Online Appendix for “Economic Research Evolves: Fields and Styles”

Joshua Angrist, MIT Department of Economics and NBER
Pierre Azoulay, Sloan School of Management and NBER
Glenn Ellison, MIT Department of Economics and NBER
Ryan Hill, MIT Department of Economics
Susan Feng Lu, Purdue University and Northwestern University

January 2017

This paper uses proprietary data from the Thomson Reuters *Web of Science* (WoS) citation database and from the American Economic Association's *EconLit*.

Appendix A The Economics Journal List

The journal list used here comes from a classification scheme developed for our study of how other scientific disciplines cite economics research (This project is described in our working paper, Angrist et al. (2017)). Each discipline's journal list is constructed by identifying the journals cited most often by a disciplinary "flagship journal" in 1968, 1978, 1988, 1998, or 2008. The economics flagship is *The American Economic Review*. We modify the initial list by moving journals between disciplines to produce a final disciplinary journal list according to rules detailed in the data appendix to our working paper. These rules associate journals that appear initially on more than one list with the discipline to which they are most important.

The final economics journal list is reproduced in Table A1 of this appendix, which shows journals sorted by the average-across-years fraction of the AER's citations they receive. Table A1 also lists this average citation rate. Journals at the bottom of the list receive few citations, suggesting our analysis should be robust to variations in the length of the journal list.

Appendix B Constructing Journal Weights

Many of our analyses use time-varying journal weights w_j^t designed to reflect the relative importance of journal j in year t. These weights are constructed as follows. First, we compute preliminary importance weights $\tilde{\mu}_k^t$ for each top six economics journal k.[1] These weights are defined via a procedure inspired by Google "page rank": Let A^t be the 6×6 matrix with entries A_{kj}^t equal to the fraction of journal j's citations to all top six journals in year t made to journal k; and let μ^t be the solution to $\mu^t = dA^t\mu^t + \frac{1-d}{6}\mathbf{1}$, i.e. $\mu^t = (I - dA^t)^{-1}\frac{1-d}{6}\mathbf{1}$, where $d = 0.85$. We next set $\tilde{w}_j^t \equiv \sum_k \mu_k^t c_{kj}^t$, where the sum is taken over the top six journals k, and c_{kj}^t is the number of citations from journal k to journal j in year t as a fraction of all year t citations from journal k to journals in our full economics list. The final w_j^t series is the five-year moving averages of the $\tilde{w}_j^t$. The resulting weights are plotted in Figure 1 in the paper.

[1] The top six journals are *American Economic Review*, *Econometrica*, *Journal of Political Economy*, *Quarterly Journal of Economics*, *Review of Economic Studies*, and *Review of Economics and Statistics*.

Appendix C Field Classification

C.1 Overview

Our field classification starts by classifying articles into one of 17 "initial fields," using the article's Journal of Economic Literature classification (JEL) codes reported in *EconLit*. We follow the mapping of JEL codes to fields used by Ellison (2002). Many papers have multiple JEL codes. We therefore use a machine learning procedure to assign a single initial field to each paper with multiple codes.

The second step uses each paper's initial field classification and the initial field of the papers each paper cites to form 10 clusters. These clusters, constructed using the k-means algorithm, become our "final fields". Information on cited papers comes from the WoS.

C.2 Data Sources

We classify *EconLit* papers published in journals on the economics journal list in the period 1970-2015. *EconLit* provides bibliographic information, JEL codes, and keywords for most of these papers. Our copy of *Econlit* has 199,520 articles published between 1886 and 2016. Restricting this file to papers published from 1970-2015 and dropping papers without JEL codes leaves a classification database containing 168,133 papers.

C.2.1 Incorporating Citation Data

The WoS includes 214,312 articles in our journal list published from 1970-2015. There is no unique identifier common to WoS and *EconLit*. We therefore start by matching each article's journal issn, publication year, volume, issue, start page number, and end page number. This generates 139,237 matches. An additional 12,110 papers are matched on title and author (after removing capitalization, punctuation, common speech articles and author first names). Finally we execute a Stata `reclink` fuzzy merge using issn, year, volume, issue, start page, end page, and author last names. We evaluate these fuzzy matches manually based on the match score and title. The final matched sample contains 153,614 articles. The analysis reported in the *Papers and Proceedings* article uses the 134,892 articles published from 1980-2015.

