Boppart, Timo; Staub, Kevin E.

Working Paper

Online accessibility of academic articles and the diversity of economics

Working Paper, No. 75

Provided in Cooperation with:
Department of Economics, University of Zurich

Suggested Citation: Boppart, Timo; Staub, Kevin E. (2012) : Online accessibility of academic articles and the diversity of economics, Working Paper, No. 75, University of Zurich, Department of Economics, Zurich,
https://doi.org/10.5167/uzh-62417

This Version is available at:
https://hdl.handle.net/10419/77555

University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 75

Online accessibility of academic articles and the diversity of economics

Timo Boppart and Kevin E. Staub

May 2012

Online accessibility of academic articles and the diversity of economics*

Timo Boppart and Kevin E. Staub†

May 9, 2012

Abstract

A key aspect of generating new ideas is drawing from different elements of preexisting knowledge and combining them into a new idea. In such a process, the diversity of ideas plays a central role. This paper examines the empirical question of how the internet affected the diversity of new research by making the existing literature accessible online. The internet marks a technological shock which affects how academic scientists search for and browse through published documents. Using article-level data from economics journals for the period 1991 to 2009, we document how online accessibility lead academic economists to draw from a more diverse set of literature, and to write articles which incorporated more diverse contents.

Keywords: Digitization, online publication, bibliometrics, knowledge production function, recombinant growth, citations, networks, scholarly communication.

JEL classification: A11, D83, O31, O33.

Acknowledgments: We are very thankful to Mary Glose from FSO, Peter Vlahakis and Luke Hospadaruk from JSTOR, and Steve Husted from EconLit for providing us with data and for their explanations. Nila Chea, Barbara Klotz and Lukas Rohrer provided excellent research assistance. A special thank goes to Christian Elsasser for his computational help. We thank Hartmut Egger, Josef Falkinger, Rainer Winkelmann, and seminar participants at the University of Bayreuth and University of Zurich for very helpful comments and suggestions. Kevin Staub gratefully acknowledges support by the Swiss National Science Foundation through fellowship PBZHP1-138692.

†University of Zurich, Department of Economics, Zürichbergstrasse 14, CH-8032 Zürich, Switzerland. E-mail: timo.boppart@econ.uzh.ch, kevin.staub@econ.uzh.ch

1 Introduction

Two elements broadly characterize academic research: (*i*) the production of knowledge in a "recombinant growth" framework (Weitzmann 1996, 1998a) where new ideas represent innovative combinations of previous ones, and (*ii*) attention as a scarce resource which limits researchers' processing capacity of existing ideas (Franck, 1999, Klamer and Van Dalen, 2002, Falkinger, 2007b).

Recombinant growth stresses the notion that the heterogeneity of existing ideas which serve as input is positively linked to knowledge accumulation.[1] The French mathematician Henri Poincaré described this idea succinctly as (Poincaré, 1910, p. 325): "Among chosen combinations the most fertile will often be those formed of elements drawn from domains which are far apart." In such a process the preservation of diversity of academic publications is important because a more diverse stock of knowledge enhances research productivity. Limited attention, on the other hand, implies that what matters specifically in this context is the diversity *perceived* by researchers. As Weitzman (1998a, p. 333) puts it: "[T]he ultimate limits to growth may lie not so much in our ability to generate new ideas, so much as in our ability to process an abundance of potentially new seed ideas into usable form."[2] This local diversity (i.e. diversity perceived by an individual researcher) depends on characteristics of researchers and on features of the technology used by researchers to learn

[1] This dependence on the stock of preexisting knowledge is often called the "standing on the shoulders of giants" effect (which goes back to a quote by Isaac Newton and now serves as an advertising slogan of Google Scholar).

[2] For a theoretical contribution how diversity can be measured and ranked see Weitzman (1992) and (1998b). Stirling (2007) emphasizes "variety", "balance" and "disparity" as three district properties of diversity. Consisting of these three components, Van den Bergh (2008) analyzes optimal diversity in a model of recombinant innovation.

about existing ideas.

In this paper, we explore the effect of one such technological aspect, the digitization of academic literature and its dissemination through the internet. The internet marks a profound technological shock which affected the way academic scientists search for and browse through published documents. On the one hand, the internet exemplifies a huge increase in the availability of scientific articles at very low (time) cost. On the other hand, the internet offers very powerful new tools such as search engines and hyperlinks to browse through this sheer amount of information. The specific empirical question we ask is how the internet affected the diversity of newly undertaken research by making the existing literature accessible online. The answer to this question is far from obvious. The new tools and, especially, search algorithms may allow researchers to find forgotten "lost pearls" and bring to their attention contemporaneous articles they would not read habitually. However, the internet also entails –almost by an empirical law– very unequal distributions of attention (Huberman, 2003). Evans (2008) documents that as more scientific articles became available online more recent articles were referenced more often and citations were more concentrated on fewer documents. In contrast to these aggregate measures of diversity, we focus on local ones measuring how diverse the ideas are a publication contains or is based on. Our results show that these local measures of diversity increase with the share of relevant literature being accessible online. These empirical findings can be linked to theoretical models of attention economies where comparative statics predict that increases in the diffusion of sender signals may diminish the equilibrium number of senders while resulting in access to a larger variety of senders for each individual receiver (Falkinger, 2007a, 2008).

The data we analyze consists of roughly two decades of publications in core eco-

nomics journals, starting in the pre-internet era and including the complete transition to full digitization. Our paper exploits the same basic exogenous variation in the date of online publication across different journals and across volumes within journals pioneered by Evans (2008), and contributes to a small and very recent literature relying on the same source of variation which explores the impact of online access for the economics profession (Depken and Ward, 2009; McCabe and Snyder, 2011).[3] So far, this literature has focused primarily on the impact articles' online access had on these articles' number of citations. Depken and Ward (2009) show that access to the online platform "Journal STORage" (JSTOR) increases the number of citations to journals contained in JSTOR as well as to older journal volumes. McCabe and Snyder (2011) document that the number of citations a publication receives increases by about 10 percent as it is included in JSTOR and that this effect is about the same both for often-cited as well as rarely-cited papers. By studying the impact on upcoming articles' contents, our research addresses quite a different aspect of the scientific process.

Section 2 introduces our two measures of diversity. The construction of our measure of online accessibility and our identification strategy are outlined in detail in section 3. Estimation results are discussed in sections 4 and 5 and section 6 concludes.

[3] A much larger literature exists on the impact of access to internet contents in more traditional market settings. In a recent review of this literature, Brynjolfsson, Hu and Smith (2010) stress that a key channel through which information technology improvements changed how consumers learn about goods and services, and how producers develop, distribute and deliver them, is through a transformation in search and recommendation tools.

2 Two measures of diversity

Our analysis considers two measures of diversity: (*i*) the distribution of pairwise geodesic distances of cited references and (*ii*) the number of Journal of Economic Literature (JEL) classification codes assigned to a publication. The geodesic distance between two items is the shortest back-in-time connection within the citation network.

The distribution of geodesic distances between the references of an article is a distinct measurement of diversity: The share of short distances is high if a publication draws only from one narrow and well-connected literature. In contrast, higher shares of large distances result if the paper connects several separated strands of the literature for the first time. The second measure of diversity considered is the JEL codes assigned to a publication by the American Economic Association's bibliography, EconLit.[4] The JEL classification indexes the contents of an article describing which fields and subfields it falls into, and thus uncovers the breadth of an article within economics. In the following, these two measures are explained in more detail.

The measures were obtained for every article published between 1991 and 2009 in 50 selected core journals of economic research. The list of journals includes all "top five,"[5] top field and second tier general interest journals (as well as their historical predecessors). Table 15 in the Appendix provides an alphabetic list of the journals.[6]

[4]It is important to emphasize that these JEL codes are not the ones declared by the authors of a paper. The codes we use are assigned by a team of economists at EconLit.

[5]American Economic Review, Econometrica, Journal of Political Economy, Quarterly Journal of Economics and Review of Economic Studies.

[6]The set of journals includes all journals considered in the standard Tilburg ranking, as well as the list considered in Palacious-Huerta and Volij (2004). Furthermore, it includes all core journals in Conroy et al. (1995), all journals used in Kalaitzidakis, Mamuneas and Stengos (2003),

2.1 Geodesic distances

Geodesic distances provide essential information about the structure of networks. Measures of network connectivity and centrality are usually characterized by functions of these distances.[7] The analysis of geodesic distances in networks of academic citations dates back at least to De Solla Price (1965). The previous literature is largely explorative and aims at describing the distribution of geodesic distances in particular citation networks (e.g. Yin et al., 2006, Franceschet, 2012).[8,9] By focusing on the distribution of geodesic distances of *articles' references*, we capture the local connectivity of references, which we interpret as a measure of articles' (local) diversity. This local connectivity specific to citation networks has not been explored in the literature so far.

The calculation of geodesic distances requires knowledge of the entire network

as well as all top 20 journals in Combes and Linnemer (2011). The list is comparable to Depken and Ward's (2009) and McCabe and Snyder's (2011) who include 79 and 63 economics journals, respectively. Using eigenfactor.org's list of over 200 economics journals, we found that our list has an eigenfactor score of around 0.75 in 1995. I.e., randomly traversing the citation network spanned by all economics journals, the list's 50 journals are selected 75 percent of the time.

[7]Bavelas (1948) and Freeman (1979) provide early foundations. See Newman (2003) for an overview.

[8]A related strand of the literature on scholarly communication studies social networks defined by co-autorship relationships (Newman 2001a,b,c), as opposed to information networks defined by citations. This literature, too, is foremost descriptive. For an exception, see Kretschmer (2004) who links co-author networks features to author productivity.

[9]The analysis of patents is also conceptually related to scholarly communication networks. For instance, social networks akin to co-authorships are defined by inventor collaborations, and information networks arise in the context of patent citations. Recent examples of papers studying geodesic distances are Balconi, Breschi and Lissoni (2004), in the former context; and Lee, Su and Wu (2010), in the latter.

of citations. We downloaded from Thomson Reuters' Web of Science the list of references of all items published between 1955 and 2009 in the 50 core economics journals. The sample does not only include articles but also notes, letters, book reviews etc., which gives rise to a total of 129,145 items. To construct the citation network, references were matched back to the published items. On average, we are able to match 36 percent of all references, and 44 percent of references in articles published 1991 to 2009. Unmatched references may refer to publications prior to 1955 or to publications in books, working papers or disregarded journals. Finally, we calculated the shortest back-in-time connection within the citation network for all binary pairs of (identified) references.

