

Political Language in Economics

Zubin Jelveh, Bruce Kogut, and Suresh Naidu *

August 19, 2017

Abstract

Do economists' political beliefs correlate with their academic research? We rely upon purely inductive methods in natural language processing and machine learning to examine patterns of implicit political ideology in economic articles. Using observed political behavior of economists and the phrases from their academic articles, we construct and validate a high-dimensional predictor of political ideology by economist. We document sorting of economists by predicted ideology into fields, departments and methodologies. Finally, we document a robust correlation between author ideology and magnitudes of reported policy relevant elasticities.

*We thank Daron Acemoglu, Jonathan Cantor, Navin Kartik, Max Kasy, Rajiv Sethi, Cosma Shalizi, Laurence Wilse-Samson, and seminar audiences at CIFAR, EMNLP, NYU, IAS, University of North Carolina, Columbia Business School, Harvard Business School, Princeton University, University of British Columbia, LSE, Clemson University, the Santa Fe Institute, and ETH-Zurich for valuable comments and feedback. Eugenie Dugoua, Natalie Carlson, Goran Laraveski, Natasha Plotkin and Qing Zhang provided excellent research assistance. We thank the AEA and JSTOR for the provision of data and are grateful to the Sanford C. Bernstein & Co. Center for Leadership and Ethics for partial funding. All errors are our own.

1 Introduction

Modern governments incorporate expert opinion into policy analysis via a wide variety of formal and informal mechanisms. Examples from economics include central bank policy, antitrust policy, and the design of taxes and regulation. Beyond economics, expertise in climate science, medicine and public health, and many engineering disciplines are of immediate relevance to policy makers. Expert opinion and judgement is often expected to be non-partisan, and yet experts may have partisan or political preferences of their own.

Whether expert opinion includes political beliefs is difficult to empirically assess. While diagnosing partisanship in media or speech is relatively straightforward, specialized technical languages may make it difficult for outsiders to distinguish partisan beliefs from expert judgement. In addition, political opinions may be shaped by expertise rather than vice-versa, or there might be unobserved experiences that shape both expert views as well as political preferences.

In this paper, we provide a purely inductive method for assessing the importance of political preferences to professional sorting in economics and to the substance of economic research using a purely inductive approach. We predict out-of-sample individual political behavior with the language from that individual's academic research papers, even adjusting for field of research. If political preferences were irrelevant for academic research in economics, this should be very difficult. Nonetheless our method generates good out-of-sample predictions of economist political behavior based on academic writing alone. We use this methodology to predict the ideology of economics papers and individual economists, and as our main application, we document a robust correlation between predicted ideology of authors and empirical estimates in policy relevant literatures.

Why focus on economics to study political preferences in research? Economics has more partisan diversity than any other social science.¹ Economics has more direct policy influence than other social sciences, and economists are the most highly paid and confident in their methodology.² In the United States, the Council of Economic Advisors has no analogue in the other social sciences, and the representation of economists in institutions such as the Congressional Budget Office, the Federal Reserve,

¹Cardiff and Klein (2005) use voter registration data in California to rank disciplines by Democrat to Republican ratios. They find that economics is the most conservative social science, with a Democrat to Republican ratio of 2.8 to 1. This can be contrasted with sociology (44 to 1), political science (6.5 to 1) and anthropology (10.5 to 1).

²Fourcade, Ollion, and Algan (2014) show that economists are the highest paid of the social scientists, and are the least likely to use interdisciplinary citations.

the Federal Trade Commission, the Department of Justice, and other agencies is far larger again than that of any other social science. Empirical work in economics informs policy proposals and evaluations, and economists often testify before Congress. More broadly, economic ideas are important for shaping economic policy by influencing the public debate and setting the range of expert opinion on various economic policy options (Rodrik 2014).

In his ‘The Politics of Political Economists’, George Stigler argued that while professional economics was conservative (in the sense of hostile to radical changes) in its orientation, advances in economic science were non-partisan due to its institutionalized incentives and norms for the dissemination of information. “The dominant influence upon the working range of economic theorists is the set of internal values and pressures of the discipline” (Stigler 1960 pg 40). Stigler believed that political and policy preferences do not drive economic research, and when they do, it is for the worse.³ This belief that economics conforms with scientific norms, articulated for example by the sociologist Robert Merton (1942), is the basis of the working consensus that is widely defended.⁴

Yet, the evidence for the view that scientific practices purge ideology from economics is surprisingly thin, relying upon surveys or subjective coding of political beliefs. We investigate the role of political preferences, or ideology, in economics with a data-driven approach. We extend methods of machine learning and of natural language processing introduced to economics by Gentzkow and Shapiro (2010). Data on individual campaign contributions and on petition signings establish a ground-truth sample of economists’ ideologies, which, linked to the text of academic articles, allows us to identify word phrases whose frequency is correlated with individual partisan political behavior. These “partisan” phrases look intuitively plausible, and are identified within a given topic of research, ensuring that we are not simply picking up different language patterns across fields of economics. We use the correlations of these phrases with partisan political behavior to predict out-of-sample economist political behavior. We validate these predictions of political preferences using held-out data, as well as confirming that they are correlated with partisan IGM responses (Gordon and Dahl 2013). Our first result is that it is indeed possible to predict partisan behavior with high-dimensional representations of academic

³Stigler continues “Often, of course, the explicit policy desires of economists have had a deleterious effect upon the theory itself... the effect of policy views on the general theory has stemmed from a feeling that the theory must adapt to widely held humanitarian impulses.” (Stigler 1960 pg 43)

⁴For example, see <http://www.nytimes.com/2013/10/21/opinion/yes-economics-is-a-science.html> Chetty 2013

writing, suggesting that distinct academic writing is associated with distinct political preferences.

The possibility of a completely apolitical economic science remains contested. In their classic book “the Making of an Economist”, Colander and Klamer (1990) argued that graduate training in economics induced conservative political beliefs, with Colander (2005 pg 177) writing that: “10 percent of first-year students considered themselves conservative; by the fourth and fifth year, this number had risen to 23 percent. There was also a large drop by year in students who considered themselves radical; that percentage fell from 13 percent of first year students to only 1 percent of fourth-year and higher students”. Or consider the view of The Economist magazine: “People drawn to the economics profession do tend to favour the market mechanism to allocate resources instead of the government”.⁵ Countering this view of economics as intrinsically libertarian, Klein and Stern (2005) use a survey of AEA members to argue that only 8% of economists are “truly free-market”. The best evidence comes from a comprehensive survey undertaken by Fuchs et al. (1998) who asked a number of labor and public finance economists their views on parameters, policies, and values. They conclude that “one of the most important empirical results of this study is the strong correlation between economists’ positions and their values, but an understanding of this relationship requires further research” (Fuchs et al., 1998, pp 1415). A series of recent papers investigate empirically the determinants of economic publication and citation patterns (Ellison 2000, 2004, Terviö 2013, Önder and Terviö 2013). Closest to our paper is the recent article by Gordon and Dahl (2013), who use the IGM survey responses to assess whether economists are divided over policy issues. None of these papers look at political ideology of economics articles, and none use the text of economics articles themselves as data, and instead analyze citation patterns or publication counts alone.⁶

Instead of these survey based methods, which may suffer from framing biases as well as selection⁷, our paper uses the correlations between patterns academic writing and observed political behavior to measure ideology. Ideology extraction from text has received attention from multiple fields including computer science, political science, and economics. Our tools most closely follow Gentzkow and Shapiro (2010) (see also Jensen et al 2013). Grimmer and Stewart (2013) provide an overview of many

⁵The Economist http://www.economist.com/blogs/freeexchange/2010/08/fiscal_policy

⁶A recent paper by Zingales (2014) looks at papers in managerial compensation, and finds that top journals are more likely to publish papers that suggest that managerial pay increases are optimal and that IGM-surveyed economists who serve on boards are more likely to disagree with the statement that CEOs are paid more than their marginal productivity.

⁷Fuchs et al. (1998) only survey economists at top 40 schools, and have only a 50% response rate. The IGM survey only looks at a small sample of “top” economists, and tends to be more left than average by our measure, as we show below.

models used in the analysis of the text of political actors like politicians and bureaucrats.⁸ Recent research in natural language processing has focused on unsupervised topic models that jointly model lexical variation due to topics and other factors such as ideology (Mei et al. (2007); Lin et al. (2008); Ahmed and Xing (2010); Paul and Girju (2010); Eisenstein et al. (2011); Wang et al. (2012)). While these models can show high predictive accuracy, they have not been applied in technical domains where ideology is not immediately apparent.

Importantly, detecting ideology in domains where institutions and norms are in place to maintain neutrality is different from predicting ideology in domains where it is overt, such as media or political speech, as all of the papers using text drawn from political actors do (Jelveh et al., 2014a). Adjusting for topics may be particularly important in highly specialized domains, where language use is tailored to very narrow audiences of other experts. Other domains with similar politics embedded in technical language could include climate science, law, or human genetics and evolution.

As an application of the usefulness of our methodology, we turn to empirical results in several key policy relevant fields in economics. The policy relevance of economics partially derives from its ability to combine economic theory (e.g. supply and demand) with parameter estimates (e.g. elasticities) to make prescriptions about optimal policies (e.g. taxes). We draw policy relevant elasticities from Fuchs et al. (1998), and locate available survey papers that compile estimates of these parameters.

From a variety of published survey papers, we collect estimates of taxable income elasticities, labor supply elasticities, minimum wage employment elasticities, intergenerational mobility elasticities, and fiscal multipliers. Using ideology predicted from papers by authors written before the reported elasticity, we find a significant correlation of ideology with reported estimates of various policy relevant parameters, with predicted liberals reporting elasticities that imply policies consistent with more interventionist ideology.

2 Conceptual Framework

In the Appendix A.1, we provide a formal framework to interpret our estimation and prediction strategy in terms of the political preferences and professional incentives, including sorting into subfields, facing

⁸Unsupervised modeling is a machine learning technique which tries to uncover patterns in data without using auxiliary data (e.g. cluster analysis or hidden Markov models).

economists. In our model, economists are indexed by a latent left-right political ideology variable. The extent of preferences for being “neutral” in academic writing are parameterized, as are the importance of professional incentives. These professional incentives can include conforming with the ideology expressed in a subfield of economics. Economists choose the degree of their political preferences to reveal in their academic writing according to their tastes, trading this off against their desire for professional success.

The model illustrates the assumptions needed to recover ideology from our empirical strategy. Importantly, our empirical strategy requires that there be no omitted variables that are correlated with both academic text as well as political behavior (like campaign contributions) besides ideology. An important potential omitted variable is field of economics, which we incorporate as an extension.

When economists are allowed to sort into fields, we have many equilibria. But an important set of equilibria involve agents sorting into distinct fields based on similar ideologies. We model fields as composed of peers, and success in a field is more likely when papers are aligned with the average ideology within the field. Indeed, in the simple 2-subfield model in the Appendix, where professional incentives push agents to sort into fields where they can express their ideology in academic articles and reviewers and peers will accept them, equilibria can arise where all agents left of the median sort into one field, and all agents right of the median sort into the other field. Besides illustrating the identification assumptions, the conceptual framework stresses the importance of adequately controlling for field, and motivates our use of both JEL codes and topic models to categorize papers.

3 Data

3.1 Linking Economists to Their Political Activity

To define our set of economists, we obtained the member directory of the American Economics Association (AEA) for the years 1993, 1997, and 2002 to 2009. From these lists, we extracted over 53,000 potential authors where for each member we have his or her name, location, email address, education, employer, and occupation.⁹ These data are used to match members across years. We then link the AEA member directory to two datasets with observed political behavior: political campaign contributions

⁹Since AEA members are drawn not only from academia, but government and the business world, not all of these individuals have produced academic research.

and petition signing activity.

