography*, 9: 439–468.

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

Card, D. (1990) Unexpected inflation, real wages, and employment determination in union contracts. *American Economic Review*, 80: 669–688.

de Chaisemartin, C., D'Haultfoeuille, X. (2020) Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110: 2964–2996.

Deryugina, T. (2017) The fiscal cost of hurricanes: disaster aid versus social insurance. *American Economic Journal: Economic Policy*, 9: 168–198.

Docquier, F., Rapoport, H. (2012) Globalization, brain drain, and development. *Journal of Economic Literature*, 50: 681–730.

Duranton, G., Puga, D. (2004) Micro-foundations of urban agglomeration economies. In J. V. Henderson and J. F. Thisse (eds) *Handbook of Regional and Urban Economics*, Vol. 4, pp. 2063–2117. Amsterdam, NL: Elsevier.

Freyaldenhoven, S., Hansen, C., Shapiro, J. M. (2019) Pre-event trends in the panel event-study design. *American Economic Review*, 109: 3307–3338.

Fujita, M., Krugman, P., Venables, A. (2000) *The Spatial Economy: Cities, Regions, and International Trade*. Cambridge, MA: The MIT Press.

Gabaix, X. (1999) Zipf's law for cities: an explanation. *Quarterly Journal of Economics*, 114: 739–767.

Glaeser, E. L. (2008) *Cities, Agglomeration, and Spatial Equilibrium*. Oxford University Press.

Goodman-Bacon, A. (2021) Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, 225: 254–277.

Goyal, S. (2007) *Introduction to Connections: An Introduction to the Economics of Networks*. Princeton, NJ: Princeton University Press.

Goyal, S., van der Leij, M. J., Moraga-González, J. L. (2006) Economics: an emerging small world. *Journal of Political Economy*, 114: 403–412.
Griliches, Z. (1992) The search for R&D spillovers. *The Scandinavian Journal of Economics*, 94: S29–S47.
Gyourko, J., Mayer, C., Sinai, T. (2013) Superstar cities. *American Economic Journal: Economic Policy*, 5: 167–199.
Hess, A. M., Rothaermel, F. T. (2011) When are assets complementary? Star scientists, strategic alliances, and innovation in the pharmaceutical industry. *Strategic Management Journal*, 32: 895–909.
Jackson, M.O. (2008) *Social and economic networks*. Princeton, NY: Princeton University Press.
Jackson, M. O., Rogers, B. W. (2005) The economics of small worlds. *Journal of the European Economic Association*, 3: 617–627.
Jackson, M. O., Rogers, B. W., Zenou, Y. (2017) The economic consequences of social-network structure. *Journal of Economic Literature*, 55: 49–95.
Jacobson, L. S., LaLonde, R. J., Sullivan, D. G. (1993) Earnings losses of displaced workers. *American Economic Review*, 83: 685–709.
Jackson, M. O., Rogers, B. W. (2007) Meeting strangers and friends of friends: how random are social networks? *American Economic Review*, 97: 890–915.
Jaffe, A. B., Trajtenberg, M., Henderson, R. (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108: 577–598.
Kahn-Lang, A., Lang, K. (2018) The promise and pitfalls of differences-in-differences: Reflections on '16 and pregnant' and other applications. Working Paper 24857, National Bureau of Economic Research.
Kapur, D. McHale J. (2006) *Give us your best and brightest: The global hunt for talent and its impact on the developing world*. Washington D.C.: Center for Global Development/Brookings Institution Press.
Kerr, William (2019) *The Gift of Global Talent: How Migration Shapes Business, Economy & Society*. Stanford: Stanford Business Book.
Khanna, R. (2021) Aftermath of a tragedy: a star's death and coauthors' subsequent productivity. *Research Policy*, 50: 104159.
Kripfganz, S. (2016) Quasi–maximum likelihood estimation of linear dynamic short-T panel-data models. *The Stata Journal*, 16: 1013–1038.
Krugman, P. (1991) Increasing returns and economic geography. *Journal of Political Economy*, 99: 483–499.
Lacetera, N., Cockburn, I. M., Henderson, R. (2004) Do firms change capabilities by hiring new people? A study of the adoption of science-based drug discovery. In J. A. C. Baum and A. M. McGahan (eds) *Business Strategy over the Industry Lifecycle*. Bingley: Emerald Group Publishing Limited.
Lotka, A. J. (1926) The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16: 317–323.
Malmendier, U., Tate, G. (2009) Superstar CEOs. *Quarterly Journal of Economics*, 124: 1593–1638.
Newman, M. E. (2001) The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98: 404–409.
Merton, R. K. (1968) The Matthew effect in science: the reward and communication systems of science are considered. *Science*, 159: 56–63.
Oettl, A. (2012) Reconceptualizing stars: scientist helpfulness and peer performance. *Management Science*, 58: 1122–1140.
Perry, M., Reny, P. J. (2016) How to count citations if you must. *American Economic Review*, 106: 2722–2741.
Price, D. J. D. S. (1963) *Little science, big science*. Columbia University Press.
Radicchi, F., Fortunato, S., Castellano, C. (2008) Universality of citation distributions: toward an objective measure of scientific impact. *Proceedings of the National Academy of Sciences of the United States of America*, 105: 17268–17272.
Rosen, S. (1981) The economics of superstars. *American Economic Review*, 71: 845–858.
Roth, J. (2021) *Pre-test with caution: event-study estimates after testing for parallel trends*. Department of Economics, Harvard University.

Schmidheiny, K., Siegloch, S. (2020) On event studies and distributed-lags in two-way fixed effects models: Identification, equivalence, and generalization. *ZEW-Centre for European Economic Research Discussion Paper*, 20–017.

Science Foundation Ireland. (2021) Shaping our future: Delivering for today, preparing for tomorrow, Science Foundation Ireland Stategy 2025, Dublin.

Sun, L., Abraham, S. (2021) Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225: 175–199.

Waldinger, F. (2012) Peer effects in science: evidence from the dismissal of scientists in Nazi Germany. *Review of Economic Studies*, 79: 838–861.

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.

Wolfers, J. (2006) Did unilateral divorce laws raise divorce rates? A reconciliation and new results. *American Economic Review*, 96: 1802–1820.

Zucker, L. G., Darby, M. R., Armstrong, J. (1998) Geographically localized knowledge: spillovers or markets? *Economic Inquiry*, 36: 65–86.