In Figure 1, this process is illustrated for a paper-and-proceedings article written by John Cochrane and Monica Piazzesi and published in 2002,[10] visualized in the graphic in Panel (a) by a red node. This article made seven references. Within our sample we can identify four of them and these items are depicted as blue nodes.[11] The four identified references give rise to six bilateral connections (unordered pairs) among them. We then calculated for each of the bilateral links the shortest back-in-time connection within the entire citation network spanned by the sample of over 100,000 items. In the example of Figure 1, one of the references, *Rudebusch (1998)*, cites another one directly –*Christiano et al. (1996)*– which implies a geodesic distance of one (Panel b). Moreover, *Christiano et al. (1996)* is linked to a third reference, *Cochrane (1989)*, via two connections; as is the case for *Rudebusch (1998)*

[10] "The Fed and Interest Rates: A High-Frequency Identification," *American Economic Review, Papers and Proceedings*, May 2002, Volume 92, Issue 2, pp. 90-95.

[11] In the following we abbreviate all the sources by authors and publication date in italic without specifying the entire reference. For the exact reference of the citations we refer the reader to the paper by John Cochrane and Monica Piazzesi.

and the fourth reference, *Clarida et al. (2000)*. Panel (c) plots these geodesic distances of order two. The shortest connection from *Clarida et al. (2000)* to *Cochrane (1989)*, and to *Christiano et al. (1996)*; as well as between *Rudebusch (1998)* and *Cochrane (1989)*, is given by three steps (Panel d). We determine the geodesic distances iteratively up to a length of 3, giving rise to a probability mass function over four categories, with the last category comprising geodesic distances strictly larger than three. In the example of Figure 1, the fraction of pairwise geodesic distances of order one, two, three and higher than three are $\frac{1}{6}$, $\frac{1}{3}$, $\frac{1}{2}$, and 0, respectively.

This simple example, chosen for illustration purposes, is not typical for the dataset. The median article has 10 identified references and thus its references' citation network comprises 45 pairs. Over the whole sample period 1991-2009, the average fractions of these geodesic distances are 25, 27, 20, and 28 percent.

— Figure 1 about here —

2.2 JEL codes

Our second measurement of diversity is the number the JEL classification codes assigned to a publication. Up to six three-digit JEL codes are assigned by EconLit to each article and we downloaded this information from the EconLit webpage. The first digit of a JEL codes is a letter which divides economics into twenty main fields, such as "public economics" or "industrial organization". The JEL classification system was introduced in 1991 and consequently we observe these classification codes only from then onwards.[12] While half the articles fall into exactly one field according to the one-digit definition, about 37 percent contribute to two fields,

[12]See Pencavel (1991), the editor's note with which the Journal of Economic Literature introduced the new system. The JEL codes replaced an earlier, more narrow classification system.

and somewhat over 10 percent have three one-digit JEL codes. There are large differences between journals; for instance, while the average article in Econometrica has about 1.1 one-digit JEL codes, the Journal of Development Economics' average article has about 2.3. The variation within journals is even larger. Relying on journal-year fixed effects, this variation within a journal is the one exploited in the empirical section. The last two digits classify the twenty fields into narrower sub- and subsubfields resulting in a very subtle measure of diversity. The median article has two three-digit JEL codes, while about one third of articles have more than two. In our analysis we consider the number of distinct one-, two- and three-digit JEL codes each as a separate dependent variable.[13]

JEL codes constitute a unique and precise categorization of articles' contents beyond its main field, which other, similarly structured applications such as patent citations lack. In such datasets the intellectual content of a patent is limited to one "patent class" only. An example is the NBER patent citations dataset (cf. Hall, Jaffe and Trajtenberg, 2001). Trajtenberg, Henderson and Jaffe (1997) introduced a variable called "originality" which is (the negative of the) Herfindahl concentration index of different patent classes a specific patent cites. Clearly, our diversity measures are closely related to this concept of originality.

[13]JEL codes have been the subject of some descriptive work which used them to characterize the evolution of economic fields or subfields over time (Kim, Morse, Zingales, 2006; Kelly and Bruestle, 2011). Previous literature using JEL codes in regression analysis has included them as control variables for the specific fields (e.g. Formby, Gunther and Sakano, 1993, Axarloglou and Theoharakis, 2003, Boschini and Sjögren, 2007).

3 Online accessibility of economics journals

For the online accessibility of publications we use data from two sources: "Fulltext Sources Online" (FSO) and JSTOR. The FSO data contains, for each year 1998-2009 and each online platform, information on which volumes of which journal were accessible online.[14] The FSO data contains this information for the journals' own webpage as well as for all major providers (such as e.g. EBSCOhost, LexisNexis, ScienceDirect or WilsonWeb) with the important exception of JSTOR.[15] Since JSTOR is one of the most important providers of online access (and has been even more important historically) we augment the dataset with the information about the date of a journal volume's first download at JSTOR.

A satisfactory measure of online accessibility based on these data should enable us to distinguish online accessibility from other secular time trends such as general internet usage. Achieving this should be possible since research projects differ in which subsets of the entire past literature are relevant to them. We assume that articles cite all relevant past works (a requirement stated in all journals' article submission guidelines) and define an article's relevant (past) literature as the set of journals in which the article's references were published.[16] Journals varied widely

[14]We assume that no volumes were accessible before 1998 on platforms covered by the FSO data. This is reasonable, since only about 2 percent of volumes were online on platforms other than JSTOR in 1998, and these accessible volumes were mainly the contemporaneous ones. For the historically most important platform of online access, JSTOR, we do have the data about the accessibility of journals even prior to 1998.

[15]JSTOR is not included in the FSO database before 2009.

[16]Evans (2008), whose units of observation are journal-years, considers the journal where an article is published as the only relevant past literature. This approach seems unsuitable for economics. In our data, on average only 7 percent of citations refer to the same journal where the article is published. Even for the journal where this ratio is highest over the whole period – the

and unsystematically regarding the date when they first went online and the pace with which their volumes' back catalogs were made accessible on the internet,[17] so that, in general, online accessibility will differ between articles with different relevant past literature even if the articles were written in the same year.

3.1 Online accessibility treatment variable

In order to measure to which degree an article's relevant literature has been accessible online, we calculate the share of online accessible volumes[18] of all cited journals at the time the paper has (presumably) been drafted. More formally, we denote the set of all journals by $\boldsymbol{\mathcal{J}}$. Suppose an article i has been published in year t and cites the subset of journals $\boldsymbol{J}_i \subset \boldsymbol{\mathcal{J}}$. Indexing the journals in $\boldsymbol{J}_i$ by j, our online treatment is given by

$$T_i(t) = \frac{\sum_{j \in \boldsymbol{J}_i} a_j(t-1)}{\sum_{j \in \boldsymbol{J}_i} h_j(t-1)}, \tag{1}$$

where $a_j(t-1)$ is the number of volumes of journal j that have been accessible online in the year $t-1$ on at least one platform and $h_j(t-1)$ is the number of existing historical volumes at date $t-1$ published in journal j. This measure of online availability of relevant literature is article-specific, since it depends on the set of cited journals and their online accessibility.[19] In the empirical section the treatment defined in (1) is called *percent online*.

Journal of Finance – it is only 20 percent.

[17] Cf. Evans (2008), Depken and Ward (2009), McCabe and Snyder (2011).

[18] We use the term "volume" to denote all issues of a journal published in the same calendar year.

[19] The online treatment defined in (1) is based on two specific assumptions regarding *(i)* the set of relevant literature and *(ii)* the time lag between first draft and publication of a paper. In section 4 we show that our results are robust to other reasonable specifications.

The default behavioral model behind our treatment variable is that an author facing zero online accessibility searches the relevant literature in print, for instance by browsing his library's collection of volumes aided by keyword or abstract indexation systems; in contrast, another author whose relevant literature is partially accessible online will browse this electronic literature by using internet tools, while still using the same methods as the previous author for the literature available only in print. In such a case, the use of internet literature browsing and searching tools coincides exactly with our online treatment variable. In practice, some deviations from such behavior are likely. For instance, very low levels of *percent online* might not induce researchers to search online; and, conversely, researchers whose literature is almost entirely online might neglect the few remaining print-only volumes. However, studies on researchers' literature searching behavior suggest that the joint use of print and electronic resources (with declining use of print) was typical for researchers during the transition to full electronic access (Tenopir, Hitchcock and Pillow, 2003; Boyce et al., 2004), so that *percent online* should be a reasonable approximation to researchers behavior.[20]

A qualification needs to be made at this point. While we can compute an article's share of relevant literature which was *accessible* online, we do not observe whether the article's authors effectively *used* the internet to search for related literature. Thus, online accessibility effects should be understood as intention-to-treat effects of online access. In section 5 we use information about aggregate time trends of subscriptions to online contents from other studies to explore the relationship between

[20] An alternative interpretation of the online treatment variable can be given assuming a different, more stylized behavioral model where there exist only two types of researchers: one group using print literature only, the other group relying exclusively on online literature. Then, *percent online* can be interpreted as the probability that the article's author is an online researcher.

accessibility and access.

A second remark relates to the question whether *percent online* could be endogenously linked to diversity. Many plausible stories of endogeneity which rely on differences in journal specific time trends are excluded by our journal-year fixed effects specification. Another concern is that the top five journals were available online from very early on. Thus, articles citing predominantly top five journals might have a high measure of *percent online*. This would bias the estimated effect if such articles were inherently more/less diverse. Similarly, researchers who tend to cite older literature (which has lower online accessibility, on average) might be inherently more/less diverse, too. Since any measure of online accessibility is bound to have such problems, we will address these concerns in the empirical section through appropriate robustness checks. For instance, we will test whether there is an online accessibility effect when comparing articles with the same share of top five journals cited or with the same age distribution of references.

A more challenging concern is that –beyond the differences between journals and over time– there might be some additional heterogeneity on the author level. Specifically, consider the hypothetical case where online accessibility has no effect on diversity, yet authors differ in the extent to which they make use of diverse sources. Then, if the propensity to adopt the internet is correlated with the diversity of an author's research agenda,[21] we would find spurious effects of the online treatment on diversity. We address this concern by including author fixed effects (in addition to the journal-year fixed effects). However, one should bear in mind that estimations with author fixed effects might be too conservative since they exclude channel

[21]For instance, younger researcher might adopt the new technology faster and might differ from their older colleagues in terms of the diversity of their research interests.

through which the effect of online accessibility works. For instance, online access could change the composition of "diverse" and "non-diverse" authors. Therefore we see the regressions with author fixed effects as an important robustness check but not our main specification.