We obtain campaign contribution data from the Federal Election Commission’s website for the years 1979 to 2012. Campaign committees are required to publicly disclose information about individuals who have contributed more than \$200 to them. These disclosures contain the contributor’s name, employer, occupation, state, city, zip code, transaction date, and transaction amount. Our goal is to match the AEA roster to these individual contributions of which there are about 20 million. This is an example of a typical record linkage or data matching problem and has been studied extensively in the science of informational retrieval.¹⁰ Ideally, we would like to compare each AEA member with each FEC contributor to determine if there is an identity match while taking into account that, in a significant proportion of matches, a person’s information will be recorded differently in the two databases. To address this, we apply a fuzzy string matching algorithm (Navarro, 2001) to member and contributor attributes. We describe the methodology and the results in full detail in Appendix A.2, and summary statistics on the campaign contributions are provided in Table A.1. Besides campaign contributions, we also proxy economist ideology through petition signings. Our data comes from Hedengren et al. (2010) who collected 35 petitions signed principally by economists. We use fuzzy string matching and manual inspection to match the signatories to our economists. Hedengren et al. (2010) classify petitions on whether they advocate for or against individual freedoms. Similarly for our purposes, many of the petitions exhibit viewpoints that are aligned with the political left or right. Examples include petitions for and against federal stimulus following the 2008 financial crisis and petitions endorsing or opposing John Kerry’s 2004 presidential campaign. Appendix Table A.2 reproduces the list of petitions from Hedengren et al. (2010) which includes their classification on the liberty scale along with an additional column indicating our classification. We drop petitions classified as neutral. Figure 1 compares the ratios of contributions to Democrats vs. Republicans against the ratio of signatures for left- and right- leaning petitions. Surprisingly, left-leaning authors make more political contributions while right-leaning authors sign more petitions.

We now take a simple approach to assigning an ideology to an economist based on their campaign contribution and petition signing behavior. Let $pet_{k,e}$ be the number of petitions signed by economist e aligned with partisanship k taking on values d (left-leaning), r (right-leaning), or u (undetermined). A

¹⁰A general probabilistic approach was formalized by Fellegi and Sunter (1969). For more recent developments, see Winkler (2006).

similar definition applies to $contrib_{k,e}$ which is the number of campaign contributions. The following logic is then applied to assigning ideologies, θ_e .

- For each economist e and ideology labels $x, y \in \{d, r\}, x \neq y$:
 - If $pet_{x,e} > pet_{y,e}$ and $contrib_{x,e} > contrib_{y,e}$ then $\theta_e = x$
 - If $pet_{x,e} > pet_{y,e}$ and $contrib_{x,e} = contrib_{y,e} = 0$ then $\theta_e = x$
 - If $pet_{x,e} = pet_{y,e} = 0$ and $contrib_{x,e} > contrib_{y,e}$ then $\theta_e = x$
 - Otherwise $\theta_e = u$

If an economist has given more times to Democrats (Republicans) and signed more left-leaning (right-leaning) petitions, then the assigned ideology is left-leaning (right-leaning). In the cases where the economist has zero contributions (or signed no petitions) then we only consider signed petitions (contributions). If there is disagreement between the signals, or one of them is indeterminate but nonzero (e.g same number of Republican and Democrat contributions), then we treat the ideology as undetermined.

Revealed ideology through campaign contributions and petition signatures is largely consistent. Table 1 displays the pattern exhibited by 441 AEA members who both signed partisan petitions and contributed to Democrats and/or Republicans. Of these, 83.4% showed agreement between their petition signatures and campaign contributions. However, these rates mask some heterogeneity. When viewed from the perspective of contributions, 76.7% of AEA members who contributed more to Democrats also signed more left-leaning petitions while 98.7% of members who contributed more to Republicans signed more right-leaning petitions. When viewed from the petition signing perspective, 98.7% of members who signed more left-leaning petitions also contributed more to Democrats while only 69.5% of members who signed more right-leaning petitions gave more times to Republicans. Economists who contribute more times to Republicans or sign more left-leaning petitions have greater consistency in their ideologies.

3.2 Economic Papers Corpus

To create our corpus of academic writings by economists, we also obtained from JSTOR the full text of 62,888 research articles published in 93 journals in economics for the years 1991 to 2008. We also

collected 17,503 working papers from the website of the National Bureau of Economic Research covering June 1973 to October 2011. These papers were downloaded in PDF format and optical character recognition software was applied to extract text. Appendix figures A.5 and A.4 show the number of JSTOR and NBER papers per year, respectively, and Table A.11 lists the individual JSTOR journals.

We remove common words and capitalization from the raw text and use a stemmer (Porter, 1980) to replace words with their morphological roots.¹¹ For example, a stemmer will resolve the words ‘measures’, ‘measuring’, and ‘measured’ to their common root ‘measur’. We construct predictors for our algorithm by combining adjacent words to create phrases of length two (bigram) and three (trigram). We drop phrases that occur less than five times. To further focus our attention on the phrase sequences that are most likely to contain ideological valence, we follow Gentzkow and Shapiro and compute Pearson’s χ^2 statistic for each remaining phrase. More explicitly, we create a ranking of phrases by partisanship by computing

$$\chi_{pl}^2 = \frac{(c_{plr}c_{\sim pld} - c_{pld}c_{\sim plr})^2}{(c_{plr} + c_{pld})(c_{plr} + c_{\sim plr})(c_{pld} + c_{\sim pld})(c_{\sim plr} + c_{\sim pld})} \quad (1)$$

where c_{pl} is the count for the number of times phrase p of length l was used by all economists of a particular ideology (d or r) and $c_{\sim pl}$ is the number of times phrases of length l that are not p were used. We calculate p-values from the χ^2 statistics and keep only those phrases where this value is ≤ 0.05 .

3.2.1 Accounting for Topics

Table 2 lists the 40 most slanted bigrams and trigrams ranked by χ^2 values. A quick glance at this table leaves the impression that the top ideological phrases are reflective of sorting into research subfields by ideology. For example, right-leaning terms like ‘bank_note’, ‘money_suppli’, and ‘feder_reserv’ are typically associated with macroeconomics and left-leaning terms ‘mental_health’, ‘medic_care’, and ‘mental_ill’ are related to health care. While sorting is an interesting phenomenon to document in and of itself, we are also interested in whether individual ideology is associated with research results. We attempt to account for field-level sorting by estimating ideology within research area. We map papers to topics and predict authors’ ideologies using topic-specific phrase counts. These predictions are combined to form a final estimate of an author’s political leaning. We are consequently removing

¹¹These common words include terms not likely to be correlated with ideology such as ‘a’, ‘the’, and ‘to’.

the effect of field ideology by estimating individual ideology within field. We also calculate individual ideology scores without adjusting for topics; these are called the 'no topic' ideology scores. Since we do not observe topics for all of the papers in our corpus, we use two methodologies from statistical natural language processing to create topic classifications for papers.

3.2.2 JEL Codes as Topics

Our first method for estimating topics takes advantage of JEL classification codes maintained by the *Journal of Economic Literature*. These codes are hierarchical markers of an article's subject area. For example, the code C51 can be read, in increasing order of specificity, as Mathematical and Quantitative Methods (C), Econometric Modeling (C5), Model Construction and Estimation (C51). Our JSTOR dataset did not include JEL codes so we obtain classifications for 539,572 published articles and the 1.4 million JEL codes assigned to them by the *Journal of Economic Literature*.

We were able to match and assign JEL codes to 37,364 of our JSTOR articles. The average paper was assigned to 1.90, 2.31, and 2.68 first-, second- and third-level JEL codes, respectively. We then use the relationship between language and topic codes in a machine learning algorithm to predict JELs for the set of papers that fall outside of the EconLit data. We predict codes for the 1st and 2nd levels and refer to these topic mappings as *JEL1* and *JEL2*.

To predict JELs for papers that don't have them, we take a "one-vs-all" (Bishop, 2006) approach to construct a series of binary classification models. For each code, we take the set of papers for which we know the actual JEL codes and construct a training set where $y_{p,j}$ equals one if paper p was assigned code j and zero otherwise and \mathbf{X} is a matrix where the (p, w) -th element is the count of the number times word w appeared in paper p . We construct a series of models¹² that estimate $\hat{y}_{p,j}$, the probability that paper p is about topic t . The models perform well with an average AUC of X. We describe the full results of these models in A.3.

3.2.3 Latent Dirichlet Allocation

In the previous section we described a method which relied on an annotated set of topics in order to predict JEL codes. We also construct topic mappings using a class of algorithms which don't rely

¹²We apply gradient boosting (Friedman, 2002)

on pre-labeled data but form topics based on patterns of co-occurring words across documents. The most common method for this task, known as Latent Dirichlet Allocation (LDA) (Blei et al., 2003), defines a probabilistic model for the joint distribution of observed data and the latent factors generating the data. In common practice, the observed data are text from documents and the latent factors are unobserved topics. A key assumption behind LDA is that documents can be about more than one topic. For example, some of the topics present in this paper are economics, political ideology, and text mining. Topics are shared by all documents in a corpus; the topic proportions are document-specific and randomly drawn from a Dirichlet distribution. LDA allows each document to exhibit multiple topics with different proportions, and it can thus capture the heterogeneity in grouped data which exhibit multiple latent patterns. One of the consequences of the Dirichlet distribution is that topic proportions are independent. To overcome this, we use a related algorithm, the correlated topic model (CTM) (Lafferty and Blei, 2006).

We ran CTM on the set of papers used in the analyses below which accounts for 57,742 papers. Mappings were created with 30, 50, and 100 topics (*CTM30*, *CTM50*, and *CTM100*). For each topic, it is possible to rank the words or phrases most relevant to that topic. These rankings can be used to qualitatively assess a real-world analogue to the algorithm-generated topics. For example, the left-most column of Tables 3 and 4 shows the top twenty bigrams for two of the topics generated by running CTM with 50 topics on the economic papers corpus and our qualitative descriptions for those topics.¹³

We use the topic distributions estimated by CTM to assign articles to topics. If there is at least a 10% probability that an article is about a topic, then we assign that article to that topic. While 10% might seem to be a lower threshold, the topic distributions estimated by CTM tend to be sparse. For example, even with 50 topics to ‘choose’ from in *CTM50* and a threshold of 10%, 99.5% of the papers would be assigned to five or fewer topics.

4 Predicting Ideology From Phrases

In this section we describe how the gathered and constructed data outlined above are used in our prediction algorithm. To recap, we have created a dataset which contains the following:

¹³The top phrases and partisan phrases for all the topics, both *JEL1* and *CTM50*, are available online at www.santafe.edu/~snaidu/topics.pdf.

- 1) A set of economists with known ground-truth ideology derived from campaign contributions and petition signatures
- 2) A set of economists with unknown ideology
- 3) The set of papers written by these economists
- 4) The n -grams and associated counts for each paper
- 5) Six mappings from papers to topics: *JEL1*, *JEL2*, *CTM30*, *CTM50*, *CTM100*, and *NoTopic*.

The *NoTopic* mapping refers to pooling all papers without regard to topic.

Our topic-adjusted algorithm for ideology prediction works as follows: We first create two sets of individuals, e_{gt} and e_{ngt} , the set of groundtruth and non-groundtruth authors, respectively. We split e_{gt} into five equally sized subsets (folds). We iteratively hold out one fold and build models with the other four folds (training set). This procedure produces out of sample predictions for each person in e_{gt} which will allow us to estimate model performance. The procedure also produces five separate predictions for each person in e_{ngt} which we combine by taking the mean.

For a given a topic mapping, we iterate through each topic t , and, for each topic, select the papers written by ground-truth authors. We now alter the χ^2 computation from equation 1 and perform it at the topic level. For a given topic t , we compute χ_{plt}^2 by only considering the set of phrases that appear in papers in t ¹⁴. Certain phrases might pass our ideological filter either by chance or because of author idiosyncrasies in our ground-truth dataset. Both of these scenarios have the effect of increasing the number phrases that are then used for the prediction model and potentially increasing the noise to signal ratio. To capture phrases that are consistently correlated with our measure of ideology, we partition the training data into five folds. We hold out one fold at a time and apply the χ^2 filter to identify significantly slanted phrases. We take the intersection across the five sets of significant phrases.