C.3 Classification into Initial Fields

Our 17 initial fields are microeconomics, macroeconomics, public finance, labor, industrial organization, development, urban economics, environmental, econometrics, finance, international, experimental (lab), economic history, political economy, productivity, law and economics, and other. Each JEL code is mapped to a field using the scheme in Ellison (2002). Each article is assigned an initial field using machine learning as described below.

C.3.1 Training Data

We assembled a training dataset that exploits the fact that between 1991 and 2004, JEL codes typically appear in *EconLit* in order of importance rather than alphabetically. We therefore assigned fields using the first JEL code for papers published in these years. Our machine learning (ML) algorithm treats fields assigned this way as a dependent variable, to be predicted using the full set of up to 7 (unordered) JEL codes as well as article titles and keywords. Training articles in widely recognized field journals (like the *Journal of Labor Economics*) were subject to a "field journal override" before running the ML classifier. Articles with a single JEL code were omitted from the training data because for these articles, the set of JEL codes is perfectly informative. Training data with these articles included would far over-represent the prevalence of single-code fields, generating a misleadingly high success rate. Although single-JEL papers are not in the training data, they were classified by the ML model to take advantage of information in titles and keywords.

C.3.2 Classification Algorithm

The training data set was used to train a random forest classifier for multi-JEL papers (Breiman, 2001). Predictors include (up to 7) fields for (up to 7) JEL codes, dummies for words occurring in the title, and dummies for keywords.[2] Words occurring in the titles and keywords of more than 50% of articles or fewer than .5% of articles were excluded. Titles were preprocessed such that words were tagged by part of speech and converted into a normal form (lemmatized) and geopolitical entities were also tagged.[3] Preprocessing uses standard procedures in the Python Natural Language Toolkit

[2] Classification and coding uses the Python "Scikit-learn" package (Pedregosa et al., 2011).

[3] Lemmatization replaces the words "is," "were," and "am" in a sentence with the word "be." Lemmatization uses the NLTK `pos-tag` procedure, converting part-of-speech tags to the WordNet format, and then uses the NLTK `wordnet.lemmatize` procedure.

(Bird, Klein and Loper, 2009). Numbers were also replaced by a word indicating their type (e.g. year, decimal, fraction, percentage, integer).

We classified papers into fields using the Random Forest algorithm because it performed well in cross-validation comparisons with other schemes.[4] Our classifier consists of 500 trees with 30% of covariates sampled for each tree, with each tree trained to classify a sample of articles drawn uniformly at random (with replacement) from the set of all articles.[5] In a 90-10 split sample test, the algorithm with these parameters classified 94.2% of training articles correctly.

C.4 Classification into Final Fields

Ten final fields were constructed by clustering the 17 initial fields using a k-means algorithm that looks at each paper's initial field and the initial fields of the papers it cites.

C.4.1 Clustering Procedure

For each article i, we generate a set of 17 dummies indicating the article's initial field ($\mathbf{1}\{\text{field} = f\}_i$) and a set of 17 variables that count the number of cited articles on article i's reference list for each field ($\#\text{cites}_{fi}$). We then weight these variables using the following procedure.

First a reference weight is defined:

$$w_i^{ref} = w^a \cdot (1 - w^b(1 - x_i))$$

where x_i is the percentage of reference list citations that were classified using the *EconLit* data. The weights w^a and w^b are preselected. After inspection of classification results, we use $w^a = 0.65$ and $w^b = 0.3$

Next we define the own-field weight:

$$w_i^{\text{own}} = 1 - w_i^{\text{ref}}$$

[4]Algorithms compared include logistic regression (with L1 and L2 penalty), support vector machines (with L1 and L2 penalty), binary classification trees, the naive bayes algorithm, and k-nearest-neighbor classification.

[5]The large number of covariates per tree, a parameter chosen to minimize classification error in a 90-10 split-sample test, is consistent with the sparsity of our dataset.

Finally, we create 17 variables own_{fi} and 17 variables ref_{fi}

$$\text{own}_{fi} = \mathbf{1}\{\text{field} = f\} \cdot ({w_i^{\text{own}}}/{17})$$
$$\text{ref}_{fi} = (\text{share}_{fi} - \overline{\text{share}_f}) \cdot ({w_i^{\text{ref}}}/{17})$$

where $\text{share}_{fi} = \frac{\#\text{cites}_{fi}}{\sum_f \#\text{cites}_{fi}}$ is the fraction of articles in field f on the article's reference list, and $\overline{\text{share}_f}$ is the average over all articles for field f.