3.2 Online accessibility and diversity over time

Our estimation sample includes the 45,553 articles or paper-and-proceeding articles published between 1991-2009 in the 50 considered core journals. Figure 2 plots, for each year, the average share of existing volumes that were accessible online on at least one platform. Online accessibility of economics journals started in 1997 on the JSTOR platform. Since then, the back volumes of the different journals were gradually scanned and uploaded, and in 2009 almost all publications were available online. Hence, the considered period covers some years of the pre-internet era as well as the entire transition to full coverage. Until the turn of the millennium online access was clearly dominated by JSTOR. Later, other platforms caught up. A large amount of back volumes of Elsevier journals (which are not included in JSTOR) were made accessible in 2005.

— Figure 2 and Figure 3 about here —

Figure 3 plots the average number of one-, two- and three-digit JEL codes assigned to an article. For the first years in our sample, the number of assigned codes is constant. Then, it rises for all JEL code digits from 1995 onwards. Figure 2 and 3 reveal a positive time correlation of online accessibility and the number of assigned JEL codes. But since the number of assigned JEL codes might not be comparable between different years we do not want to overstate this correlation. For instance,

some additional JEL codes were added after 1991. Moreover, the assigning process might have changed over time.[22]

With 0.25, 0.27, 0.20 and 0.28, the four different categories of geodesic distances have about the same relative prevalence (the summary statistics of all variables can be found in Table 14 in the Appendix). But these averages mask huge cross-sectional variations. In the case of the geodesic distances, the time trend is even harder to interpret than it is in the case for the number of assigned JEL codes. The way geodesic distances are constructed generates an inherent time trend, since the citation network is more comprehensive for later years where our dataset covers more back volumes. This makes the share of geodesic distances higher than three falling by construction.[23] It is the aim of our empirical strategy to exploit this cross-sectional variation while controlling for any secular time trends in order to estimate the effect of online access on the measures of diversity.

3.3 Identification strategy

As explained above, in the distribution of geodesic distances, time trends emerge by construction. Such inherent time trends are present in the assignment of JEL codes,

[22]E.g. the dint in the number of codes in 2006 might be explained by a change of EconLit's managing director. The assigning process is, however, consistent within a given year.

[23]It is possible to try and capture this by partialling out some time trend. However, this requires assumptions for the trend's functional form. In Figure 4 in the Appendix a possible correction has been applied. In that graph the average fractions of short geodesic distances (order one and two) fall over time, while those of order higher than three increase. This would be in line with an increase in diversity over the period. In contrast to Figure 4, our regression framework presented in the next section, which does include time fixed effects, does not require arbitrary assumptions about the time trend.

too. For instance, some three- and two-digit codes, and even a one-digit code, have been introduced after 1991. Moreover, until the mid 1990's the production process set an upper bound of five for the number of codes assigned to an article. For all these reasons it is indispensable to control for year fixed effects to disentangle the causal effect of online accessibility from other ongoing trends.[24]

There are substantial differences in diversity measures between journals. While it is not obvious that this journal-specific diversity is related to online accessibility, such a correlation could arise if the relevant literature predominant in some journals was accessible online later or earlier than in others. To allow for changing time- and journal-specific heterogeneity in the most flexible way, we account for journal-year fixed effects. Thus, the variation which we use to estimate the effect stems from cross-article differences in the share of the relevant literature that is online within a given year and a given journal. Since the structure of the data consists of articles in journal-years forming an unbalanced pseudo-panel, we use the (linear) panel specification

$$Y_{iv} = \alpha T_{iv} + \boldsymbol{X}'_{iv}\boldsymbol{\beta} + \mu_v + u_{iv}, \tag{2}$$

where i indexes articles and $v = \tilde{v}(j,t)$ journal-years. Thus, Y_{iv} represents the diversity measure of article i which was published in journal j and year t. With slight abuse of notation, T_{iv} stands for the online treatment defined in (1). $\boldsymbol{X}_{iv}$ is a vector of possible article-specific control variables to be discussed below, and $\boldsymbol{\beta}$ is a conformable parameter vector. μ_v denote fixed effects specific to journals and years. Finally, u_{iv} is an idiosyncratic shock. Under mean independence of u_{iv} from T_{iv}, $\boldsymbol{X}_{iv}$, and μ_v, the coefficient α measures the causal effect of a marginal increase

[24] See also McCabe and Snyder (2011) who illustrate the empirical importance of flexible controls of time trends.

in the share of literature online on the diversity measure Y_{iv}. Equation (2) can be estimated conveniently using the OLS within-estimator.

4 The effect of online accessibility on diversity

4.1 Baseline results

Estimation results for the baseline model (2) are collected in Table 1. Panel I depicts the coefficient of the treatment variable *percent online* for regressions on the number of one-, two- and three-digit JEL codes (first three columns) and on the fraction of geodesic distances equal to one, two and three (last three columns). In all regressions, the panel dimension of the OLS within-estimator is journal-years, of which there are 859 unique groups. The standard errors shown are robust to heteroskedasticity and clustering at the journal-year level.

Since the average of *percent online* varies from zero in 1991 to one in 2009, the coefficients of the first three columns can be read as the total change in the average number of JEL codes comparing a world without any online access to one which provides full access to all 50 journals. The effect is substantial: the coefficients —about 0.2 for one-digit JEL codes and about 0.35 for two- and three-digit JEL codes— correspond to 39, 45 and 32 percent of the increase of one-, two- and three-digit JEL codes in the data in the observed period.

— Table 1 about here —

Online accessibility has a diversity-enhancing effect on the distribution of geodesic distances, too: The fractions of low geodesic distances ($g = 1, 2$) are reduced and higher geodesic distances ($g = 3$, $g > 3$) are increased as a consequence of online

accessibility. The coefficients 0.067, 0.0027, 0.0589 and 0.0108, respectively can be read as percentage-point changes in the fractions of geodesic distances.[25] Thus, the fraction of shortest geodesic distances (whose average over the entire period is 25 percent) is estimated to have shrunken by about 6.7 percentage points due to online accessibility of the literature.

While the regressions in Panel I control for any confounding journal-year-specific characteristics, there might be some further heterogeneity within journal-years correlated to the treatment *percent online* which could drive the effect. This is a stronger concern for the regressions on geodesic distances. For instance, higher shares of long geodesic distances can result from citing two types of relatively unconnected work. The first possibility is that the references in question, while well-connected to other literatures, are relatively unconnected between them. The second possibility is that the references in question are relatively unconnected at all. While we would readily interpret the first case as a sign of diversity, some might want to exclude the second case from counting as diversity. To address this issue, Panel II adds to the specification the average number of citations received by a reference and the average number of references contained in a reference. In this way, the online accessibility effect is computed for similarly well-connected reference networks. The list of other control variables in Panel II includes a papers-and-proceedings dummy, the number of authors, number of pages, number of references made, number of distinct journals referenced, percent references found in the data, and percent of self-references.

The effects in Panel II remain large and statistically significant. The coefficients

[25]The coefficient of *percent online* in a regression on $\Pr(g > 3)$ is not shown in the table, but it can be easily obtained from the three numbers depicted in the last three rows of the table, since it is equal to the negative of the sum of the coefficients in the regressions for $g = 1, 2, 3$ (because shifts in the probability mass of the distribution add up to zero).

for the JEL code regressions are somewhat smaller than before. This is mainly the result of controlling for number of pages and number of different journals referenced, two variables mediating the effect of online accessibility on diversity. Whether these variables are part of the causal effect and should not be controlled for is to a large extent a matter of taste and interpretation. The effect on the distribution of geodesic distances is slightly larger overall, with about 7.5 percentage points being shifted from the lower part of the distribution ($g = 1, 2$) to the right tail ($g = 3$, $g > 3$). While without controls the shift was mainly from $g = 1$ to $g = 3$, now the effect is more evenly distributed among the four categories of g.[26]

4.2 Robustness checks

An important first robustness check for our results relates to the appropriate lag of the treatment. The time when an article's references were collected is unknown and has to be inferred from the date of publication; there is also bound to be differences in length of the publication process across articles. In (1) we made the informed guess that the best approximation is the online accessibility faced by authors one year prior to publication.[27] Table 8 in the Appendix explores alternative lags of zero, two and three years. Given the heterogeneity in publication process length, it would be worrisome to find that the results in Table 1 hold only under the one year lag. Comfortingly, the results remain qualitatively the same for all lags explored, although the effects are strongest for the one- and two-year lag, which is in line with our expectations that a majority of the articles' time from draft to publication lies

[26] The full set of estimates is shown in Table 7 in the Appendix; summary statistics for dependent variables and all regressors are in Table 14 in the Appendix.

[27] Evans (2008) uses the same lag specification, while Depken and Ward (2009) and McCabe and Snyder (2011) use a lag of zero.

in the one-to-two-year range.

Next, we set out to assess the robustness of our treatment by exploring other ways of capturing online accessibility. Implicitly, the treatment *percent online* gives more weight to long-standing journals (with many volumes) because the percentage is calculated over the sum of all cited journals' volumes. An alternative which weights journals equally is to construct the treatment as the percent online in the average journal cited.[28] Similarly, treatment can be defined as the percent of an article's references that were online one year prior to publication. This weights the journals by their share in the reference list. Finally, instead of focusing on percentages, treatments can also be constructed based on the *absolute number* of volumes online (an approach related to Evans, 2008). Table 9 in the Appendix documents that the baseline results from Table 1 remain valid for any of these alternative treatments.

Another robustness check is with respect to the data sources of online access of the different platforms. Note that our measure of online accessibility combines information obtained directly from JSTOR with the one collected by FSO. Detailedness and quality of these two sources varies. Whereas FSO collects its data twice a year, JSTOR's database is very precise.[29] To make sure that such differences between data sources are not influencing our results, we constructed two treatments: one taking into account access provided by JSTOR only, the second taking into account access on the remaining online platforms contained in the FSO data. The results (displayed in Table 12 in the Appendix) show that disaggregating the treatment by data source delivers estimates that are very similar to the aggregate treatment in the baseline specification.

[28] Formally, the treatment is then calculated as $\sum_{j \in \boldsymbol{J}_i} \frac{a_j(t-1)}{h_j(t-1)}$ instead of (1).

[29] In fact, we know from JSTOR for each journal issue the exact date of first user access.

The OLS estimator used in Table 1 gives the best linear fit for our model of diversity and online accessibility without relying on strong distributional assumptions. It also has the attractive property that the effects for the geodesic distances add up to zero. In Table 10 in the Appendix we explore an alternative, constant-elasticity specification which we estimate both by OLS (using the logarithm of the dependent variables) and by Poisson Pseudo-Maximum Likelihood. Again, the effects (which are now to be interpreted as approximate percental changes) are similar.