We construct the count frequency matrix \mathbf{F}_t where the (e, p) -th entry is the number of times economist e used partisan phrase p . For papers with multiple authors, each author gets the same count of phrases. As our analysis is at the author-level, economists' phrases are aggregated across their papers in a topic. To predict ideology, we then use decision trees, a non-parametric machine learning

¹⁴In order for a paper to be classified as TTTT

algorithm which work by recursively partitioning the input space into subspaces that seek to maximize the homogeneity of outcome variable in each subspace. A short-coming of decision trees is that they can over-fit data, i.e. find signal where there is actually noise. To overcome this we apply Random Forest (Breiman, 2001), a model averaging algorithm which combines the output of many decision via the application two model averaging techniques: bootstrap aggregation (also referred to as bagging) and attribute bagging. With bagging, samples of the original data are drawn with replacement to form a new dataset. In our case, we sample with replacement from the rows \mathbf{F}_t . With attribute bagging, the sampling is done at the level of the predictor (i.e. attribute). In our case, the resulting dataset would be created by sampling without replacement from the columns of \mathbf{F}_t .¹⁵ Each decision tree within a Random Forest model can be viewed as a vote on whether an author is left- or right-leaning. We calculate the vote as follows.

Our algorithm results in a three-dimensional array with the (e, t, c) -th entry representing the number of votes economist e received in topic t for ideology c . A final prediction is computed as the weighted average of the percentage of right-leaning votes received in each topic:

$$\hat{\theta}_e = \sum_{t=1}^T w_{e,t} \frac{r_{e,t}}{r_{e,t} + d_{e,t}}$$

where $r_{e,t}$ and $d_{e,t}$ are the number of right- and left-leaning votes economist e received in topic t , respectively, and $w_{e,t}$ is the topic- and economist-specific weight. We let $w_{e,t}$ equal the share of all votes e received in t

$$w_{e,t} = \frac{r_{e,t} + d_{e,t}}{\sum_{t=1}^T r_{e,t} + d_{e,t}}$$

which simplifies predicted ideology to

$$\hat{\theta}_e = \frac{\sum_{t=1}^T r_{e,t}}{\sum_{t=1}^T r_{e,t} + d_{e,t}}.$$

Topics with more votes have a greater say in determining the final prediction. Ideology values closer to zero are associated with a left-leaning ideology and values closer to one are associated with a rightward lean. To get back to the $[-1, 1]$ range, we transform $\hat{\theta}_e$ by multiplying by two and subtracting

¹⁵We sample twice the square root of the number of columns in \mathbf{F}_t

by one. For example, if $\hat{\theta}_e = .5$, we multiple this number by 2 and subtract 1, returning the value of 0. Thus, our ideology scores are centered in theory at 0 with a maximum value of 1 and minimum value of -1. The empirical mean will deviate from 0 depending on the sampling.

4.1 Validation

Our analysis in subsequent sections involves studying attributes of academic economists at American colleges and universities. We know the ground-truth ideology for a subset of these economists and use this information to evaluate the predictive ability of the algorithm presented in the previous section. In terms of our model, we are comparing observed θ_i to predicted $\hat{\theta}_e$. We stress here that the information from this group of economists is never used as a predictor in our algorithm (i.e. it is held out) so we are not contaminating our estimate of $\hat{\theta}_e$ with θ_i itself. This means that the to-be-analyzed authors phrase counts are not used in the χ^2 filter step or as input into the prediction algorithm. Additionally, we also eliminate the contamination from the case where we predict the ideology of an economist who has coauthored a paper with someone in the ground-truth dataset. When we construct the vector of phrase counts for this author, we do not include the phrase counts from the coauthored paper.¹⁶

We assess the performance of our algorithm by employing a summary statistic that is commonly used in binary prediction problems: the area under the receiver operating curve (AUC) (Fawcett, 2006). To see how this curve is constructed, note that our algorithm produces a probability that an author is right- or left-leaning. We translate these probabilities to binary predictions by setting a threshold (e.g. 25%, 50%, etc.) and assigning an author to be right-leaning if their predicted ideology is above this threshold and left-leaning otherwise. From each possible threshold, we compute and plot the true positive rate (the proportion of correctly predicted right-leaning authors) and the true negative rate (the proportion of correctly predicted left-leaning authors). By connecting these points, a Lorenz-like curve is created. The area under this curve can range from zero to one and tells us about the predictive accuracy of our algorithm. An AUC of one means the classifier can perfectly separate positive from negative cases, an AUC of 0.5 means the classifier does no better than random guessing, and AUCs below 0.5 imply the model actually does worse than random guessing. The AUC is equivalent to the probability that a binary classifier will rank a randomly chosen right-leaning author higher than a

¹⁶We also ran a version of the algorithm where these types of coauthored papers were dropped from the dataset but our results were unaffected.

randomly chosen left-leaning author, where the rank is based on the percentage of right-leaning votes received.

There are two primary benefits to employing AUC as a performance metric. First, the AUC is less sensitive to asymmetry in the outcome distribution than a simple measure of accuracy. To see this, imagine the extreme case where we had 90 left-leaning and 10 right-leaning economists in the test set. If all authors were predicted to be left-leaning, our accuracy would be a seemingly strong 90% even though the algorithm itself was quite dumb. The second benefit is that algorithm performance is not a function of just one threshold but many. For example, a naive way of converting the predicted probabilities to ideology assignments would be to assign authors as right-leaning if their predicted probability is greater 50% and left-leaning otherwise. But it may be the case that the best separation between left- and right-leaning authors occurs at some other threshold.

Figure 2 shows the AUC plots and Table 5 the relative performance for our various topic mappings. While *CTM50* provides the best performance, many of the models show similar results in terms of AUC and correlation with ground truth ideology. The maximum correlation between predicted and ground truth ideology is 0.412. For comparison, the out-of-sample correlation reported by Gentzkow and Shapiro between their ideology measure and one obtained from another source of newspaper slant was 0.40.

We can also see from Table 5 that a model without an ensemble component performs worse than all other models except for *JEL2*. The likely reason for the under-performance of *JEL2* is that the combination of a large number of topics and a low number of topics assigned to each paper lead to a small dataset size by which to estimate PLS in each *JEL2* topic. There are about two topics assigned to each paper in the *JEL2* mapping. For comparison, the *CTM* topic mappings have about four topics per paper. *JEL1* also has about two papers per topic, but since the number of *JEL1* topics is about 15% of the size of *JEL2* topics, each *JEL1* topic still has many papers.

Adjusting for topics does not appear to provide a performance improvement versus the unadjusted model. But this result may be due to the 4-fold increase in the number of authors used to construct the prediction algorithm for the *No Topics* mapping. For comparison, we also constructed a reduced version of the *No Topics* mapping by down-sampling the number of authors to mimic that of the topic-adjusted mappings. To do so, we constructed the *No Topics* version by sampling 400 of the 1,812

available authors and computing the prediction metrics. We repeated this 30 times by taking a different sub-sample of 400 authors and averaging the results. As shown in Table 5, the predictive performance declines in this scenario, suggesting that the larger sample size on which the full *No Topics* version is built is driving some of its accuracy and that adjusting for topics does provide some performance gains.

For further insight into how well our model generalizes, we use data from Gordon and Dahl (2013) to compare our predicted and ground-truth ideologies to responses provided by economists for a survey conducted by the Chicago Booth School of Business through October 30, 2012. The panel sets out to capture a diverse set of views from economists at top departments in the United States. Each question asks for an economist’s opinion on a particular statement. The questions reflect issues of contemporary and/or long-standing importance such as taxation, minimum wages, or the debt ceiling. Valid responses are: Did not answer, No Opinion, Strongly Disagree, Disagree, Uncertain, Agree, Strongly Agree.¹⁷ Of importance here is that Gordon and Dahl (2013) categorize a set of questions where agreement with the statement implies belief in ‘Chicago price theory’ and disagreement implies concern with market failure. The former of these also implies a rightward lean while the latter is consistent with left-leaning beliefs.

While Gordon and Dahl (2013) found no evidence of a conservative/liberal divide in the survey responses, we find a significant correlation between the responses and our predicted ideologies. We also know the ground-truth ideology of 20 members on the panel and the correlation between ground-truth ideologies and survey responses is also significant. Figure 3 shows binned scatterplots from a linear probability specification, conditional on question fixed effects, for each of our 4 ideology measures. There is a clear correlation between the conservativeness of the predicted ideology scores and the IGM measure of conservativeness.

In order to examine this more formally, Table 6 Panels A-D further presents results from logit and ordered logit regressions of the following form:

$$Pr(response_{i,j} = C) = \Lambda(\tau_j - X_{ij}) - \Lambda(\tau_{j-1} - X_{ij}), \quad (2)$$

where Λ is the logistic cumulative distribution function, τ represents cut points dividing the density

¹⁷For further details on the data see Gordon and Dahl (2013) and Sapienza and Zingales (2013). The latter show that the IGM panel answers to the questions are far away from the answers of a random sample of the public.

into regions corresponding to survey responses, and $X_{ij} = \beta_1 \hat{\theta}_i + \beta_2 \text{question}_j$. Hats denote predicted values. In the logistic version (columns 1-3), $\text{response}_{i,j}$ is a binary variable indicating whether the panelist agreed with the conservative viewpoint or not.¹⁸ In the ordered logistic version (columns 4-6) the response variable is coded with the following order: Strongly Disagree, Disagree, Uncertain, Agree, Strongly Agree.¹⁹ As seen in Table 6, the coefficients between the ideology variable and the conservative viewpoint are all in the expected directions and all are significant. The magnitude of the relationship varies between the models. For the ground-truth model, the probability of switching from liberal to conservative increases by about 5% when a person’s ideology switches from far left to far right. Other models put the probability at between 14% to 48%. Across all the different topic adjustments, the logit and ordered logit results in Table 6 show a significant positive relationship between our ideology variables and the probability of being in an increasingly conservative category. Columns 3 and 6 in each panel add the same controls as Gordon and Dahl (2013), which are the years of the awarding of a Ph.D. and the indicator variables for Ph.D. institution, NBER membership, gender, and experience in federal government. It is worthwhile to note the small increase in log-likelihood when controls are added, suggesting that our ideology scores are much better predictors of IGM responses than demographic and professional controls.²⁰

4.2 Sorting by Professional Characteristics

We link CVs of economists to our ideology prediction and document cross-sectional patterns of ideology. We start by first describing these descriptive patterns of ideology across fields of economics as well as school and career characteristics. We collect data from CVs of economists at top 25 departments and top 10 business schools in Spring 2011. The list of schools, the number of economists, and mean ideology is provided in Table A.6. We collect year and department of Ph.D. and all subsequent employers, nationality and birthplace where available, and use self-reported field of specialization. As Proposition 1 suggests above, we are interested in the political behavior of economists by subfield. In particular, looking at self-declared primary fields, we examine labor economics, public economics,

¹⁸Uncertain, No Opinion, and Did not answer responses were dropped for the binary logistic analysis.

¹⁹No Opinion and Did not answer responses were dropped for ordered logit analysis.

²⁰As an additional validation exercise, we run our algorithm on a corpus of editorials written by Israeli and Palestinian authors and show that we can achieve high prediction accuracy. We discuss our performance relative to other political scaling methods more completely in our companion paper Jelveh et al. (2014b).

financial economics (including corporate finance), international economics, and macroeconomics as determinants of political behavior, as these are among the most policy relevant fields in economics, but we also examine a number of other fields. We classify each department as saltwater or freshwater or neither following Önder and Terviö (2012). An economist is saltwater or freshwater if either went to grad school, had their first job, or had their current job at a saltwater or freshwater school.

We are interested to see if there are significant correlations between political ideology and field of research. Note that even though our ideology scores are adjusted for topic, self-reported fields of individuals vary independently of topic-adjusted paper ideologies. So it very well could be that financial economists who write on monetary policy adopt conservative language within that topic. Secondly, we are interested in institutional affiliations. We construct a variable for being at a business school, a Top 5 department, as well as our indicator for “freshwater” and “saltwater” schools. Finally, we consider a set of demographic and professional characteristics such as Latin American origin, European origin, and doctoral degree year, years between undergraduate degree and economics phd, and number of different employers per year since obtaining the Ph.D. We present summary statistics in the appendix Table ??.

We then look at the correlation between author ideology and various CV characteristics. The estimating equation is:

$$\hat{\theta}_i = \sum \delta_{Field} Field_i + \gamma X_i + \delta_{DoctoralInstitution \times Year} + \epsilon_i \quad (3)$$

Here $\hat{\theta}_i$ denotes predicted ideology, $Field_i$ is a set of indicator variables for different fields of economics, and X_i is a vector of other economist characteristics. We also control for fixed effects by Ph.D. Institution X year, to see if the correlations remain robust within Ph.D. cohorts. Standard errors are clustered at the department level. We vary this specification with different sets of controls, including department fixed effects, university fixed effects (there are 15 business schools in the same university as economics departments in our sample).