The variables own_{fi} and ref_{fi} are used as features in the k-means clustering algorithm (see Bishop (2006) for more on k-means). We used the Matlab package `kmeans`. A set of 18,423 articles with no references to other papers in our merged sample are clustered using only their initial own-field classification.

C.4.2 Classification of Development and Political Economy

We successfully classified the overwhelming majority of papers in fields that focus on roughly the same sorts of topics over time (Labor, Macroeconomics, Econometrics, etc.) Fields that have shifted focus proved harder to classify. We especially struggled with development and political economy; many recent development papers were initially classified as labor or public finance, while our ML routine classified many studies that are now considered political economy as macro or public finance. We believe this problem arises from the evolution of topics within these fields. Development economics has moved from studying growth and institutions in developing countries to a much broader set of topics. Modern development authors cite earlier development papers little, instead citing methodologically similar studies in labor and public finance. JEL codes are often chosen from these other fields as well. Political economy has also seen a sea change towards empirical papers that often make little or no connection with earlier work in the field.

To improve classification of development and political economy, we override the initial ML-assigned fields with a supplemental training sample. Specifically, we recoded the initial ML-assigned fields of some papers before processing them through the k-means algorithm. Papers with a JEL code beginning O1 or O2 were given a composite initial field that is .83 development and .17 whatever field the ML algorithm chose. Likewise, papers with a JEL code of D02 or D72-D78 were given an initial code of political economy using the same weighting scheme. These weights reflect our judgement of the intervention needed to classify modern papers in these fields correctly. In total we recode 13,050 articles published since 1990 (when the current alphanumeric JEL codes were introduced). The recoded papers were fed to k-means along with the rest of the papers classified initially to generate final fields.[6]

Appendix D Classification of Styles

D.1 Overview

We classify economics articles into three styles of research: (1) empirical, (2) theoretical, and (3) econometrics. Papers classified in the econometrics field are assigned the econometrics style. Remaining papers are classified as empirical or theoretical. As with classification into fields, style classification uses machine learning and a training data set. Specifically, style classification uses logistic ridge regression with inputs article titles, journal identifiers, fields, JEL codes, keywords, publication decade, and abstracts (where available). Also as in the field classification procedure, this algorithm was chosen after comparison of several algorithms.[7] The sample of papers classified into styles is a subset of those classified into fields, starting with papers published since 1980.

D.2 Training Data

Our training dataset contains a sample opf 5,850 hand-classified articles over-representing top journals. The training data include:

1. Articles originally classified by Ellison (2002). These papers are from 'top 6' economics journals and published from 1971-1998: 1,507 articles.

2. A sample of articles from the AER, JPE, and Econometrica:

    - AER, 1992-2004: 436 articles
    - Econometrica, 1998-2013: 822 articles
    - JPE, 1987-2014: 933 articles

3. Fifteen randomly chosen articles from each journal in our list published 1980-1989: 1,080 articles

[6]Examples affected by these overrides include Duflo, Hanna and Rya (2012), which our ML routine originally classified as labor and Acemoglu et al. (2008), which our ML routine originally classified as macro. The override moves these papers to development and political economy, .

[7]Algorithms compared include logistic regression (with L1 and L2 penalty), support vector machines (with L1 and L2 penalty), binary classification trees, the naive-Bayes algorithm, k-nearest-neighbor classification (with both standard and `word2vec` embeddings), and classification using a shallow convolutional neural network (Kim, 2014). We also compared the performance of various dimension reduction techniques, including filtering by the (univariate) ANOVA F-statistic, filtering by the χ^2-statistic for binary covariates, using LASSO for variable selection, and principal component analysis.

4. Fifteen randomly selected articles per journal per decade (1990-1999, 2000-2013) for top-20 journals based on cites from the AER. Five randomly selected articles per journal per decade for all other journals: 1,172 articles

D.3 Classification

The classification routine was trained to identify empirical papers. After empirical papers are identified, econometrics papers are removed, and remaining papers are classified as theoretical.

Roughly 30% of the articles in our classification dataset have no abstract. Not surprisingly, classification is more accurate with an abstract. We therefore first classified the full sample without using abstracts, then separately classified the subset of papers with abstracts using abstracts as a feature. The final classification gives precedence to the with-abstract classification where available.