The last issue we explore is a refinement of the fixed effects. While defining fixed effects at a more detailed level can purge the online accessibility effect from more confounding through unobserved heterogeneity, there is a trade-off to be considered since such an approach entails a loss of precision because of the higher number of fixed effects. In a first step, we treated papers-and-proceedings issues of a journal as a separate journal. Since most journals publish such issues, the number of panel units for these regressions increased to $1,456$. As a logical consequence of such an approach we can go even one step further and define a separate fixed effect for every single issue published in every journal in the period. This gives over 4,800 fixed effects. With both specifications, as the estimates in Table 11 in the Appendix show, the results are only marginally affected.

4.3 Author fixed effects

A more fundamental refinement of the fixed effects changes the panel dimension to a much less aggregated unit: the authors. While in our baseline regressions we exploit the variation in online accessibility between articles of a particular journal in a given year, a different source of variation comes from repeated publications of the same authors.

Exploiting only the variation for a given author (group) changes the interpretation of the coefficients, as the online accessibility effect being estimated excludes some channels which are part of the effect using the within journal-year variation. For instance, the availability of online literature may have an impact on the composition of the pool of authors, increasing the share of authors which are efficient users of online tools. In the estimation with journal-years fixed effects this margin is part of the causal effect as the pool of authors is not kept constant and changes with the spread of online accessibility. While ultimately we favor this approach, the specification with author fixed effects provides an important alternative view which shows the effect of online accessibility for authors publishing repeatedly during this period.

We approach this issue from two perspectives. In the first take, we extract from the EconLit database all author names which appear at least in two articles, leaving us with $12,165$ distinct authors.[30] We cloned articles with multiple authors to create one record for every author, and obtained a total of $67,903$ observations. Observations corresponding to the same article are clearly not independent, and the reported standard errors account for this correlation. Indeed, we used two-way clustered standard errors (Cameron, Gelbach and Miller, 2011) which are robust to heteroskedasticity and clustering at the article level, as well as at the author level. Moreover, in addition to the control variables, we included a full set of journal-year indicator variables to account for these fixed effects, too. The results are printed in Panel I of Table 2. Panel II contains results where the panel units are co-author-groups (including groups of size one, i.e. single authors). There are $7,307$ such

[30] We used data from EconLit as we found it significantly more reliable than Thomson Reuters', which contained numerous inconsistencies in the coding of author names.

unique co-author-groups which have published more than twice in our data. The total number of articles they have written is 21,767. As before, we additionally include over 800 journal-year fixed effects and our list of control variables. Standard errors are robust to clustering at the co-author-group level.

— Table 2 about here —

The results in Table 2 are substantially less precise. Given the substantial loss of degrees of freedom, this does not come as a surprise. It is the more remarkable, therefore, that the results in this table reveal the same patterns than those from the baseline regressions. To be sure, the coefficients are attenuated compared to the baseline; still, we find that online accessibility significantly increased the number of JEL codes, and that it transferred probability mass from the lower end of geodesic distances' distribution ($g = 1, 2$) to its right tail.

5 Effect heterogeneity and further results

Having established the robustness of the effect of online accessibility on the diversity of academic articles in economics, this section explores the heterogeneity of the effect and possible channels through which it affects the diversity measures.

5.1 The effect over time

A potential source of heterogeneity in the effect is time. While there are many potential factors with a time trend, one of them has been highlighted in the literature as particularly relevant for online accessibility: institutional subscription to platforms providing online contents of economics journals (i.e. effective online

access). For instance, Depken and Ward (2009) and McCabe and Snyder (2011) document that the number of institutions subscribing to JSTOR (and to Elsevier's online contents) increased in the considered period almost linearly (cf. Depken and Ward, 2009, Fig. 1, McCabe and Snyder, 2011, Fig. 7). Table 3 shows estimation results for a specification which adds an interaction of the treatment with a linear time trend, which is bound to capture this effect of increasing online access. The time trend was normalized to zero in 1997, so that the coefficient on *percent online* gives the effect in that year. The effect on the JEL codes is indeed moderate in the beginning of the period and shows an increasing time trend, suggesting that as more researchers gained access to online contents the effect of online access on the number of JEL codes became more prominent. However, results are less clear-cut for the geodesic distance regressions, where the absence of a time trend cannot be rejected.

— Table 3 about here —

5.2 The effect across journals

The effect of online accessibility found in our baseline regressions could vary greatly for different journals. While for the average journal the effect on diversity is positive, it could be that this aggregation masks negative effects for certain classes of journals. To explore this issue we estimated a specification with interactions for three classes of journals: the "top five" journals, general interest journals, and field journals (Table 4).[31] We find the same kinds of effects as in the baseline regressions for every journal category. The effects are most pronounced for second tier general interest journals, but the effects are large for all three categories.

[31]Note that uninteracted level effects are subsumed in the journal-year fixed effects.

— Table 4 about here —

5.3 The composition of references' journals and publication years

One way in which online accessibility may have influenced diversity is by reducing the bibliographic importance of the journal an article appeared in. The correlation between reading a particular journal and contributing to it might have been weakened by the internet, leading to a more diverse pool of influences. A second way in which online accessibility may have influenced diversity is by increasing the importance of the "top five" journals, which are journals publishing diversely to begin with. Two characteristics of these journals are that they have a long publication history and that they were among the first to be put online. Thus, researchers relying on online sources were likely to rely on these journals. Table 5 addresses this issue. The specifications for which results are shown add two regressors to the model: the percent of an article's references that were published in the journal where the article appeared, and the percent of an article's references that were published in one of the "top five" journals. The results indicate that these two channels explain some of the online accessibility effect. The variables themselves are highly significant, the percent references to "top five" journals increasing diversity, the percent references to the own journal decreasing it. The online accessibility effect on JEL codes is reduced by about 30 to 50 percent. It remains significant, however, suggesting that there are further channels at work as well. The pattern is similar but less accentuated for geodesic distances, where the joint reduction in the fraction of distances one and two is about 20 percent.

— Table 5 and Table 6 about here —

Finally, we set out to quantify the importance of the age distribution of an article's reference for the effect on diversity. The results from regressions including average citation lag (i.e. the difference in years between the article's publication year and that of its average reference) are shown in Table 6. Average citation lag has been used as the primary dependent variable in previous work analyzing the impact of online accessibility on academic research (Evans, 2008 and Depken and Ward, 2009).[32] The fact that our coefficients of interest remain virtually unaffected in size and statistical significance when including citation lag shows that our direct measures of diversity go substantially beyond the heterogeneity of references' age distribution.

6 Concluding remarks

This paper documents how online accessibility of articles lead to an increase in the diversity of upcoming economic research. We do so by considering local measures of diversity, i.e. the diversity of ideas *a single article* touches on or is based on. This is a sharp contrast to the aggregate measures of diversity considered in Evans (2008) or McCabe and Snyder (2011). It can well be, that the local diversity increases at the same time as the number of overall cited articles decreases and the concentration of cited articles increases (as suggested by Evans, 2008). For instance, different fields of economics may get tighter connected, whereas in each field some "superstars" emerge. However, the results of McCabe and Snyder (2011) suggest that in the

[32] Table 13 contains further regressions including the median and standard deviation of references' publication year. These results lead to the same conclusion that our measures of diversity capture a different dimension of heterogeneity than the distribution of citation lags.

case of economics, online access did not skew the distribution of citations and did lead to an decline in the number of uncited papers. Whether local or aggregate measures of diversity are of interest depends on the context. And it is unclear whether more diversity is a priori desirable. In the introduction, we provided one example of a setting where such local diversity matters and is desirable – a model of scientific research based on recombinant growth and limited attention. In any case, our results suggest that online access did not narrow but broaden the mind of economic researchers.

References

Axarloglou K., and V. Theoharakis (2003), "Diversity in Economics: An Analysis of Journal Quality Perceptions", *Journal of the European Economic Association*, **1**(6), 1402-1423.

Balconi M, S. Breschi, and F. Lissoni (2004), "Networks of inventors and the role of academia: an exploration of Italian patent data", *Research Policy*, **33**(1), 127-145.

Bavelas, A. (1948), "A mathematical model for group structures", *Human Organization*, **7**(3), 16-30.

Boschini A., and A. Sjögren (2007), "Is Team Formation Gender Neutral? Evidence from Coauthorship Patterns", *Journal of Labor Economics*, **25**(2), 325-365.

Boyce P., D. W. King, C. Montgomery and C. Tenopir (2004), "How Electronic Journals Are Changing Patterns of Use", *The Serials Librarian*, **46**(1-2), 121-141.

Brynjolfsson E, Y. Hu, and M. D. Smith (2010), "Long Tails Versus Superstars: The Effect of IT on Product Variety and Sales Concentration Patterns", *Information Systems Research*, **21**(4), 736-747.

Cameron, A. C., J. B. Gelbach and D. L. Miller (2011), "Robust Inference with Multiway Clustering", *Journal of Business and Economic Statistics*, **29**(2), 238-249.

Combes P.-P., and L. Linnemer (2011), "Inferring Missing Citations: A Quantitative Multi-Criteria Ranking of all Journals in Economics," HAL Working Paper Series, Nr. 520325.

Conroy M. E., R. Dusansky, D. Drukker and A. Kildegaard (1995), "The Productivity of Economics Departments in the U.S.: Publications in the Core Journals" *Journal of Economic Literature*, **33**(4), 1966-1971.

De Solla Price D. J. (1965), "Networks of scientific papers", *Science*, **149**(3683), 510-515.

Depken, C. A., and M. R. Ward (2009), "Sited, Sighted, and Cited: The Effect of JSTOR in Economic Research", *SSRN Working Paper Series*, Working Paper No. 1472063.

Evans, J. A. (2008), "Electronic Publication and the Narrowing of Science and Scholarship", *Science*, **321**(5887), 395-399.

Falkinger J. (2007a), "Attention Economies", *Journal of Economic Theory*, **133**(1), 266-294.

Falkinger J. (2007b), "Distribution and Use of Knowledge Under the Laws of the Web", *CESifo Working Paper*, No. 2154.

Falkinger J. (2008), "Limited Attention as a Scarce Resource in Information-Rich Economies", *Economic Journal*, **118**(532), 1596-1620.

Formby J. P., W. D. Gunther, and R. Sakano (1993), "Entry level salaries of academic economists: does gender of age matter?" *Economic Inquiry*, **31**(1), 128-138.

Franceschet M. (2012), "Large-Scale Structure of Journal Citation Networks", *Journal of the American Society for Information Science and Technology*, **64**(4), 837-842.