Figure 4 summarizes the results from the baseline specification when the CTM50 measure of predicted ideology is used. We see that the fields of finance, macroeconomics and international trade are more conservative, while labor is considerably more liberal than the average. Other fields, such as development and econometrics, show less political valence. We further see that faculty at business schools are more conservative, as are professors affiliated with “freshwater” schools, while “saltwater”

schools have a left-wing bent. Professors of European origin also seem to be somewhat more conservative, and there seems to be no effect of Latin American origin, full professor rank or Top 5 department ranking.

The bottom set of coefficients in the plot shows results from specifications where the groundtruth variable included as a covariate. Virtually all coefficients become smaller and insignificant (save 2). This is reassuring: if our results are truly driven by the true ideology and not prediction error, the correlation between characteristics and the predicted ideology should disappear when true ideology is included as a control.

We show analogous results for the CTM-100 ideology measure in Table 7 with the *Notopic* and *JELI* adjusted ideology scores shown in Appendix tables A.8 and A.7. Column 1 shows the basic regression with no fixed effects.

Note that this is estimated from topic-adjusted ideologies, so it is not simply selection into area of research. While this could indicate that our topic adjustment strategy is performing poorly, it could also imply that self reported fields are a significant predictor of ideology even within a field. It could very well be that a financial economist who writes on monetary policy adopts conservative language within the field of monetary economics. While not reported to save space, there is no robust evidence of significant ideology for economists who declare their primary fields as microeconomic theory, econometrics, development, or economic history.

It is natural to hypothesize that faculty in business schools lean conservative, as sympathy with business interests is either induced or selected on by institutions that educate business leaders. Our methodology finds more conservative ideology for economists at business schools. This is true controlling for both self-reported field as well as controlling for university fixed effects, and so suggests that there is some professional affinity between business schools and conservative ideology.

The finding that both the finance subfield and business schools tend to attract (or produce) economists with more conservative predicted ideology is interesting in light of the patterns documented in Fourcade et al. (2014), who show that there has been a pronounced increase in economists with business school affiliations as well as in the importance of financial economics as a subfield within economics over the past few decades. These two trends, together with the political preferences documented here, may have contributed to the perception that economics is a “conservative” field.

We also test the saltwater-freshwater divide. One natural hypothesis is that saltwater economists are more left wing than freshwater economists. While this appears to be the case, it is only because there is no significant correlation between freshwater economists and ideology, so the saltwater-freshwater methodological divide, insofar as it is political, appears to be one sided.

The magnitudes of all these coefficients should be interpreted as effects on the expected ideology of the economist. For example, a coefficient of 0.2 indicates that the author was 10 percentage points (20 divided by the 2 that we rescale all the ideology scores by) more likely to be classified as a Republican by our prediction algorithm.

We also find that predicted ideology is persistent within individuals. As documented fully in Table A.1, we split authors by their first 50% of publications and their second 50%. We then predict ideology separately for each set of publications, and find that the correlation between early predicted ideology and late predicted ideology is quite high.²¹

5 Ideology And Policy Elasticities

Part of economists' influence on policy is arguably its quantitative precision. Economic theory identifies important empirical estimates that in turn imply particular optimal policies. Introductory microeconomics teaches thousands of students every semester about supply and demand elasticities, and how knowing the magnitude of the relevant elasticity tells you about the economic incidence of various policies. Economic literatures have thus developed around key empirical estimates of behavioral responses to policy. These elasticities are then used to argue, either formally or informally, for various policies. For example, the labor demand elasticity for low-wage workers can tell policy makers what the costs and benefits of the minimum wage are, and empirical fiscal multipliers gauge the efficacy of government stimulus spending. Various government agencies, such as the Congressional Budget Office, the Federal Reserve, and the Federal Trade Commission actively incorporate empirical economic research into policy evaluations.

²¹In a previous version of the paper, we examined the relationship between journal editors and journal ideology, measured as the mean ideology of the articles. While predicted editor ideology is strongly correlated with journal ideology in both the cross-section and the pooled panel data, this relationship disappears once journal fixed effects are included as a control, with precise standard errors. While this may suggest that editors have no effect on journal ideology, it may also suggest that the ideological matching between editors and journals is quite assortative, and so there is little variation in ideology across editorial changes within a journal.

This marriage of economic theory and data is well-articulated, again, by Stigler: “In general there is no position, to repeat, which cannot be reached by a competent use of respectable economic theory. The reason this does not happen more often than it does is that there is a general consensus among economists that some relationships are stronger than others and some magnitudes are larger than others. This consensus rests in part, to be sure, on empirical research.” (Stigler 1959 pg 531).

Recently, the focus on key behavioral elasticities as sufficient for optimal policy has been reinvigorated in applied fields such as public finance, labor economics, industrial organization, and trade. (Chetty 2009, Weyl and Fabinger 2013, Costinot and Rodriguez-Clare 2014). This approach suggests that a variety of models incorporate similar fundamental economic intuition, which can then be encoded in a few empirical estimates. The magnitudes of these estimates, together with formulas yielded by economic theory, discipline the policy prescriptions of economists.

An important question, therefore, is if author political ideology predicts the magnitude of an elasticity reported in a published paper in these policy relevant literatures. If it does, it may suggest that economists are selecting into methodologies and variation that yield elasticities consistent with political beliefs. However, there is a possibility of reverse causation, whereby economists who discover elasticities that suggest that market interference is highly costly are moved to contribute to the Republican party or become conservative on other issues as well. It is very difficult to causally identify any effect of political ideology on empirical estimates, as any exogenous shock to political ideology could also influence the decision to be an economist, as well as the selection into what field of economics to work in. Therefore, we limit ourselves to a descriptive analysis. In a robustness exercise below, we mitigate endogeneity concerns by using only ideology estimated from the first 50% of an author’s writing.

We select elasticities drawing on Fuchs et al. (1998) (henceforth FKP). FKP survey labor and public finance economists about their views on policy and parameters. FKP estimate the correlation between policy preferences and beliefs about parameter values. They provide a mapping from policy preferences to economic parameters from labor and public that implicitly gives each parameter estimate a policy implication that is easy to map into a partisan direction. For example, beliefs about the empirical effect of unions on productivity might influence preferences towards increased unionization. Similarly, the female labor supply elasticity may influence beliefs about the desirability increasing Aid to Families with Dependent Children. The mapping between estimates and policies, as well as the implicit partisan

leaning, is provided in table 8. There is one elasticity, the labor demand elasticity, that FKP did not assign to a clear policy, and so we denote it non-“policy-relevant”. Indeed one can imagine a high labor demand elasticity being both favored by (conservative) skeptics of labor market interventions such as the minimum wage, as well as (liberal) skeptics about welfare reform.

We focus on estimated rather than calibrated or simulated parameters, which are mostly from the labor economics literature. We then looked through the literature for meta-analyses of these parameters, obtained the data from the authors where available, and then merged the authors of each estimate in each meta-analysis to our predicted slant measures. The list of meta-analyses is also in Table 8. In addition, we obtained a number of other meta-analyses from the meta-analysis archive maintained at Deakin University by Chris Doucougliasis²², enabling a placebo exercise, where we check the correlation between author ideology and non-policy relevant parameters. We expect the correlation between predicted ideology of the authors and policy-irrelevant parameters to be insignificant.

Meta-analyses necessarily rely on the judgements of the authors about what to include and what to exclude.²³ With such diverse literatures, we take the datasets as they are, and do not process them extensively. One exception is the female gender gap, where the literature reports both the total gender gap as well as the unexplained gender gap. We transform this to be the ratio of the unexplained to the total, to better account for idiosyncracies in choices of control variables.

There are often many estimates from a single paper. When standard errors are provided, we weight estimates by the inverse of the standard error, otherwise we take the simple average of estimates. These gives a single estimate from each paper. We show robustness to unweighted estimates below. We adjust the sign of each category of estimate so that higher is more conservative, following FKP, and present this in Table 8.

Meta-analyses may have distributions of estimates that are skewed or truncated (as shown in Andrews and Kasy (2017)) and so our primary measure is the rank of the coefficient within the category (multiplied by 100). We also look at a binary indicator for a coefficient being greater than the median. Finally, in order to give quantitative interpretations to our point estimates, we further normalize each paper-level estimate within the survey paper, taking the Z-score of its value using the mean and the standard deviation of the elasticities reported in the survey paper.

²²At <http://www.deakin.edu.au/buslaw/aef/meta-analysis/>, accessed March 6, 2016.

²³A recent paper by Andrews and Kasy (2017) examines the econometrics of meta-analyses rigorously.

As many estimates have multiple coauthors, we average the predicted author ideology to construct an estimated average author ideology for each paper. Let $coefrank_{js}$ denote the elasticity from paper j in survey paper s . Our baseline regression equation is given by:

$$coefrank_{sj} = \gamma \bar{\theta}_j + \delta_s + \epsilon_{sj} \quad (4)$$

where $\bar{\theta}_j = \frac{1}{|N_j|} \sum_{e \in N_j} \theta_e$ is the mean of the authors N_j of paper j ideology predictions from our methodology above. δ_s is a meta-analysis fixed effect, which will be included in all specifications, and ϵ_{sj} is an error term. We illustrate the basic variation using binned scatter plots in Figure 5, which shows that there is a strong correlation between our ideology measures and the coefficient rank, adjusting for meta-analysis fixed effects. This is true across our different topic adjustment, and in fact there is a positive correlation between groundtruth ideology and coefficient estimates. Table 9 shows estimates of γ from 4. Panel A shows results for no topic adjustment, Panel B for CTM100 adjustment, and Panel C for JEL1 adjustment. Column 1 shows results with the coefficient rank as the outcome variable, while column 2 shows γ when the outcome is the binary indicator variable. All estimates are positive and significant. A 1 standard deviation in average author conservative ideology results in a 16-25 point increase in rank (out of 100), a 30-60 percent increase in the probability that the coefficient is greater than the median in its category, and between .4 and .7 standard deviations within the category. ?? Explain this more.

For comparison, Panel D shows results with the Groundtruth measure of ideology. While all the coefficients are positive and comparable in magnitude to the results in Panel A, the sample of elasticities is, at $N=31$, quite small, and the resulting standard errors make the estimates insignificant at conventional levels. This shows the utility of our text-based measure: with only the groundtruth measure constructed from campaign contributions and petition signings we would not be able to estimate the ideology of very many economists, but using the groundtruth measure together with academic text allows us to predict ideology for many more economists, and thus expand the sample used in this regression considerably.

We explore a number of robustness tests in Tables 10 and 11, across our different coefficient magnitude measures. Column 1 includes fixed effects for category interacted with 5-year intervals capturing date in which the estimate was published, in order to capture observed heterogeneity in

methods, data, or simple improvements in estimates over time. Column 2 uses a measure that ignores the standard errors attached to estimates, and instead uses the simple unweighted average of estimates within a paper. Results in both cases are qualitatively identical to the baseline specification. Column 3 adds an indicator variable for whether the estimate was obtained on US data. While US estimates seem to be in a more conservative direction, the effect of predicted author ideology remains statistically significant with all three measures (albeit sometimes at only 10% significant).

In column 4 we restrict attention to predictions made using the first 50% of the words written by authors, to minimize reverse causality running from empirical results to predicted ideology. These predictions are necessarily going to have more error, as they use less of the available text for each economist. Indeed, 5 papers in our sample are lost as none of the authors have enough text in the first 50% of their writings to estimate ideology. Nonetheless, the results remain positive and statistically significant, despite the attenuation we would expect from the additional prediction error.

In column 5 we adapt split-sample instrumental variables to deal with possible prediction error in our main estimates. While this instrumental variables strategy does not handle endogeneity, it can help address prediction error. Because our independent variable is a prediction of ideology, it has an error, akin to measurement error. We split each author's writings into 2 random samples, and predict ideology in both. Under the assumption that prediction error is orthogonal to the true ideology, then using the ideology in one sample to instrument for the ideology in the other sample will eliminate the resulting attenuation bias. Formally, if the true second stage equation is 4, but we have prediction error in the main independent variable, we will have:

$$\bar{\theta}_j = \bar{\theta}_j^{True} + \eta_{sj}$$

Where η_{sj} is the mean prediction error, $\eta_{sj} = \frac{1}{|N_j|} \sum_{a \in N_j} e_a$, akin to measurement error. And even if η_{sj} is uncorrelated with either the true value of the independent variable or any omitted variable, the estimated coefficient $\hat{\gamma}$ will be attenuated by the well-known factor $\frac{\text{var}(\bar{\theta}^{True})}{\text{var}(\bar{\theta}^{True}) + \text{var}(\eta)} < 1$.²⁴ Thus our coefficients will be too small, relative to the truth.