Other data used by our classifier includes dummies for words occurring in $.001 - 50\%$ of titles, whether the title contained a question mark, keywords, fields assigned by the field classification procedure, journal names, and journal decade interactions. We also coded term-frequency minus inverse-document-frequency (TF-IDF) for words appearing in $.1 - 50\%$ of all abstracts, using only those articles that had an abstract. TF-IDF is a metric formed by dividing the frequency a word appears in, say, an article's title or abstract, by the frequency the word appears in titles or abstracts overall (Wu et al., 2008).[8]

We then fit a model of topics to the coded title and keyword data using Latent Dirichlet Allocation (LDA) (Blei, Ng and Jordan, 2003). Since titles contain only 10-15 words drawn from a vocabulary of about 20,000, they are highly sparse, and many informative words never appear in the training data. LDA is a popular dimension-reduction tool used in this scenario to better capture similarity between documents (in this case, titles). We fit a model of 10, 30, 50, 70, 90, 110, 130, and 200 topics, following past work in the natural language processing literature on the classification of short text (Chen, Jin and Shen, 2011). The resulting topic data was used in classification both with and without abstracts.

Finally, using these predictors, articles were classified using ridge logistic regression, with regularization parameter $\lambda = .0003$ for classification with abstract data (respectively $\lambda = .0005$ without abstract data). The regularization parameter was chosen to maximize accuracy in a 90-10 split sam-

[8] We compared the performance a number of data representations including TF-IDF, dummies for each word, and sums of `word2vec` embeddings (Mikolov et al., 2013) for the naive-Bayes algorithm, support vector machines, and logistic regression, before settling on our chosen representation. Comparisons were performed using a 90-10 split-sample test, as elsewhere.

ple validation test; the experiment was repeated 100 times for each potential choice of regularization parameter λ and the one producing the highest average accuracy was selected. For the 90-10 split sample test, our accuracy was 81.16% for classification without abstracts, and 87.14% with abstracts.

Classification accuracy was additionally checked by sampling 250 articles at random from the full sample and classifying these articles by hand to check the algorithm's output. Our success rate averaged 87% accurate with abstracts and 83% without. The average overall accuracy is 85.8%.

Table A2 reports the joint distribution of fields and styles for the sample of economics publications described in our figures. This table shows that papers in the microeconomics field are mostly (though not entirely) classified as theoretical, while papers in the "applied micro" fields of labor, development, and public finance are mostly empirical. On the other hand, papers in IO, also an applied micro field, tilt towards theory. Both the macro and international fields are somewhat more empirical, but each have a large theoretical share. The collection of smaller fields grouped under the miscellaneous heading (environmental, lab experiments, history, law and economics, political economy, productivity, urban, and unclassified) are nearly two-thirds empirical.

Table A1: Economics Journal List

Economics		
Journal	First Year Indexed	Importance
AMER ECON REV	1916	0.264
J POLIT ECON	1919	0.128
ECONOMETRICA	1934	0.088
QUART J ECON	1902	0.079
REV ECON STUD	1936	0.047
REV ECON STATIST	1950	0.032
J MONETARY ECON	1976	0.031
J ECON THEOR	1969	0.030
ECON J	1902	0.022
J ECON PERSPECT	1988	0.022
BELL J ECON	1970	0.022
J PUBLIC ECON	1976	0.019
RAND J ECON	1984	0.019
J ECON LIT	1969	0.018
J INT ECON	1972	0.014
J LAW ECON	1958	0.014
GAME ECON BEHAV	1991	0.013
J LABOR ECON	1983	0.011
ECONOMICA	1927	0.011
INT ECON REV	1960	0.011
J EUR ECON ASSOC	2005	0.010
J HUM RESOUR	1966	0.010
EUR ECON REV	1969	0.009
ECON INQ	1974	0.009
BROOKINGS PAP ECON ACTIV	1970	0.009
J ECONOMETRICS	1980	0.008
ECON LETT	1978	0.008
J ECON BEHAV ORGAN	1980	0.008
J MONEY CREDIT BANKING	1976	0.007
ANN ECON SOC MEAS	1974	0.007
J ECON HIST	1945	0.007
SOUTHERN ECON J	1956	0.006
REV ECON DYN	2001	0.006
IND LABOR RELAT REV	1956	0.005
CAN J ECON	1973	0.005
CARN ROCH CONF SERIES PUBLIC	1976	0.005
J LAW ECON ORGAN	1989	0.005
NAT TAX J	1956	0.005
J ECON DYN CONTROL	1980	0.004
J URBAN ECON	1974	0.004
J BUS ECON STAT	1985	0.004
J IND ECON	1956	0.004
J HEALTH ECON	1983	0.004
ECONOMIC THEORY	1995	0.004
OXFORD ECON PAP-NEW SER	1966	0.004
NBER MACROECON ANN	1987	0.004
J ENVIRON ECON MANAGE	1974	0.004
J LEGAL STUD	1973	0.003
INT J IND ORGAN	1987	0.003
J ECON MANAGE STRATEGY	1995	0.003
BELL J ECON MANAGE SCI	1971	0.003
AMER J AGR ECON	1968	0.003
EXPLOR ECON HIST	1969	0.002
KYKLOS	1956	0.002
ECON DEVELOP CULT CHANGE	1955	0.002
INT J GAME THEORY	1987	0.002
REV RADICAL POLIT ECON	1970	0.002
J REG SCI	1958	0.002
WORLD DEVELOP	1976	0.002
QUART REV ECON BUS	1966	0.002
PUBLIC POLICY	1956	0.002
SOC CHOICE WELFARE	1984	0.002
J MATH ECON	1980	0.002
J INT MONEY FINAN	1983	0.002
J ECON ISSUE	1967	0.002
AMER ECON	1970	0.002
ECON REC	1966	0.002
OXFORD BULL ECON STAT	1956	0.002
APPL ECON	1969	0.002
INT LAB REV	1932	0.001
THEOR DECIS	1970	0.001
REV INCOME WEALTH	1985	0.001
QUART REV ECON FINANC	1992	0.001
J INST THEOR ECON	1987	0.001
ENERGY J	1987	0.001
REV SOC ECON	1956	0.001
J REGUL ECON	1990	0.001
FED RESERVE BANK ST LOUIS REV	2004	0.001
ECONOMET THEORY	1988	0.001
J PROD ANAL	1994	0.001