Franck G. F. (1999), "Scientific Communication–A Vanity Fair?", *Science*, **286**(5437), 53-55.

Freeman L.C. (1979), "Centrality in Social Networks - Conceptual Clarification", *Social Networks*, **1**(3), 215-239.

Hall, B. H., A. B. Jaffe, and M. Trajtenberg (2001), "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools", *NBER Working Paper*, No. 8498.

Huberman B. (2003), "The Laws of the Web: Patterns in the Ecology of Information", MIT Press.

Jaffe, A. B., Trajtenberg, M. and R. Henderson (1993), "Geographic localization of knowledge spillovers as evidenced by patent citations", *Quarterly Journal of Economics*, **108**(3), 577-598.

Kalaitzidakis P., T. P. Mamuneas, and T. Stengos (2003), "Rankings of academic journals and institutions in economics", *Journal of the European Economic Association*, **1**(6), 1346 -1366.

Kelly M. A., and S. Bruestle, "Trends of subjects published in economics journals 1969-2007", *Economic Inquiry*, **49**(3), 658-673.

Kim E. H., A. Morse, and L. Zingales (2006), "What Has Mattered to Economics Since 1970", *Journal of Economic Perspectives*, **20**(4), 189-202.

Klamer A. and H. P. Van Dalen (2002), "Attention and the art of scientific publishing", *Journal of Economic Methodology*, **9**(3), 289-315.

Kretschmer H. (2004), "Author productivity and geodesic distance in bibliographic co-autorship networks, and visibility on the Web", *Scientometrics*, **60**(3), 409-420.

Lee P.-C., H.-N. Su, and F.-S. Wu (2010), "Quantitative mapping of patented technology - The case of electrical conducting polymer nano composite", *Technonoligical Forecasting and Social Change*, **77**(3), 466-478.

McCabe, M. J. and C. M. Snyder (2011), "Did online access change the economics literature?", *SSRN Working Paper Series*, Working Paper No. 1746243.

Newman M. E. J. (2001a), "The structure of scientific collaboration networks", *Proceedings of the National Academy of Sciences of the U.S.A.*, **98**(2), 404-409.

Newman M. E. J. (2001b), "Scientific collaboration networks. I. Network construction and fundamental results", *Physical Review E*, **64**(1), 016131.

Newman M. E. J. (2001c), "Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality", *Physical Review E*, **64**(1), 016132.

Newman M. E. J. (2003), "The Structure and Function of Complex Networks", *Society for Industrial and Applied Mathematics Review*, **45**(2), 167-256.

Palacios-Huerta I. and Volij O. (2004), "The Measurement of Intellectual Influence", *Econometrica*, **72**(3), 963–977.

Pencavel, J. (1991), "Editor's note", *Journal of Economic Literature*, **29**(1), v.

Poincaré H. (1910), "Mathematical Creation", *The Monist*, **20**(3), 321-335.

Stirling A. (2007), "A general framework for analysing diversity in science, technology and society", *Journal of Royal Society Interface*, **4**, 707-719.

Tenopir C., B. Hitchcock and A. Pillow (2003), *Use and Users of Electronic Library Resources: An Overview and Analysis of Recent Research Studies*, Washington, D.C.: Council on Library and Information Resources.

Trajtenberg M., R. Henderson, A. B. Jaffe (1992), "Ivory Tower Versus Corporate Lab: An Empirical Study of Basic Research and Appropriability", *NBER Working Paper*, 4146.

Van den Bergh J. (2008) "Optimal diversity: Increasing returns versus recombinant innovation", *Journal of Economic Behavior & Organization*, **68**(3–4), 565–580.

Weitzman M. L. (1992), "On Diversity", *Quarterly Journal of Economics*, **107**(2), 363-405.

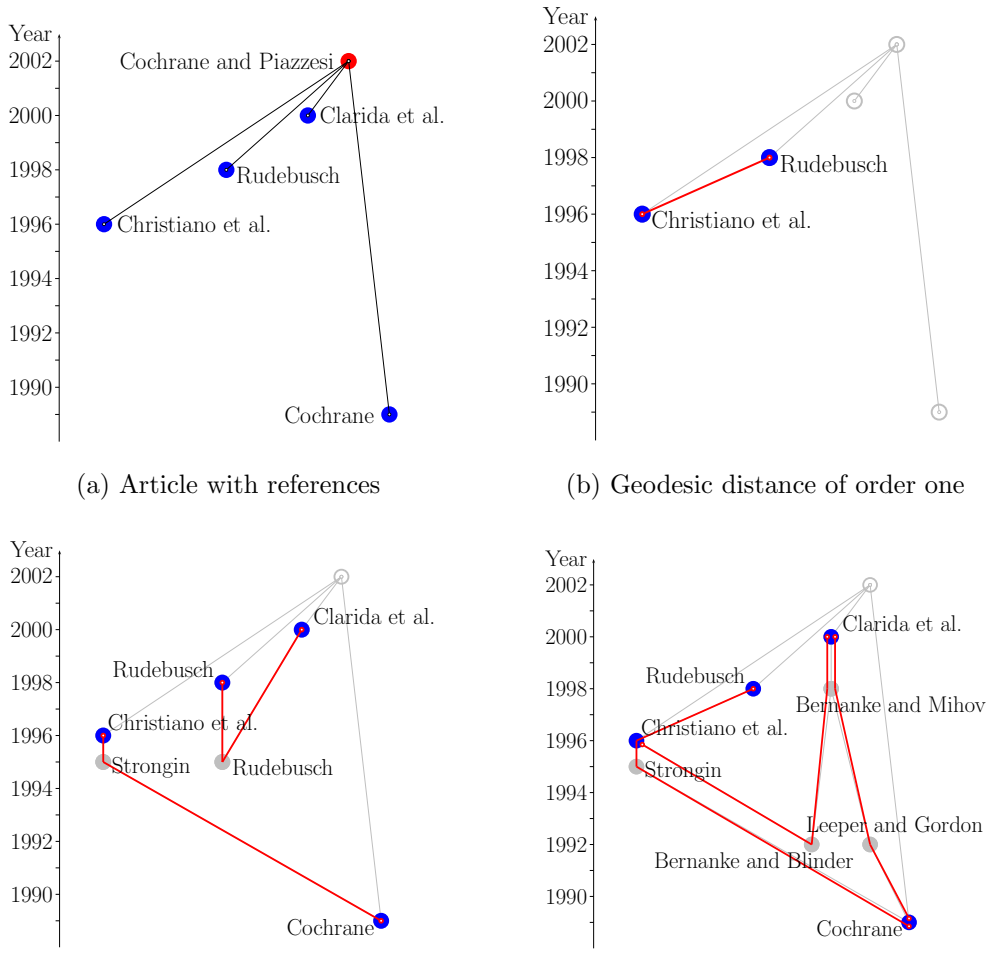
Weitzman M. L. (1996), "Hybridizing Growth Theory", *American Economic Review*, **86**(2), 207-212.

Weitzman M. L. (1998a), "Recombinant Growth", *Quarterly Journal of Economics*, **113**(2), 331-360.

Weitzman M. L. (1998b), "The Noah's Ark Problem", *Econometrica*, **66**(6), 1279-1298.

Yin L.-C., H. Kretschmer, R. A. Hanneman, and Z.-Y. Liu (2006), "The evolution of a citation network topology: The development of the journal Scientometrics", in: *Proceedings of the International Workshop on Webometrics, Informetrics and Scientometrics, and Seventh COLLNET meeting*, Nancy, France: SRDI-INIST-CNRS, 92-113.

Figures



(a) Article with references

(b) Geodesic distance of order one

(c) Geodesic distances of order two

(d) Geodesic distances of order three

Figure 1: **Geodesic distances of an article's references**

Notes: The Figure illustrates how geodesic distances of an article's references are obtained, using the article by John Cochrane and Monica Piazzesi, "The Fed and Interest Rates: A High-Frequency Identification", *American Economic Review, Papers and Proceedings*, May 2002, Volume 92, Issue 2, pp. 90-95. Panel (a) plots the article as a red node and the four references identified in the data as blue nodes. Panels (b), (c) and (d) plot shortest back-in-time citation paths between blue nodes (geodesic distances) as red lines. Blue nodes' references relevant for these paths are plotted in grey.

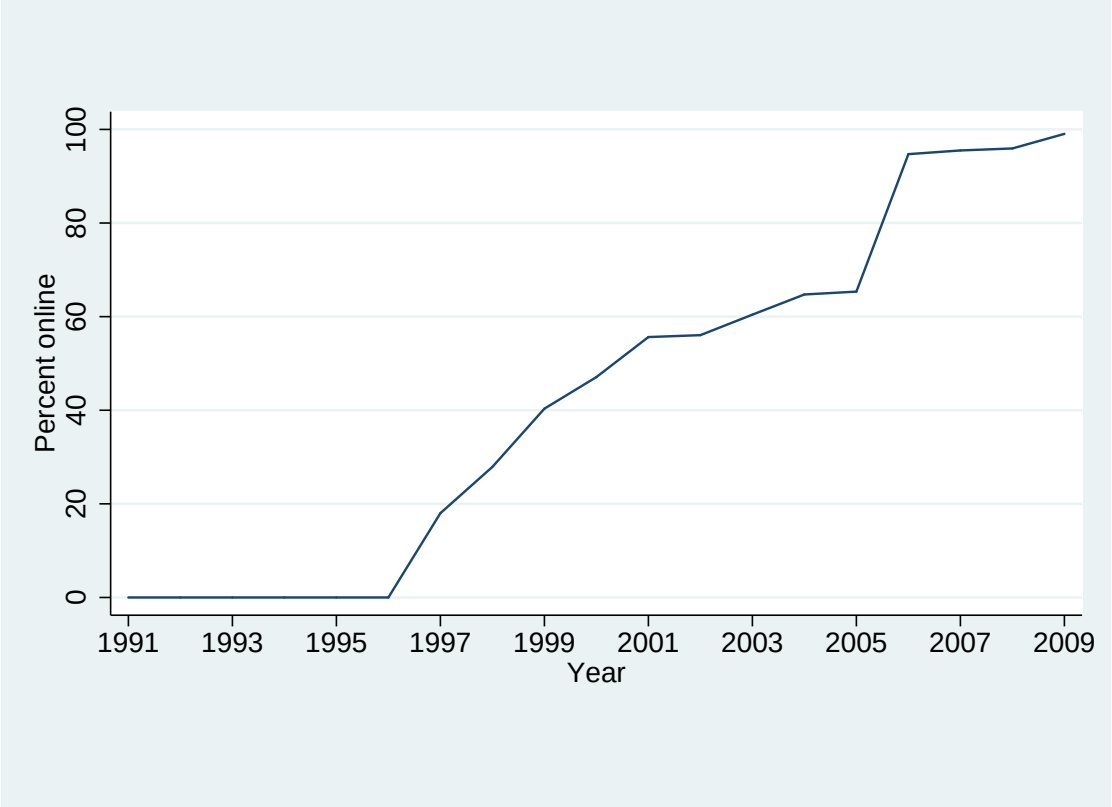


Figure 2: **The share of economics journal volumes accessible online**

Notes: The Figure plots for each year 1991-2009 the average share of existing volumes published in the 50 selected economics journals which was accessible online on at least one platform.