Our IV strategy mitigates this problem. We split the words used by each author into 2 samples,

²⁴Even though our groundtruth measure is a binary measure, our prediction is continuous, so the measurement error can still be classical, which would not be the case if our prediction was binary.

and estimate two separate, independent predictions of ideology, $\overline{\theta}_j^0$ and $\overline{\theta}_j^1$. Unsurprisingly, both of these measures are highly correlated with each other. To show that the IV eliminates the influence of prediction error.

$$\overline{\theta}_j^g = \overline{\theta}_j^{True} + \delta_s + \eta_{sj}^g, g = 0, 1$$

We then use the $g = 1$ prediction as an instrument for the $g = 0$ prediction. Keeping the covariates δ_s implicit, this results in an IV coefficient given by:

$$\begin{aligned} \gamma^{IV} &= \frac{Cov(\text{coefrank}_{sj}, \overline{\theta}_j^0)}{Cov(\overline{\theta}_j^0, \overline{\theta}_j^1)} \\ &= \frac{Cov(\gamma(\overline{\theta}_j^{True} + \eta_{sj}^g) + \epsilon_{sj}, \overline{\theta}_j^0)}{Cov(\overline{\theta}_j^0, \overline{\theta}_j^1)} = \gamma \end{aligned}$$

since ϵ_{sj} , η_{sj}^1 and η_{sj}^0 are all independent, and thus we have $Cov(\overline{\theta}_j^0, \overline{\theta}_j^1) = Cov(\overline{\theta}_j^{True}, \overline{\theta}_j^1) = Var(\overline{\theta}_j^{True})$.

Turning to the results, the first-stage F-statistic is unsurprisingly extremely strong, and the coefficients are generally larger than the OLS estimates, with somewhat larger standard errors. An exception to this is the JEL-1 topic adjusted estimate, where the IV coefficient is slightly smaller than the OLS estimate and marginally insignificant at conventional levels.

Finally, in column 6 we conduct an identical exercise using “non-political” elasticities, described above. These elasticities are beta convergence, the value of alternative fuels, the effect of institutions on growth, the value of recreational area, and the labor demand elasticity. We again calculate rank within each category of elasticity and estimate the correlation with mean author ideology. We find no significant correlation between predicted ideology and these elasticities, and the point estimates are an order of magnitude smaller than the same specification estimated on the “political” elasticities.

As another check on the general validity of our estimates, in Figure 6 we show results from dropping each category of elasticities one at a time, in order to confirm that no one set of elasticities is driving our result. Across all our different ideology measures, the correlation between mean author ideology and average reported elasticity remains significant at 5%, regardless of which category is dropped.

What to make of the magnitude of these estimates? Consider the labor supply elasticity as a particu-

lar example. Building on Saez (2001), Saez and Diamond (2011) suggest top tax rates of $\tau^* = \frac{1}{1+1.5\epsilon}$, where ϵ is the taxable income elasticity of top income earners. The mean of the Chetty et al. survey on the labor supply elasticity is 0.31, suggesting a top tax rate of 68%. However, the mean (JEL1) ideology among people who estimate taxable income elasticities in this sample is -0.22, more left than average, but with a considerable range, from -0.66 to 0.55. Using our estimate, moving from the most left wing to the most right wing within this sample would change the elasticity by .2 points, changing the optimal top tax rate from 77% to 63%. Extrapolating to the most liberal ideology of -1 to the most conservative ideology of 1, we end up with optimal tax rates from 81% to around 57%. The same standard optimal taxation formula may yield quite different prescriptions depending on the ideology of the researcher producing the elasticity.

If ideology is associated with sorting into fields and methodologies, then policy makers may wish to consider the sensitivity of parameters to partisanship. Following Manski (2003), one might consider constructing “ideological bounds” around an estimate, adjusting for the sorting by ideology into particular fields.

6 Conclusion

There is a robust correlation between patterns of academic writing and political behavior. If in fact partisan political behavior was completely irrelevant to academic economic writing, then academic writing would be a very poor predictor of political ideology. However, our within-topic ideological phrases are not only intuitive, they also predict political behavior well out-of-sample, and even predict the partisanship calculated from completely unrelated Gordon and Dahl IGM survey data. The patterns of individual ideology we document are also of interest, as they suggest that there are in fact professional patterns of ideology in economics, across universities and subfields. While we cannot claim causal identification, we believe our methodology for measuring ideology and the correlations with academic outcomes we have uncovered are informative.

These estimates do not imply that economists are deliberately altering empirical work in favor of preconceived political ideas. Firstly, these correlations could be driven by omitted variables. While we have used ideology measured using previously published papers, if past research findings drive both measured ideology as well as current research results then that would confound these estimates.

However, given the stability of our ideology scores over careers, we think it more likely that ideology is driving selection into methodologies that generate particular estimates. Economists with predetermined policy preferences could select into methodologies that yield parameter estimates that justify those policy preferences. Or it could be other, omitted factors that determine both political behavior and parameter estimates. For example, our results could be driven by methodology or field-specific human capital. If particular skill-sets and cognitive abilities yield comparative advantages in the use of certain methodologies, and if they are also associated with different worldviews and political beliefs, then the sorting of political ideology across research findings could be relatively efficient in the context of that literature. Attempting to alter patterns of political ideology across fields could in fact worsen research productivity.

Of course, economists may not know themselves if their work is partisan. The advantage of our approach is that we do not need to rely solely on direct expert advice to discriminate phrases by ideological orientation. A drawback is that we instead use variation in observed political behavior among economists, which may be both a coarse projection of complex underlying beliefs, as well as missing ideological beliefs that do not vary across economists in our sample.

Theoretical research on the determinants of ideology in academic research would be welcome. A promising place to start could be the literature on self-censorship and political correctness (Loury 1994, Morris 2001), where academic writing does not just reveal the results of research, but also implicit loyalties and beliefs. As academic economic articles have potentially multiple audiences, from specialists to general interest economists to policy makers and journalists, modelling the resulting trade-offs in choosing what to say and how to explain ideas, methods, and results could be a fruitful area of research.

One potential route for combining theory with the empirical approach in this paper is to develop methods for “ideological adjustments” that incorporate the effects of sorting into summaries of parameter estimates, such as weighting results counter to an author’s ideology more highly. However, we are skeptical that any purely technical solution to this fundamentally political problem can be found. Debates in economics about the extent of intervention in the market or the merits of various policies will not be resolved by better methodologies alone. A simpler alternative is to understand partisanship in economic arguments as part of the democratic process of policy making, and economics itself as not

above politics.

References

- Ahmed, Amr and Eric P. Xing**, “Staying informed: supervised and semi-supervised multi-view topical analysis of ideological perspective,” in “Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing” Association for Computational Linguistics 2010, pp. 1140–1150.
- Andrews, Isaiah and Maximilian Kasy**, “Identification of and correction for publication bias,” Technical Report, National Bureau of Economic Research 2017.
- Archak, N., A. Ghose, and P. G. Ipeirotis**, “Deriving the pricing power of product features by mining consumer reviews,” *Management Science*, 2011, 57 (8), 1485–1509.
- Biau, Gerard**, “Analysis of a random forests model,” *The Journal of Machine Learning Research*, 2012, 98888, 1063–1095.
- Bishop, Christopher M**, *Pattern recognition and machine learning*, springer, 2006.
- Blaug, Mark**, *Economic theory in retrospect*, Cambridge university press, 1990.
- Blei, D. M.**, “Probabilistic topic models,” *Communications of the ACM*, 2012, 55 (4), 77–84.
- **and J. D. Lafferty**, “Topic models,” *Text mining: classification, clustering, and applications*, 2009, 10, 71.
- **and J. D. McAuliffe**, “Supervised topic models,” *arXiv preprint arXiv:1003.0783*, 2010.
- Blei, David M., Andrew Y. Ng, and Michael I. Jordan**, “Latent dirichlet allocation,” *the Journal of machine Learning research*, 2003, 3, 993–1022.
- Bonica, A.**, “Estimating Ideological Positions of Candidates and Contributors from Campaign Finance Records,” 2011. Stanford University Mimeo.
- , “Ideology and interests in the political marketplace,” *Social Science Research Network*, 2012.
- Breiman, Leo**, “Random forests,” *Machine learning*, 2001, 45 (1), 5–32.
- , “Random Forests,” *Machine Learning*, 2001-10-01, 45 (1), 5–32.
- Cardiff, Christopher F and Daniel B Klein**, “Faculty partisan affiliations in all disciplines: A voter-registration study,” *Critical Review*, 2005, 17 (3), 237–255.
- Chetty, Raj**, “Sufficient Statistics for Welfare Analysis A Bridge Between Structural and Reduced-Form Methods,” *Annual Review of Economics*, 2009, 1, 451–488.
- , “Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply,” *Econometrica*, 2012, 80 (3), 969–1018.
- , “Yes, Economics Is a Science,” *The New York Times*, 2013-10-20.

- Chun, Hyonho and Sunduz Keles**, “Sparse partial least squares regression for simultaneous dimension reduction and variable selection,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2010, 72 (1), 3–25.
- Coats, Alfred William**, “The role of scholarly journals in the history of economics: An essay,” *Journal of Economic Literature*, 1971, 9 (1), 29–44.
- Coats, AW**, “The politics of political economists: comment,” *The Quarterly Journal of Economics*, 1960, pp. 666–669.
- Colander, D.**, “The making of an economist redux,” *The Journal of Economic Perspectives*, 2005, 19 (1), 175–198.
- Corak, Miles**, “Do Poor Children Become Poor Adults? Lessons from a Cross Country Comparison of Generational Earnings Mobility,” 2006.
- Costinot, Arnaud and Andrés Rodríguez-Clare**, “Trade Theory with Numbers: Quantifying the Consequences of Globalization,” *Handbook of International Economics*, 2014, 4, 197.
- Diamond, Peter and Emmanuel Saez**, “The case for a progressive tax: from basic research to policy recommendations,” *The Journal of Economic Perspectives*, 2011, pp. 165–190.
- Dorfman, Robert**, “Thomas Robert Malthus and David Ricardo,” *The Journal of Economic Perspectives*, 1989, pp. 153–164.
- Duppe, Till and E Roy Weintraub**, “Siting the New Economic Science: The Cowles Commission’s Activity Analysis Conference of June 1949,” *Science in Context*, 2013.
- Eisenstein, Jacob, Amr Ahmed, and Eric P. Xing**, “Sparse additive generative models of text,” in “Proceedings of the 28th International Conference on Machine Learning (ICML-11)” 2011, pp. 1041–1048.
- Ellison, Glenn**, “The Slowdown of the Economics Publication Process,” *Journal of Political Economy*, 2002, 110 (5).
- , “How does the market use citation data? The Hirsch index in economics,” 2010.
- Fawcett, Tom**, “An introduction to ROC analysis,” *Pattern Recognition Letters*, 2006, 27 (8), 861–874.
- Feldstein, Martin**, “Tax avoidance and the deadweight loss of the income tax,” *Review of Economics and Statistics*, 1999, 81 (4), 674–680.
- Fellegi, Ivan and Alan Sunter**, “A Theory for Record Linkage,” *Journal of the American Statistical Association*, 1969, 64 (328), 1183–1210.
- Fisher, Irving**, “Economists in Public Service: Annual Address of the President,” *The American Economic Review*, 1919, 9 (1), 5–21.
- Flach, P.**, “ROC analysis,” *Encyclopedia of Machine Learning*, 2010, pp. 869–74.