Table A2: Classification of fields and styles

Economics Field	Research Style			
	Empirical	Metrics	Theoretial	Total
	(1)	(2)	(3)	(4)
Development Economics	9,075		1,523	10,598
Econometrics		8,820		8,820
Finance	4,346		2,947	7,293
Industrial Organization	5,911		6,655	12,566
International Economics	5,326		3,543	8,869
Labor Economics	10,776		2,520	13,296
Macroeconomics	11,446		8,875	20,321
Microeconomics	2,659		16,946	19,605
Public Finance	6,996		4,287	11,283
Miscellaneous	14,207		8,034	22,241
Total	70,742	8,820	55,330	134,892

Notes: Field by style distribution of papers published in major economics journals between 1980-2015.

References

Acemoglu, Daron, Simon Johnson, James A Robinson, and Pierre Yared. 2008. "Income and democracy." *The American Economic Review*, 98(3): 808–842.

Angrist, Josh, Pierre Azoulay, Glenn Ellison, and Susan Feng Lu. 2017. "Inside Job or Deep Impact? Using Extramural Citations to Assess Economic Scholarship." Unpublished manuscript.

Bird, Steven, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python.* " O'Reilly Media, Inc.".

Bishop, Christopher M. 2006. "Pattern recognition." *Machine Learning*, 128.

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. "Latent dirichlet allocation." *Journal of machine Learning research*, 3(Jan): 993–1022.

Breiman, Leo. 2001. "Random forests." *Machine learning*, 45(1): 5–32.

Chen, Mengen, Xiaoming Jin, and Dou Shen. 2011. "Short text classification improved by learning multi-granularity topics." 1776–1781, Citeseer.

Duflo, Esther, Rema Hanna, and Stephen P Ryan. 2012. "Incentives work: Getting teachers to come to school." *The American Economic Review*, 102(4): 1241–1278.

Ellison, Glenn. 2002. "The Slowdown of the Economics Publishing Process." *Journal of Political Economy*, 110(5): 947–993.

Kim, Yoon. 2014. "Convolutional neural networks for sentence classification." *arXiv preprint arXiv:1408.5882*.

Mikolov, Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. "Efficient estimation of word representations in vector space." *arXiv preprint arXiv:1301.3781*.

Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. "Scikit-learn: Machine learning in Python." *Journal of Machine Learning Research*, 12(Oct): 2825–2830.

Wu, Ho Chung, Robert Wing Pong Luk, Kam Fai Wong, and Kui Lam Kwok. 2008. "Interpreting tf-idf term weights as making relevance decisions." *ACM Transactions on Information Systems (TOIS)*, 26(3): 13.