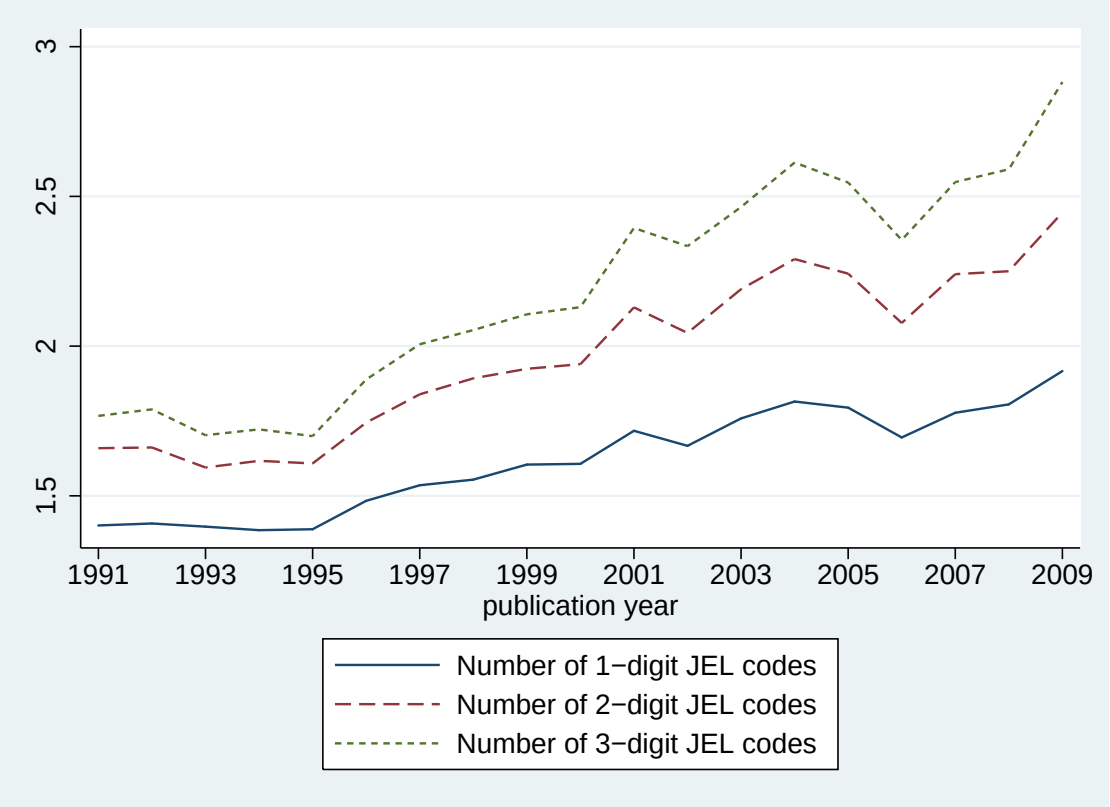


Figure 3: **The average number of JEL codes over time**

Notes: The figure plots the average number of one, two and three digit JEL codes. The sample includes the 45,553 articles published between 1991-2009 in the considered journals.

Tables

Table 1: Fixed effects regressions of percent online on diversity variables, N=45,553

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr(g=1)	Pr(g=2)	Pr(g=3)
I. *Regressions on percent online*						
Percent online	0.2208***	0.3426***	0.3565***	-0.0670***	-0.0027	0.0589***
	(0.0274)	(0.0360)	(0.0408)	(0.0120)	(0.0104)	(0.0079)
R^2	0.0482	0.0679	0.0921	0.0044	0.0055	0.0219
II. *Regressions on percent online and further control variables*						
Percent online	0.1792***	0.2823***	0.2955***	-0.0404***	-0.0344***	0.0356***
	(0.0265)	(0.0350)	(0.0395)	(0.0113)	(0.0101)	(0.0076)
R^2	0.0727	0.0928	0.1135	0.1460	0.1018	0.0854

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator accounting for journal-year fixed effects (859 groups). Robust standard errors clustered at journal-year level in parentheses. R^2 is the squared correlation between dependent variable and prediction. Further control variables in Panel II: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

Table 2: Author and co-author-groups fixed effects

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr($g=1$)	Pr($g=2$)	Pr($g=3$)
I. *Regressions with author fixed effects,* $N = 67,903$						
Percent online	0.0606**	0.1409***	0.1473***	-0.0213*	-0.0277**	0.0329***
	(0.0300)	(0.0385)	(0.0420)	(0.0123)	(0.0108)	(0.0092)
R^2	0.1617	0.1893	0.2283	0.1887	0.1676	0.1118
II. *Regressions with author-group fixed effects,* $N = 21,767$						
Percent online	0.0512	0.1068*	0.1612**	-0.0009	-0.0165	0.0242
	(0.0418)	(0.0571)	(0.0640)	(0.0188)	(0.0157)	(0.0148)
R^2	0.1044	0.1404	0.1789	0.1475	0.1031	0.0685

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator. Regressions in Panel I account for author fixed effects (12,165 groups). Panel I standard errors (in parentheses) robust to heteroskedasticity and clustering at author (12,165 groups) and article level (41,441 groups). Regressions in Panel II account for co-author-group fixed effects (7,307 groups). Panel II standard errors (in parentheses) robust to heteroskedasticity and clustering at co-author-group level. Further control variables in both Panels: complete set of journal-year indicators, paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

Table 3: Treatment interacted with time trend, $N = 45,553$

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr($g=1$)	Pr($g=2$)	Pr($g=3$)
Percent online	0.0366	0.1101*	0.0941	-0.0479**	-0.0265	0.0195
	(0.0484)	(0.0635)	(0.0696)	(0.0213)	(0.0162)	(0.0141)
Perc. online × year	0.0319***	0.0385***	0.0450***	0.0017	-0.0018	0.0036
	(0.0088)	(0.0122)	(0.0142)	(0.0044)	(0.0034)	(0.0030)
F-statistic	32.05	36.85	29.98	6.72	5.78	11.11
p-value	0.0000	0.0000	0.0000	0.0013	0.0032	0.0000
R^2	0.0705	0.0948	0.1247	0.1483	0.0966	0.0827

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator accounting for journal-year fixed effects (859 groups). Robust standard errors clustered at journal-year level in parentheses. The variable "Perc. online × year" is normalized to zero in 1997. F-statistics and p-values are for joint significance tests on coefficients of "Percent online" and "Perc. online × year". R^2 is the squared correlation between dependent variable and prediction. Further control variables: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

Table 4: Treatment interacted with journal type, $N = 45,553$

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr($g=1$)	Pr($g=2$)	Pr($g=3$)
Perc. online × top 5	0.1519***	0.2344***	0.2780***	-0.0239	-0.0498**	0.0170
	(0.0458)	(0.0577)	(0.0687)	(0.0214)	(0.0227)	(0.0133)
Perc. online × gen. interest	0.2613***	0.4154***	0.3856***	-0.0608**	-0.0311*	0.0391***
	(0.0673)	(0.1015)	(0.1102)	(0.0246)	(0.0184)	(0.0129)
Perc. online × field	0.1727***	0.2739***	0.2813***	-0.0443***	-0.0266**	0.0452***
	(0.0375)	(0.0481)	(0.0542)	(0.0149)	(0.0122)	(0.0107)
R^2	0.0699	0.0907	0.1130	0.1458	0.1021	0.0829

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator accounting for journal-year fixed effects (859 groups). Robust standard errors clustered at journal-year level in parentheses. R^2 is the squared correlation between dependent variable and prediction. Further control variables: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

Table 5: Citing top 5 journals and own journal, $N = 45,553$

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr($g=1$)	Pr($g=2$)	Pr($g=3$)
Percent online	0.0928***	0.1742***	0.2183***	-0.0461***	-0.0096	0.0310***
	(0.0293)	(0.0380)	(0.0428)	(0.0115)	(0.0106)	(0.0079)
Percent refs. to top 5	0.2478***	0.3589***	0.2448***	0.0431***	-0.0982***	0.0137
	(0.0420)	(0.0529)	(0.0569)	(0.0119)	(0.0101)	(0.0085)
Perc. refs. to own journal	-0.2895***	-0.1569**	-0.1597**	0.0929***	-0.0313**	-0.0136
	(0.0465)	(0.0609)	(0.0663)	(0.0158)	(0.0128)	(0.0107)
R^2	0.0728	0.0929	0.1090	0.1432	0.1218	0.0859

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator accounting for journal-year fixed effects (859 groups). Robust standard errors clustered at journal-year level in parentheses. R^2 is the squared correlation between dependent variable and prediction. Further control variables: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

Table 6: Average citation lag, $N = 45,553$

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr($g=1$)	Pr($g=2$)	Pr($g=3$)
Percent online	0.1820***	0.2878***	0.3052***	-0.0396***	-0.0302***	0.0380***
	(0.0266)	(0.0351)	(0.0395)	(0.0113)	(0.0101)	(0.0076)
Average citation lag	-0.0019***	-0.0037***	-0.0067***	-0.0005**	-0.0029***	-0.0016***
	(0.0007)	(0.0009)	(0.0011)	(0.0002)	(0.0002)	(0.0002)
R^2	0.0732	0.0937	0.1154	0.1453	0.1075	0.0917

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator accounting for journal-year fixed effects (859 groups). Robust standard errors clustered at journal-year level in parentheses. R^2 is the squared correlation between dependent variable and prediction. Further control variables: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