- Fourcade, Marion, Etienne Ollion, and Yann Algan**, “The Superiority of Economists,” *Journal of Economic Perspectives*, 2014, 28 (4), 1–29.
- Friedman, Jerome H.**, “Stochastic gradient boosting,” *Computational Statistics & Data Analysis*, 2002, 38 (4), 367–378.
- Friedman, Milton**, *Essays in positive economics*, University of Chicago Press, 1953.
- Fuchs, V. R., A. B. Krueger, and J. M. Poterba**, “Economists’ views about parameters, values, and policies: Survey results in labor and public economics,” *Journal of Economic Literature*, 1998, 36 (3), 1387–1425.
- Gans, Joshua**, *Publishing economics: Analyses of the academic journal market in economics*, Edward Elgar Publishing, 2000.
- Garcia, E.**, *A Tutorial on Correlation Coefficients*, URL, 2010.
- Gentzkow, Matthew and Jesse M. Shapiro**, “What Drives Media Slant? Evidence From U.S. Daily Newspapers,” *Econometrica*, 2010, 78 (1), 35–71.
- Gordon, Roger and Gordon B Dahl**, “Views among Economists: Professional Consensus or Point-Counterpoint?,” *American Economic Review*, 2013, 103 (3), 629–635.
- Gormley, Matthew R., Mark Dredze, Benjamin Van Durme, and Jason Eisner**, “Shared components topic models,” in “Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies” Association for Computational Linguistics 2012, pp. 783–792.
- Grimmer, J. and B. M. Stewart**, “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts,” *Political Analysis*, 2013, 21 (3), 267–297.
- Grun, B. and K. Hornik**, “topicmodels: An R package for fitting topic models,” *Journal of Statistical Software*, 2011, 40 (13), 1–30.
- Guthrie, David, Ben Allison, Wei Liu, Louise Guthrie, and Yorick Wilks**, “A closer look at skip-gram modelling,” in “Proceedings of the 5th international Conference on Language Resources and Evaluation (LREC-2006)” 2006, pp. 1–4.
- Haenlein, M. and A. M. Kaplan**, “A beginner’s guide to partial least squares analysis,” *Understanding statistics*, 2004, 3 (4), 283–297.
- Hastie, Trevor, Robert Tibshirani, and J. Jerome H. Friedman**, “The elements of statistical learning,” 2001.
- , —, and **Jerome Friedman**, *The Elements of Statistical Learning* Springer Series in Statistics, New York, NY, USA: Springer New York Inc., 2001.
- Hedengren, David, Daniel B. Klein, and Carrie Milton**, “Economist Petitions: Ideology Revealed,” *Econ Journal Watch*, 2010, 7 (3), 288–319.

- Hoffman, M. D., D. M. Blei, and F. Bach**, “Online learning for latent dirichlet allocation,” *Advances in Neural Information Processing Systems*, 2010, 23, 856–864.
- Hong, H. and J. D. Kubik**, “Analyzing the analysts: Career concerns and biased earnings forecasts,” *The Journal of Finance*, 2003, 58 (1), 313–351.
- Jelveh, Zubin, Bruce Kogut, and Suresh Naidu**, “Detecting Latent Ideology in Expert Text: Evidence From Academic Papers in Economics,” in “Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing” Association for Computational Linguistics 2014.
- , —, and —, “Detecting Latent Ideology in Expert Text: Evidence From Academic Papers in Economics.,” in “EMNLP” 2014, pp. 1804–1809.
- , —, and —, “Detecting Political Ideology In Economics,” *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1804–1809.
- Jensen, Jacob, Suresh Naidu, Ethan Kaplan, and Laurence Wilse-Samson**, “Political Polarization and the Dynamics of Political Language: Evidence from 130 Years of Partisan Speech [with Comments and Discussion],” *Brookings Papers on Economic Activity*, 2012, pp. 1–81.
- Keane, Michael P**, “Labor supply and taxes: A survey,” *Journal of Economic Literature*, 2011, 49 (4), 961–1075.
- Klamer, Arjo and David C Colander**, *The making of an economist*, Westview Press Boulder, 1990.
- Klein, Daniel B and Charlotta Stern**, “Is There a Free-Market Economist in the House? The Policy Views of American Economic Association Members,” *American Journal of Economics and Sociology*, 2007, 66 (2), 309–334.
- Lafferty, John D and David M Blei**, “Correlated topic models,” in “Advances in neural information processing systems” 2006, pp. 147–154.
- Lin, Wei-Hao, Eric Xing, and Alexander Hauptmann**, “A joint topic and perspective model for ideological discourse,” in “Machine Learning and Knowledge Discovery in Databases,” Springer, 2008, pp. 17–32.
- Loury, Glenn C**, “Self-Censorship in Public Discourse A Theory of “Political Correctness” and Related Phenomena,” *Rationality and Society*, 1994, 6 (4), 428–461.
- Maclin, Richard and David Opitz**, “Popular ensemble methods: An empirical study,” *arXiv preprint arXiv:1106.0257*, 2011.
- Manning, C. D. and H. Schatze**, *Foundations of statistical natural language processing*, MIT press, 1999.
- Manski, Charles F**, *Public Policy in an Uncertain World: Analysis and Decisions*, Harvard University Press, 2013.
- Margo, Robert A**, “The Economic History of the ‘American Economic Review’: A Century’s Explosion of Economics Research,” *The American Economic Review*, 2011, pp. 9–35.

- Margolin, Drew, Yu-Ru Lin, and David Lazer**, “Why so Similar?: Identifying Semantic Organizing Processes in Large Textual Corpora,” *Identifying Semantic Organizing Processes in Large Textual Corpora*, 2013.
- Marneffe, Marie-Catherine De and Christopher D. Manning**, “The Stanford typed dependencies representation,” in “Coling 2008: Proceedings of the workshop on Cross-Framework and Cross-Domain Parser Evaluation” 2008, pp. 1–8.
- Mathur, Aparna, Sita Slavov, and Michael R. Strain**, “Should the Top Marginal Income Tax Rate Be 73 Percent?,” *Tax Notes*, 2012, 137, 676.
- McClelland, Robert and Shannon Mok**, “A review of recent research on labor supply elasticities,” *Working Paper Series Congressional Budget Office*, 2012.
- Mei, Qiaozhu, Xu Ling, Matthew Wondra, Hang Su, and ChengXiang Zhai**, “Topic sentiment mixture: modeling facets and opinions in weblogs,” in “Proceedings of the 16th international conference on World Wide Web” ACM 2007, pp. 171–180.
- Merton, Robert K.**, “Science and technology in a democratic order,” *Journal of legal and political sociology*, 1942, 1 (1-2), 115–126.
- Miller, H. J. and J. Han**, *Geographic data mining and knowledge discovery*, CRC, 2009.
- Monroe, B. L., M. P. Colaresi, and K. M. Quinn**, “Fightin’ words: Lexical feature selection and evaluation for identifying the content of political conflict,” *Political Analysis*, 2008, 16 (4), 372–403.
- Morris, Stephen**, “Political correctness,” *Journal of political Economy*, 2001, 109 (2), 231–265.
- National Bureau of Economic Research**, “Amended and Restated By-Laws of National Bureau of Economic Research, Inc.,” 2010.
- Navarro, Gonzalo**, “A guided tour to approximate string matching,” *ACM Comput. Surv.*, 2001, 33 (1), 31–88.
- Neumark, David and William Wascher**, “Minimum wages and employment: A review of evidence from the new minimum wage research,” 2006.
- Önder, Ali Sina and Marko Terviö**, “Is economics a house divided? Analysis of citation networks,” Technical Report, Working Paper, Department of Economics, Uppsala University 2013.
- Oyer, P.**, “The Macro-Foundations of Microeconomics: Initial Labor Market Conditions and Long-Term Outcomes for Economists,” 2006.
- Pang, B. and L. Lee**, *Opinion mining and sentiment analysis*, Now Pub, 2008.
- Paul, Michael and Roxana Girju**, “A two-dimensional topic-aspect model for discovering multi-faceted topics,” *Urbana*, 2010, 51, 61801.
- Paul, Michael J. and Mark Dredze**, “Factorial LDA: Sparse Multi-Dimensional Text Models.,” in “NIPS” 2012, pp. 2591–2599.

- and —, “Drug extraction from the web: Summarizing drug experiences with multi-dimensional topic models,” in “Proceedings of NAACL-HLT” 2013, pp. 168–178.
- Pennacchiotti, M. and A. M. Popescu**, “Democrats, republicans and starbucks aficionados: user classification in twitter,” in “Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining” 2011, pp. 430–438.
- Perito, Daniele, Claude Castelluccia, Mohamed Ali Kaafar, and Pere Manils**, “How Unique and Traceable Are Usernames?,” in Simone Fischer-Hübner and Nicholas Hopper, eds., *Privacy Enhancing Technologies*, number 6794. In ‘Lecture Notes in Computer Science.’, Springer Berlin Heidelberg, 2011, pp. 1–17.
- Perlich, Claudia, Foster Provost, and Jeffrey S. Simonoff**, “Tree induction vs. logistic regression: a learning-curve analysis,” *The Journal of Machine Learning Research*, 2003, 4, 211–255.
- Pfleiderer, Paul**, “Chameleons: The Misuse of Theoretical Models in Finance and Economics,” 2014.
- Piketty, T. and E. Saez**, “How progressive is the US federal tax system? A historical and international perspective,” 2006.
- Poole, Keith T. and Howard Rosenthal**, “A Spatial Model for Legislative Roll Call Analysis,” *American Journal of Political Science*, 1985, 29 (2), 357–384.
- Poon, H. and P. Domingos**, “Unsupervised semantic parsing,” in “Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1” 2009, pp. 1–10.
- Porter, M. F.**, “An algorithm for suffix stripping,” *Program: electronic library and information systems*, 1980, 14 (3), 130–137.
- Rabinovich, Maxim and David Blei**, “The Inverse Regression Topic Model,” in “Proceedings of The 31st International Conference on Machine Learning” 2014, pp. 199–207.
- Rahman, M. Shafiur**, *Food properties handbook*, CRC press, 2010.
- Ramey, Valerie A.**, “Can Government Purchases Stimulate the Economy?,” *Journal of Economic Literature*, 2011-09, 49 (3), 673–85.
- Ratcliff, John W. and D.E. Metzener**, “Pattern Matching: the Gestalt Approach,” *Dr. Dobb’s Journal*, 1988.
- Rivlin, Alice M.**, “Economics and the political process,” *The American Economic Review*, 1987, pp. 1–10.
- Rodrik, Dani**, “When Ideas Trump Interests: Preferences, Worldviews, and Policy Innovations,” *Journal of Economic Perspectives*, 2014-02, 28 (1), 189–208.
- Ross, Dorothy**, *The origins of American social science*, Vol. 19, Cambridge University Press, 1992.
- Rothbard, Murray N.**, “The politics of political economists: Comment,” *The Quarterly Journal of Economics*, 1960, pp. 659–665.

- Saez, Emmanuel**, “Using elasticities to derive optimal income tax rates,” *The review of economic studies*, 2001, 68 (1), 205–229.
- Sapienza, Paola and Luigi Zingales**, “Economic Experts versus Average Americans,” *The American Economic Review*, 2013, 103 (3), 636–642.
- Schumpeter, Joseph A**, “Science and ideology,” *American Economic Review*, 1949, 39 (2), 345–359.
- Stigler, G. J.**, “The economist and the state,” *The American Economic Review*, 1965, 55 (1), 1–18.
- Stigler, George**, “The Politics of Political Economists,” *The Quarterly Journal of Economics*, 1959, 73 (4), 522–532.
- Stigler, George J**, “The influence of events and policies on economic theory,” *The American Economic Review*, 1960, pp. 36–45.
- **and Claire Friedland**, “The citation practices of doctorates in economics,” *The Journal of Political Economy*, 1975, pp. 477–507.
- Taddy, Matt**, “Multinomial Inverse Regression for Text Analysis,” *Journal of the American Statistical Association*, 2013, 108 (503), 755–770.
- Terviö, Marko**, “Divisions within academia: evidence from faculty hiring and placement,” *Review of Economics and Statistics*, 2011, 93 (3), 1053–1062.
- Tetlock, P. C.**, “Giving content to investor sentiment: The role of media in the stock market,” *The Journal of Finance*, 2007, 62 (3), 1139–1168.
- , “All the news that’s fit to reprint: Do investors react to stale information?,” *Review of Financial Studies*, 2011, 24 (5), 1481–1512.
- , **M. Saar-Tsechansky, and S. Macskassy**, “More than words: Quantifying language to measure firms’ fundamentals,” *The Journal of Finance*, 2008, 63 (3), 1437–1467.
- Timmermann, Allan**, “Forecast combinations,” *Handbook of economic forecasting*, 2006, 1, 135–196.
- Varian, Hal R**, “Big data: New tricks for econometrics,” *The Journal of Economic Perspectives*, 2014, 28 (2), 3–27.
- Wang, William Yang, Elijah Mayfield, Suresh Naidu, and Jeremiah Dittmar**, “Historical analysis of legal opinions with a sparse mixed-effects latent variable model,” in “Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1” Association for Computational Linguistics 2012, pp. 740–749.
- Weinstein, Michael**, “Economists and the Media,” *The Journal of Economic Perspectives*, 1992, 6 (3), 73–77.
- Weyl, E Glen and Michal Fabinger**, “Pass-through as an economic tool: principles of incidence under imperfect competition,” *Journal of Political Economy*, 2013, 121 (3), 528–583.