Appendix

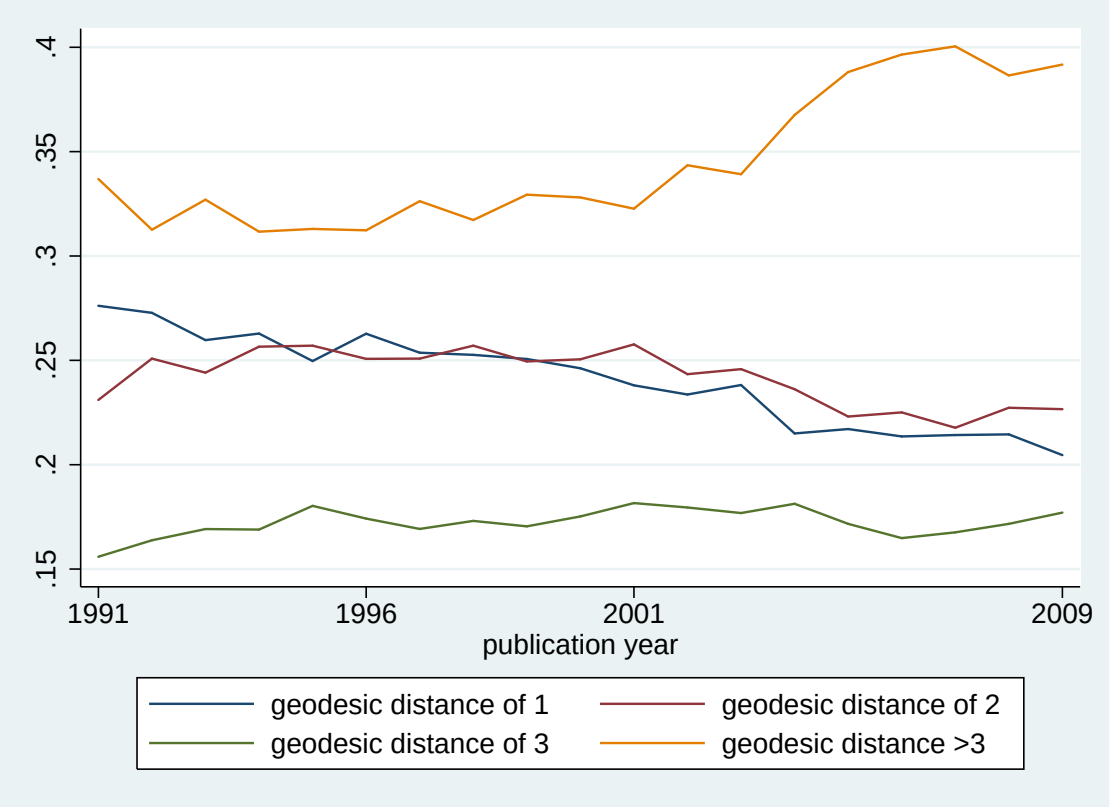


Figure 4: **The average shares of geodesic distances over time**

Notes: The figure plots the average share of the different geodesic distances after controlling for the time trend explained by the fraction of items for which we observe less then 2 references (i.e. geodesic distances cannot be calculated). The sample includes all items published between 1991-2009 without missing values. More precisely, the figure plots the residuals of regressing (in the full sample 1955-2009) the average share of geodesic distances of length $s = 1, 2, 3, > 3$ in year t on a constant and the share of items in year t with less than 2 observed references. In order to visualize the level of the series the predicted value of the simple regression in the year 1991 has been added to the residuals (i.e. in 1991 the values correspond to the one in the raw data).

Table 7: Baseline regression (Panel II of Table 1) – Full output, $N = 45,553$

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr($g=1$)	Pr($g=2$)	Pr($g=3$)
Percent online	0.1792***	0.2823***	0.2955***	-0.0404***	-0.0344***	0.0356***
	(0.0265)	(0.0350)	(0.0395)	(0.0113)	(0.0101)	(0.0076)
Proceedings paper	0.0074	0.0409	0.0442	-0.0035	-0.0021	0.0066**
	(0.0178)	(0.0256)	(0.0299)	(0.0039)	(0.0033)	(0.0028)
No. authors	0.0080*	-0.0009	-0.0026	-0.0050***	0.0007	0.0023**
	(0.0045)	(0.0055)	(0.0059)	(0.0012)	(0.0009)	(0.0009)
No. pages	0.0044***	0.0084***	0.0115***	-0.0015***	0.0004***	0.0009***
	(0.0006)	(0.0008)	(0.0009)	(0.0001)	(0.0001)	(0.0001)
No. references	-0.0015***	-0.0011**	-0.0008	-0.0008***	0.0006***	0.0001
	(0.0004)	(0.0005)	(0.0006)	(0.0001)	(0.0001)	(0.0001)
No. journals referenced	0.0296***	0.0342***	0.0330***	-0.0211***	-0.0011*	0.0093***
	(0.0024)	(0.0030)	(0.0033)	(0.0007)	(0.0006)	(0.0004)
Perc. refs. in data	-0.2641***	-0.2537***	-0.2710***	0.1201***	0.2428***	0.0193***
	(0.0283)	(0.0359)	(0.0382)	(0.0085)	(0.0082)	(0.0067)
Perc. self-refs.	-0.1449***	-0.2143***	-0.2955***	0.3043***	0.0077	-0.0943***
	(0.0533)	(0.0700)	(0.0753)	(0.0224)	(0.0162)	(0.0131)
Avg. ref.'s cit. $\times 10^{-2}$	-0.0032**	-0.0061***	-0.0100***	0.0037***	0.0026***	-0.0008*
	(0.0014)	(0.0018)	(0.0021)	(0.0006)	(0.0005)	(0.0004)
Avg. ref.'s refs. $\times 10^{-1}$	0.0212***	0.0237***	0.0222***	0.0104***	0.0207***	0.0092***
	(0.0034)	(0.0042)	(0.0044)	(0.0013)	(0.0012)	(0.0010)
R^2	0.0727	0.0928	0.1135	0.1460	0.1018	0.0854

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator accounting for journal-year fixed effects (859 groups). Robust standard errors clustered at journal-year level in parentheses. R^2 is the squared correlation between dependent variable and prediction.

Table 8: Different lags of treatment, $N = 45,553$

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr($g=1$)	Pr($g=2$)	Pr($g=3$)
I. *Regressions using contemporaneous treatment*						
Percent online	0.1518***	0.2452***	0.2662***	-0.0453***	-0.0311***	0.0308***
	(0.0267)	(0.0344)	(0.0391)	(0.0110)	(0.0104)	(0.0080)
R^2	0.0696	0.0886	0.1078	0.1450	0.1039	0.0854
II. *Regressions using treatment lagged by one year (baseline treatment)*						
Percent online	0.1792***	0.2823***	0.2955***	-0.0404***	-0.0344***	0.0356***
	(0.0265)	(0.0350)	(0.0395)	(0.0113)	(0.0101)	(0.0076)
R^2	0.0727	0.0928	0.1135	0.1460	0.1018	0.0854
III. *Regressions using treatment lagged by two years*						
Percent online	0.1738***	0.2659***	0.2890***	-0.0308***	-0.0372***	0.0343***
	(0.0264)	(0.0334)	(0.0378)	(0.0104)	(0.0100)	(0.0076)
R^2	0.0727	0.0920	0.1140	0.1491	0.1009	0.0858
IV. *Regressions using treatment lagged by three years*						
Percent online	0.1522***	0.2347***	0.2401***	-0.0176*	-0.0370***	0.0338***
	(0.0237)	(0.0307)	(0.0357)	(0.0093)	(0.0081)	(0.0068)
R^2	0.0705	0.0891	0.1077	0.1524	0.1025	0.0860

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator accounting for journal-year fixed effects (859 groups). Robust standard errors clustered at journal-year level in parentheses. R^2 is the squared correlation between dependent variable and prediction. Further control variables: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

Table 9: Alternative treatments, $N = 45,553$

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	$\Pr(g=1)$	$\Pr(g=2)$	$\Pr(g=3)$
I. *Regressions using average percent online as treatment*						
Avg. perc. online	0.1736***	0.2673***	0.2812***	-0.0175*	-0.0473***	0.0255***
	(0.0272)	(0.0368)	(0.0420)	(0.0100)	(0.0091)	(0.0073)
R^2	0.0723	0.0923	0.1129	0.1523	0.0963	0.0865
II. *Regressions using percent of references online as treatment*						
Perc. refs. online	0.0884***	0.1273***	0.1415***	-0.0025	-0.0291***	0.0109*
	(0.0243)	(0.0306)	(0.0355)	(0.0088)	(0.0082)	(0.0060)
R^2	0.0627	0.0762	0.0912	0.1544	0.1057	0.0848
III. *Regressions using number of volumes online as treatment*						
Vols. online $\times 10^{-1}$	0.0046***	0.0074***	0.0087***	0.0000	-0.0008***	0.0002***
	(0.0005)	(0.0006)	(0.0007)	(0.0001)	(0.0001)	(0.0001)
R^2	0.0777	0.1009	0.1256	0.1546	0.1087	0.0843
IV. *Regressions using average number of volumes online as treatment*						
Avg. vols. online	0.0033***	0.0048***	0.0051***	-0.0002**	-0.0005***	0.0003***
	(0.0003)	(0.0004)	(0.0004)	(0.0001)	(0.0001)	(0.0001)
R^2	0.0838	0.1064	0.1282	0.1536	0.1064	0.0861

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator accounting for journal-year fixed effects (859 groups). Robust standard errors clustered at journal-year level in parentheses. R^2 is the squared correlation between dependent variable and prediction. Further control variables: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

Table 10: Constant elasticity models

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr($g=1$)	Pr($g=2$)	Pr($g=3$)
I. *Poisson PML estimates,* $N = 45,553$						
Percent online	0.1165***	0.1544***	0.1448***	-0.1034***	-0.1304***	0.2400***
	(0.0173)	(0.0191)	(0.0194)	(0.0359)	(0.0395)	(0.0478)
R^2	0.0731	0.0940	0.1159	0.1484	0.1008	0.0828
II. *OLS estimates of logarithmized dependent variable*						
Percent online	0.1067***	0.1527***	0.1459***	-0.2326***	-0.2459***	0.0536
	(0.0154)	(0.0178)	(0.0189)	(0.0262)	(0.0306)	(0.0380)
R^2	0.0714	0.0915	0.1139	0.2631	0.0467	0.0273
N	45,553	45,553	45,553	41,828	40,208	37,705

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. Panel I regressions estimated by the fixed effects Poisson Pseudo-Maximum Likelihood estimator, Panel II estimated by the OLS within-estimator. Both estimators account for journal-years fixed effects (859 groups). Robust standard errors in parentheses, clustered at journal-year level in Panel II. R^2 is the squared correlation between dependent variable and prediction. Further control variables: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

Table 11: Refining journal fixed effects

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr($g=1$)	Pr($g=2$)	Pr($g=3$)
I. *Regressions with journal-document-type-year fixed effects,* $N = 45,553$						
Percent online	0.1705***	0.2691***	0.2845***	-0.0397***	-0.0333***	0.0358***
	(0.0268)	(0.0356)	(0.0400)	(0.0114)	(0.0102)	(0.0077)
R^2	0.0733	0.0931	0.1140	0.1461	0.1023	0.0851
II. *Regressions with journal-issue-year fixed effects,* $N = 44,937$						
Percent online	0.1543***	0.2564***	0.2779***	-0.0340***	-0.0345***	0.0340***
	(0.0284)	(0.0369)	(0.0418)	(0.0116)	(0.0104)	(0.0080)
R^2	0.0695	0.0888	0.1094	0.1466	0.1018	0.0827

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator. Panel I accounts for journal-document-type-year fixed effects (1,456 groups), Panel II for journal-issues-years fixed effects (4,817 groups). Robust standard errors clustered at journal-year level in parentheses. R^2 is the squared correlation between dependent variable and prediction. Further control variables: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

Table 12: Treatment defined over different online platforms, $N = 45,553$

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr($g=1$)	Pr($g=2$)	Pr($g=3$)
Perc. online, JSTOR	0.1975***	0.3070***	0.3209***	-0.0274***	-0.0351***	0.0297***
	(0.0234)	(0.0308)	(0.0344)	(0.0088)	(0.0083)	(0.0064)
Perc. online, FSO	0.2043***	0.2560***	0.3108***	-0.0521***	-0.0338***	0.0161*
	(0.0480)	(0.0635)	(0.0722)	(0.0145)	(0.0123)	(0.0095)
R^2	0.0802	0.1034	0.1317	0.1417	0.0947	0.0865

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator accounting for journal-year fixed effects (859 groups). Robust standard errors clustered at journal-year level in parentheses. R^2 is the squared correlation between dependent variable and prediction. Further control variables: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

Table 13: References' age distribution, $N = 45,553$

	JEL codes			Geodesic distances		
	1-digit	2-digit	3-digit	Pr($g=1$)	Pr($g=2$)	Pr($g=3$)
I. *Average citation lag and citation lag's standard deviation*						
Percent online	0.1827***	0.2883***	0.3058***	-0.0398***	-0.0296***	0.0381***
	(0.0267)	(0.0351)	(0.0395)	(0.0113)	(0.0101)	(0.0076)
Average citation lag	-0.0030***	-0.0045***	-0.0077***	-0.0003	-0.0038***	-0.0018***
	(0.0009)	(0.0012)	(0.0014)	(0.0003)	(0.0003)	(0.0002)
Lag's std. dev.$\times 10^{-2}$	0.0033**	0.0025	0.0030	-0.0008	0.0028***	0.0004
	(0.0016)	(0.0023)	(0.0026)	(0.0006)	(0.0005)	(0.0003)
R^2	0.0737	0.0939	0.1158	0.1456	0.1069	0.0917
II. *Median citation lag*						
Percent online	0.1822***	0.2867***	0.3026***	-0.0401***	-0.0309***	0.0375***
	(0.0266)	(0.0351)	(0.0396)	(0.0113)	(0.0102)	(0.0077)
Median citation lag	-0.0024***	-0.0036***	-0.0058***	-0.0002	-0.0028***	-0.0015***
	(0.0007)	(0.0010)	(0.0011)	(0.0002)	(0.0002)	(0.0002)
R^2	0.0736	0.0939	0.1156	0.1457	0.1056	0.0899

Notes: *, ** and *** indicate statistical significance at 10%, 5% and 1% level. All regressions estimated by the OLS within-estimator accounting for journal-year fixed effects (859 groups). Robust standard errors clustered at journal-year level in parentheses. R^2 is the squared correlation between dependent variable and prediction. Further control variables: paper-and-proceedings indicator, number of authors, number of pages, number of references, number of journals referenced, percent references in data, percent self-references, average number of references' citations, average number of references' references.

Table 14: Descriptive statistics of data used in estimation, $N = 45,553$

Variable	Short description	Mean	Std. Dev.	Min./Max.
I. *Diversity measures (dependent variables)*				
1-digit JEL code	Number of 1-digit JEL codes assigned to article	1.64	0.75	1 / 6
2-digits JEL code	Number of 2-digit JEL codes assigned to article	2.01	0.99	1 / 7
3-digits JEL code	Number of 3-digit JEL codes assigned to article	2.25	1.14	1 / 8
Pr($g=1$)	Fraction of article's references with geodesic distance 1	0.2464	0.2055	0 / 1
Pr($g=2$)	Fraction of article's references with geodesic distance 2	0.2707	0.1792	0 / 1
Pr($g=3$)	Fraction of article's references with geodesic distance 3	0.1983	0.1523	0 / 1
Pr($g>3$)	Fraction of article's refs. with geod. dist. greater than 3	0.2845	0.2511	0 / 1
II. *Online accessibility measures (treatment variables)*				
Percent online	Fraction of volumes accessible online in year prior to article's publication out of all existing volumes in distinct journals cited by article	0.5731	0.4169	0 / 1
Percent online, JSTOR	Percent online through JSTOR	0.4929	0.3660	0 / 1
Percent online, FSO	Percent online through other major online platforms	0.2542	0.3069	0 / 1
Percent online, no lag	Percent online in article's publication year	0.624	0.4068	0 / 1
Perc. online, 2-year lag	Percent online two years prior to article's publication	0.5188	0.4188	0 / 1
Perc. online, 3-year lag	Percent online three years prior to article's publication	0.4556	0.4086	0 / 1
Average percent online	Percent online in average journal cited by article	0.5005	0.3965	0 / 1
No. volumes online	Number of volumes accessible online in year prior to publication in all distinct journals cited by article	209.71	206.10	0 / 1,073
Avg. no. volumes online	No. volumes online in average journal cited by article	34.27	28.18	0 / 123
Percent refs. online	Fraction of article's references accessible online in year prior to article's publication	0.4922	0.4068	0 / 1
III. *Control variables*				
Proceedings paper	=1 if article is a proceedings paper	0.0937	0.2914	0 / 1
No. authors	Article's number of authors	1.8287	0.8609	1 / 26
No. pages	Article's number of pages	20.10	10.12	1 / 96
No. journals referenced	Number of distinct journals cited in article	5.70	2.81	1 / 21
No. references	Number of references cited in article	27.15	16.30	2 / 537
Percent refs. in data	Fraction of article's references contained in data	0.4444	0.1886	0.017 / 1
Percent self-references	Fraction of references w. at least 1 of article's author	0.0355	0.0596	0 / 1
Top 5 journal	=1 if article was published in top5 journal	0.2018	0.4014	0 / 1

Continued on next page...

... table 14 continued

Variable		Mean	Std. Dev.	Min./Max.
General interest journal	=1 if article was published in general interest journal	0.2425	0.4286	0 / 1
Field journal	=1 if article was published in field journal	0.5557	0.4969	0 / 1
Percent refs. to top 5	Fraction of article's refs. published in top 5 journals	0.1881	0.1396	0 / 1
Perc. refs. to own journal	Fraction of article's refs. published in same journal as article	0.0710	0.0930	0 / 1
Average citation lag	Avg. difference between article's and refs.' publication year	11.63	5.74	0 / 100.2
Median citation lag	Median diff. between article's and refs.' publication year	8.84	5.11	0 / 89
Std. dev. of cit. lag	Std. dev. of diff. between article's and refs.' publication year	126.32	276.21	0 / 9,528.9
Average ref.'s references	Average number of references of a reference in article	27.85	11.40	0.50 / 290.5
Average ref.'s citations	Avg. number of citations received by a reference in article	227.51	241.52	1 / 4,104

Table 15: Alphabetic list of the selected journals

No.	Journal	No. of items
1	American Economic Review	11,246
2	Bell Journal of Economics	542
3	Econometric Theory	1,288
4	Econometrica	6,039
5	Economic Inquiry	1,892
6	Economic Journal	9,397
7	Economic Theory	1,402
8	Economics Letters	7,115
9	European Economic Review	2,992
10	Games and Economic Behavior	1,436
11	International Economic Review	1,898
12	International Journal of Game Theory	745
13	Journal of Applied Econometrics	1,029
14	Journal of Business Economic Statistics	1,334
15	Journal of Development Economics	2,356
16	Journal of Econometrics	2,705
17	Journal of Economic Behavior and Organization	2,464
18	Journal of Economic Dynamics and Control	2,072
19	Journal of Economic Growth	146
20	Journal of Economic History	10,355
21	Journal of Economic Literature	6,916
22	Journal of Economic Perspectives	1,329
23	Journal of Economic Theory	3,359
24	Journal of Environmental Economics and Management	1,434
25	Journal of Finance	6,979
26	Journal of Financial and Quantitative Analysis	1,988
27	Journal of Financial Economics	1,629
28	Journal of Financial Intermediation	290
29	Journal of Health Economics	1,208
30	Journal of Human Resources	1,786
31	Journal of International Economics	2,305
32	Journal of Labor Economics	834
33	Journal of Mathematical Economics	1,228
34	Journal of Monetary Economics	2,056
35	Journal of Political Economy	4,880
36	Journal of Public Economics	2,633
37	Journal of Risk and Uncertainty	566
38	Journal of the European Economic Association	331
39	Journal of Urban Economics	1,684
40	Oxford Bulletin of Economics and Statistics	1,449
41	Quarterly Journal of Economics	2,677
42	RAND Journal of Economics	1,113
43	Review of Economic Studies	2,338
44	Review of Economics and Statistics	4,429
45	Review of Financial Studies	916
46	Scandinavian Journal of Economics	1,547
47	Social Choice and Welfare	1,143
48	Swedish Journal of Economics	326
49	Western Economic Journal	672
50	World Bank Economic Review	647
	Total number of items 1955-2009	129,145

Notes: "Bell Journal of Economics" includes its predecessor "The Bell Journal of Economics and Management Science" 1970-1974. "Oxford Bulletin of Economics and Statistics" includes its predecessor "Oxford University Bulletin of the Institute of Economics and Statistics" 1939-1972. Our sample includes three historical, non-successive journals: "Bell Journal of Economics" 1970-1983, "Swedish Journal of Economics" 1965-1975 and the "Western Economic Journal" 1962-1972. We ignore the non-English predecessor of the "Swedish Journal of Economics" - "Ekonomisk Tidskrift" - which goes back to 1899.
The list includes all journals considered in the standard Tilbourg ranking as well as the list considered in Palacios-Huerta and Volij (2004).
Some isolated publication years are missing, since they are not included in the Web of Science.