Wilson, T., J. Wiebe, and P. Hoffmann, “Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis,” *Computational linguistics*, 2009, 35 (3), 399–433.

Winkler, William E., “Overview of record linkage and current research directions,” 2006. Background Paper.

Wold, Herman, “Soft modeling by latent variables: The nonlinear iterative partial least squares approach,” *Perspectives in probability and statistics, papers in honour of MS Bartlett*, 1975, pp. 520–540.

Zingales, Luigi, “Preventing Economists Capture,” in Daniel "Carpenter and David" Moss, eds., *Preventing Regulatory Capture: Special Interest Influence and How to Limit it.*, 2014.

7 Tables

Table 1: Petition and contribution patterns for 441 AEA members present in both datasets.

| Contributions | Petitions | | |
|----------------------|-------------------|------------------|--------------------|
| | Left-Leaning (+1) | Undetermined (-) | Right-Leaning (-1) |
| Left-Leaning (+1) | 220 | 5 | 62 |
| Undetermined (-) | 1 | 0 | 3 |
| Right-Leaning (-1) | 2 | 0 | 148 |

Table 2: Top 50 bigrams and trigrams by strength of χ^2 correlation with no topic adjustment.

| Left-Leaning Bigrams | Right-Leaning Bigrams | Left-Leaning Trigrams | Right-Leaning Trigrams |
|----------------------|-----------------------|-------------------------|--------------------------|
| mental_health | public_choic | post_keynesian_econom | yes_yes_yes |
| post_keynesian | stock_return | public_polic_i_analys | journal_law_econom |
| child_care | feder_reserv | polic_i_analys_politiqu | journal_financi_econom |
| labor_market | yes_yes | analys_politiqu_vol | anna_j_schwartz |
| health_care | market_valu | journal_post_keynesian | initi_public_offer |
| work_time | journal_financi | paper_econom_activ | polit_scienc_review |
| keynesian_econom | bank_note | brook_paper_econom | american_polit_scienc |
| high_school | money_suppli | industri_labor_relat | money_credit_bank |
| polic_i_analys | free_bank | mental_health_care | journal_monetari_econom |
| analys_politiqu | liquid_effect | journal_econom_issu | monetari_gold_stock |
| politiqu_vol | journal_financ | low_birth_weight | american_journal_polit |
| birth_weight | median_voter | high_perform_work | georg_mason_univers |
| labor_forc | law_econom | high_school_graduat | journal_polit_scienc |
| journal_post | vote_share | mental_health_servic | under_bretton_wood |
| latin_america | war_spend | labor_relat_review | academ_publish_print |
| mental_ill | journal_law | canadian_public_polic_i | resal_price_mainten |
| medic_care | money_demand | intern_labor_market | journal_money_credit |
| labour_market | gold_reserv | labor_market_outcom | springer_public_choic |
| social_capit | anna_j | politiqu_vol_xxix | kluwer_academ_publish |
| singl_mother | switch_cost | econom_issu_vol | literatur_vol_xxxvi |
| brook_paper | mutual_fund | robust_standard_error | southern_econom_journal |
| human_resourc | polit_scienc | health_servic_research | yes_no_yes |
| paper_econom | financi_econom | vol_xxix_no | bank_hold_compani |
| substanc_abus | transact_cost | health_care_system | rate_tax_rate |
| african_american | price_level | labor_forc_particip | financi_statist_yearbook |
| wage_inequ | insid_trade | labor_product_growth | jame_m_buchanan |
| statist_canada | j_schwartz | capit_account_liber | risk_free_rate |
| men_women | money_credit | cambridg_ma_nber | vol_xxxvi_decemb |
| hazard_wast | rent_seek | journal_human_resourc | gold_standard_period |
| psychiatr_disord | note_issu | current_account_balanc | money_suppli_shock |
| cohort_size | monetari_econom | labor_forc_growth | voter_ideal_point |
| unemploy_rate | supra_note | incom_tax_schedul | studi_public_choic |
| minimum_wage | custom_union | econom_polic_i_institut | yes_yes_no |
| welfar_reform | initi_public | live_wage_ordin | buchanan_jame_m |
| industri_labor | fiat_money | low_incom_famili | aggreg_demand_shock |
| labour_suppli | pecuniari_extern | journal_econom_perspect | month_quarter_annual |
| reserv_wage | stock_price | effect_child_care | review_financi_studi |
| new_keynesian | journal_polit | high_school_dropout | uniform_state_law |
| labor_relat | abnorm_return | institut_intern_econom | secur_exchang_commiss |
| labor_suppli | base_money | signif_percent_level | monetari_polic_i_shock |

Table 3: Phrases from CTM-50 Topic 34: Wages

| Topic Phrases | Left-Leaning Phrases | Right-Leaning Phrases |
|------------------|----------------------|-----------------------|
| minimum_wage | child_care | overtim_pay |
| hour_work | work_time | school_year |
| child_care | lone_mother | overtim_hour |
| food_stamp | singl_mother | public_hous |
| labor_suppli | labour_market | hous_program |
| work_hour | new_orlean | year_employ |
| welfar_reform | mental_health | overtim_premium |
| wage_increas | welfar_reform | voucher_program |
| control_group | welfar_recipi | peopl_disabl |
| comparison_group | polic_i_analys | hous_assist |
| welfar_benefit | analys_politiqu | opportun_cost |
| child_support | live_wage | incom_limit |
| effect_minimum | politiqu_vol | support_payment |
| welfar_recipi | labour_suppli | administr_data |
| time_limit | public_assist | great_depress |
| wage_rate | singl_parent | work_council |
| singl_mother | marri_mother | effect_school |
| hour_week | fix_cost | hous_subsidi |
| journal_human | center_care | work_overtim |
| estim_effect | public_polic_i | substitut_effect |

Table 4: Phrases from CTM-50 Topic 49: Business cycles

| Topic Phrases | Left-Leaning Phrases | Right-Leaning Phrases |
|-------------------|----------------------|-----------------------|
| stedi_state | post_keynesian | social_secur |
| busi_cycl | keynesian_econom | period_t |
| journal_econom | labor_market | fiat_money |
| doe_not | journal_post | laissez_fair |
| adjust_cost | new_keynesian | money_hold |
| valu_function | effect_demand | public_choic |
| econom_review | long_run | capit_stock |
| american_econom | fiscal_polic | pecuniari_extern |
| technolog_shock | general_theori | public_good |
| gener_equilibrium | firm_size | price_path |
| decis_rule | real_wage | tax_rate |
| econom_theori | aggreg_demand | price_distort |
| journal_polit | keynesian_theori | monetari_econom |
| consumpt_good | industri_relat | tax_system |
| econom_studi | market_power | hold_period |
| dynam_model | labor_demand | durabl_good |
| polit_economi | keyn_p | govern_debt |
| review_econom | labor_forc | factor_input |
| equilibrium_model | modern_technolog | rate_return |
| market_clear | gross_invest | wealth_transfer |

Table 5: Predictive Performance of Topic-Adjusted Models

| Panel A: Full Groundtruth Dataset | | | | | |
|--------------------------------------|---------------|--------------------|------------|-----------------|---------------------|
| (1) Model | (2) Topics | (3) Correlation | (4) AUC | (5) 95% C.I. | (6) Mean Phrases |
| JEL1 | 19 | 0.461 | 0.773 | (0.754, 0.792) | 25, 721 |
| JEL2 | 93 | 0.405 | 0.742 | (0.722, 0.762) | 7, 631 |
| CTM30 | 28 | 0.455 | 0.770 | (0.751, 0.789) | 18, 276 |
| CTM50 | 47 | 0.438 | 0.761 | (0.741, 0.78) | 10, 387 |
| CTM100 | 90 | 0.413 | 0.749 | (0.729, 0.769) | 4, 848 |
| No Topic | 1 | 0.476 | 0.776 | (0.757, 0.795) | 124, 477 |
| Panel B: First 50% of Papers Dataset | | | | | |
| (1) Model | (2) Topics | (3) Correlation | (4) AUC | (5) 95% C.I. | (6) Mean Phrases |
| JEL1 | 18 | 0.410 | 0.741 | (0.72, 0.761) | 13, 895 |
| JEL2 | 78 | 0.358 | 0.710 | (0.689, 0.732) | 4, 091 |
| CTM30 | 28 | 0.400 | 0.734 | (0.713, 0.754) | 9, 725 |
| CTM50 | 46 | 0.387 | 0.728 | (0.707, 0.749) | 5, 582 |
| CTM100 | 84 | 0.358 | 0.710 | (0.688, 0.731) | 2, 529 |
| No Topic | 1 | 0.440 | 0.753 | (0.733, 0.773) | 78, 209 |

This table compares predictive performance between topic mappings. Listed are (1) the model name (2) the number of topics in the mapping used for prediction (3) the correlation between ground-truth and predicted out-of-sample ideologies (4) the Area Under the Curve (5) the bootstrapped confidence interval for (4), and (6) the average number of phrases per topic that pass the χ^2 filter.

Table 6: Correlation Between Author Ideology and IGM Responses

| Panel A: Ideology (No Topic) Correlation with IGM Responses | | | | | | |
|---|----------------------|---------------------|---------------------|----------------------|----------------------|---------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| main | | | | | | |
| Ideology (No Topic) | 0.535*** (0.154) | 1.240*** (0.415) | 1.021*** (0.387) | 0.437*** (0.137) | 0.590*** (0.201) | 0.699*** (0.145) |
| Question FE | No | Yes | Yes | No | Yes | Yes |
| Controls | No | No | Yes | No | No | Yes |
| Log-Likelihood | -383.2 | -138.1 | -126.4 | -1077.8 | -763.2 | -744.1 |
| Observations | 598 | 438 | 438 | 715 | 715 | 715 |
| Individuals | 39 | 39 | 39 | 39 | 39 | 39 |
| Panel B: Ideology (JEL 1) Correlation with IGM Responses | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| main | | | | | | |
| Ideology (JEL 1) | 0.903*** (0.255) | 2.349*** (0.727) | 1.619*** (0.603) | 0.761*** (0.215) | 1.091*** (0.327) | 1.084*** (0.269) |
| Question FE | No | Yes | Yes | No | Yes | Yes |
| Controls | No | No | Yes | No | No | Yes |
| Log-Likelihood | -383.5 | -137.4 | -126.4 | -1077.9 | -762.6 | -744.3 |
| Observations | 598 | 438 | 438 | 715 | 715 | 715 |
| Individuals | 39 | 39 | 39 | 39 | 39 | 39 |
| Panel C: Ideology (CTM-50) Correlation with IGM Responses | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| main | | | | | | |
| Ideology (LDA 100) | 1.216*** (0.406) | 3.269*** (1.153) | 2.657*** (0.957) | 0.988*** (0.359) | 1.547*** (0.564) | 1.710*** (0.378) |
| Question FE | No | Yes | Yes | No | Yes | Yes |
| Controls | No | No | Yes | No | No | Yes |
| Log-Likelihood | -384.1 | -138.1 | -126.1 | -1078.6 | -762.9 | -743.8 |
| Observations | 598 | 438 | 438 | 715 | 715 | 715 |
| Individuals | 39 | 39 | 39 | 39 | 39 | 39 |
| Panel D: groundtruth Correlation with IGM Responses | | | | | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| main | | | | | | |
| Groundtruth Ideology | 0.274*** (0.0681) | 0.843*** (0.220) | 10.17*** (3.100) | 0.266*** (0.0640) | 0.393*** (0.0819) | 2.187*** (0.417) |
| Question FE | No | Yes | Yes | No | Yes | Yes |
| Controls | No | No | Yes | No | No | Yes |
| Log-Likelihood | -214.2 | -67.63 | -57.92 | -588.1 | -405.8 | -395.3 |
| Observations | 334 | 199 | 199 | 394 | 394 | 394 |
| Individuals | 20 | 20 | 20 | 20 | 20 | 20 |

Standard errors are clustered by economist. Controls include year of Ph.D., indicators for gender, phd university, and washington experience. Columns 1-3 are logit regressions predicting the author as conservative as measured by Gordon and Dahl (2013), while Columns 4-6 are ordered logit regressions using the 5 different levels of agreement with statements coded by Gordon and Dahl (2013) conservative.

Table 7: Correlation Between Author Ideology (Idea100) and CV Characteristics

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|--------------------------|------------------------|-----------------------|-----------------------|-----------------------|------------------------|------------------------|------------------------|----------------------|
| Financial Economics | 0.137*** (0.0291) | 0.116*** (0.0297) | | | 0.110*** (0.0288) | 0.117*** (0.0284) | 0.115*** (0.0291) | 0.0879 (0.0645) |
| Macroeconomics | 0.124*** (0.0199) | 0.108*** (0.0206) | | | 0.119*** (0.0203) | 0.109*** (0.0200) | 0.121*** (0.0203) | 0.106** (0.0502) |
| International Trade | 0.0899*** (0.0249) | 0.0929*** (0.0266) | | | 0.0863*** (0.0247) | 0.0916*** (0.0248) | 0.0866*** (0.0245) | 0.0310 (0.0632) |
| Labor Economics | -0.0892*** (0.0204) | -0.108*** (0.0209) | | | -0.0864*** (0.0202) | -0.0962*** (0.0201) | -0.0848*** (0.0203) | -0.0898 (0.0610) |
| Public Economics | 0.0272 (0.0248) | 0.0327 (0.0251) | | | 0.0225 (0.0245) | 0.0262 (0.0248) | 0.0254 (0.0244) | 0.0711 (0.0558) |
| Development | -0.0380 (0.0252) | -0.0370 (0.0245) | | | -0.0357 (0.0245) | -0.0353 (0.0250) | -0.0353 (0.0245) | -0.0779 (0.0620) |
| Econometrics | 0.0115 (0.0268) | 0.0111 (0.0268) | | | 0.00960 (0.0264) | 0.00307 (0.0261) | 0.00850 (0.0263) | 0.0186 (0.0643) |
| Microeconomic Theory | -0.0419* (0.0250) | -0.0504** (0.0243) | | | -0.0452* (0.0250) | -0.0472* (0.0242) | -0.0465* (0.0249) | -0.0815 (0.0654) |
| Business School | | | 0.0990*** (0.0272) | 0.0401 (0.0290) | 0.0666*** (0.0214) | 0.0477* (0.0252) | 0.0673*** (0.0216) | 0.0926* (0.0517) |
| Saltwater Ever (Tervio) | | | -0.0213 (0.0451) | -0.0368 (0.0408) | -0.0246 (0.0331) | -0.0383 (0.0362) | -0.0223 (0.0330) | -0.0154 (0.130) |
| Freshwater Ever (Tervio) | | | 0.110*** (0.0384) | 0.0939** (0.0393) | 0.0780*** (0.0300) | 0.0758** (0.0352) | 0.0778*** (0.0297) | 0.0392 (0.0715) |
| Latin American Origin | | | 0.113* (0.0578) | 0.0401 (0.0716) | 0.0434 (0.0680) | 0.0327 (0.0628) | 0.0413 (0.0677) | 0.131 (0.107) |
| European Origin | | | 0.0926*** (0.0285) | 0.0704*** (0.0271) | 0.0662*** (0.0246) | 0.0532** (0.0245) | 0.0670*** (0.0245) | 0.0592 (0.0858) |
| Full Professor | | | -0.0307 (0.0226) | 0.000288 (0.0219) | -0.000215 (0.0169) | -0.0102 (0.0201) | -0.000688 (0.0169) | -0.0268 (0.0452) |
| Top 5 Economics Dept | | | -0.00721 (0.0573) | 0.0269 (0.0513) | -0.0243 (0.0194) | 0.00326 (0.0496) | -0.0239 (0.0194) | -0.0719 (0.0435) |
| Groundtruth Ideology | | | | | | | | 0.0522** (0.0233) |
| Department FE | No | Yes | No | No | No | No | No | No |
| University FE | No | No | Yes | Yes | No | Yes | No | No |
| Field FE | No | No | No | Yes | No | No | No | No |
| PhD Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 894 | 894 | 894 | 894 | 894 | 894 | 894 | 262 |

Robust standard errors in parenthesis. Fields refer to self-reported fields on CVs.

Saltwater and Freshwater are constructed from Tervio and are described in the text.

Table 8: Fuchs et al. (1998) Elasticities, Meta-Analyses, and Political Orientations

| Labor/Public | Type of elasticity | Surveys found | Usable data? | Policy Relevant | Political Orientation |
|--------------|-------------------------------|-----------------------------|--------------|-----------------|-----------------------|
| Labor | Job Training | Card. et al. 2015 | No | Yes | - |
| Labor | Job Training | Heckman et al. 1999 | Some | Yes | - |
| Labor | Labor Supply | Bargain & Peichl 2013 | Some | Yes | + |
| Labor | Labor Supply | Chetty et al. 2011 | Yes | Yes | + |
| Labor | Labor Supply | McClelland & Mok 2012 | Some | Yes | + |
| Labor | Labor Supply | Reichling & Whalen 2012 | No | Yes | + |
| Labor | Minimum Wage | Neumark & Wascher 2006 | Yes | Yes | - |
| Labor | Minimum Wage | Belman & Wolfson 2014 | Yes | Yes | - |
| Labor | Union Productivity | Belman & Voos 2004 | No | Yes | - |
| Labor | Union Productivity | Hirsch 2004 | No | Yes | - |
| Labor | Union Productivity | Jarrell & Stanley 1990 | No | Yes | - |
| Labor | Union Productivity | Doucouliagos & Laroche 2000 | Yes | Yes | - |
| Labor | Gender Wage Gap | Stanley & Jarrell 1998 | No | Yes | - |
| Labor | Gender Wage Gap | Stanley & Jarrell 2003 | No | Yes | - |
| Labor | Gender Wage Gap | Weichselbaumer et al. 2005 | Some | Yes | - |
| Labor | Labour Demand | Lichter et al. 2014 | Yes | No | - |
| Public | Elasticity of Gasoline Demand | Brons et al. 2008 | No | Yes | + |
| Public | Elasticity of Gasoline Demand | Espey 1996 | Yes | Yes | + |
| Public | Elasticity of Gasoline Demand | Espey 1998 | Yes | Yes | + |

Summary of meta-analyses used.

Table 9: Correlation Between Author Ideology and Policy-Relevant Elasticity Coefficient Rank

| Panel A: Ideology (No Topic) | | | |
|-----------------------------------|-----------------------|---------------------------------|-----------------------|
| | (1) | (2) | (3) |
| | Coef. Rank (weighted) | High Coef. Indicator (weighted) | Std. Coef. (weighted) |
| Mean Pred. Ideology (notopic) | 0.164*** (0.055) | 0.289*** (0.096) | 0.402** (0.179) |
| Meta-Analysis FE | Yes | Yes | Yes |
| R-squared | 0.05 | 0.08 | 0.03 |
| Observations | 198 | 198 | 198 |
| Ideology Range | 1.94 | 1.94 | 1.94 |
| Panel B: Ideology (CTM100) | | | |
| | (1) | (2) | (3) |
| | Coef. Rank (weighted) | High Coef. Indicator (weighted) | Std. Coef. (weighted) |
| Mean Pred. Ideology (lda100) | 0.244** (0.119) | 0.627*** (0.212) | 0.750* (0.389) |
| Meta-Analysis FE | Yes | Yes | Yes |
| R-squared | 0.04 | 0.09 | 0.03 |
| Observations | 197 | 197 | 197 |
| Ideology Range | 0.97 | 0.97 | 0.97 |
| Panel C: Ideology (JEL1) | | | |
| | (1) | (2) | (3) |
| | Coef. Rank (weighted) | High Coef. Indicator (weighted) | Std. Coef. (weighted) |
| Mean Pred. Ideology (jel1) | 0.221*** (0.085) | 0.403*** (0.151) | 0.702** (0.269) |
| Meta-Analysis FE | Yes | Yes | Yes |
| R-squared | 0.05 | 0.07 | 0.03 |
| Observations | 198 | 198 | 198 |
| Ideology Range | 1.33 | 1.33 | 1.33 |
| Panel D: Ideology (Groundtruth) | | | |
| | (1) | (2) | (3) |
| | Coef. Rank (weighted) | High Coef. Indicator (weighted) | Std. Coef. (weighted) |
| Mean Pred. Ideology (groundtruth) | 0.134 (0.083) | 0.260** (0.093) | 0.318 (0.290) |
| Meta-Analysis FE | Yes | Yes | Yes |
| R-squared | 0.46 | 0.50 | 0.25 |
| Observations | 30 | 30 | 30 |
| Ideology Range | 2.00 | 2.00 | 2.00 |

Robust Standard Errors reported in parenthesis. Ideology is calculated as the mean ideology of the authors, using ideology predicted from papers written prior to the published estimate. Coefficient rank is the rank of the average elasticity reported in the paper in the set of elasticities of the same category. High coefficient is an indicator variable for the paper elasticity being higher than the median elasticity within the same category. Standardized coefficient value is the paper's elasticity normalized by the mean and standard deviation within category.

Table 10: Correlation Between Author Ideology and Policy-Relevant Elasticity High Coefficient Indicator-Robustness

| Panel A: Ideology (No Topic) | | | | | | |
|---|-------------------------|---------------------|---------------------|--------------------|---------------------|------------------|
| | (1) Cat. X 5-Year FE | (2) Unweighted | (3) US control | (4) Early Pred. | (5) IV | (6) Placebo |
| Mean Pred. Ideology (notopic) | 0.298*** (0.103) | 0.166*** (0.055) | 0.294*** (0.109) | | | 0.123 (0.118) |
| US Estimate | | | 0.086 (0.093) | | | |
| Mean Early Pred. Ideology (notopic) | | | | 0.260** (0.118) | | |
| Mean Pred. Ideology (notopic) Sample 2-IV | | | | | 0.545*** (0.193) | |
| Meta-Analysis FE | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.13 | 0.05 | 0.05 | 0.05 | 0.05 | 0.01 |
| Observations | 198 | 198 | 179 | 193 | 198 | 229 |
| Ideology Range | 1.94 | 1.94 | 1.94 | 1.94 | 1.94 | 1.50 |
| F-stat | | | | | 426.37 | |
| Panel B: Ideology (CTM100) | | | | | | |
| | (1) Cat. X 5-Year FE | (2) Unweighted | (3) US control | (4) Early Pred. | (5) IV | (6) Placebo |
| Mean Pred. Ideology (lda100) | 0.620*** (0.218) | 0.256** (0.117) | 0.620*** (0.232) | | | 0.022 (0.186) |
| US Estimate | | | 0.083 (0.093) | | | |
| Mean Early Pred. Ideology (lda100) | | | | 0.462** (0.223) | | |
| Mean Pred. Ideology (lda100) Sample 2-IV | | | | | 0.722*** (0.234) | |
| Meta-Analysis FE | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.13 | 0.04 | 0.06 | 0.05 | 0.07 | 0.00 |
| Observations | 197 | 197 | 178 | 192 | 197 | 227 |
| Ideology Range | 0.97 | 0.97 | 0.89 | 0.97 | 0.97 | 1.20 |
| F-stat | | | | | 967.83 | |
| Panel C: Ideology (JEL1) | | | | | | |
| | (1) Cat. X 5-Year FE | (2) Unweighted | (3) US control | (4) Early Pred. | (5) IV | (6) Placebo |
| Mean Pred. Ideology (jel1) | 0.477*** (0.153) | 0.228*** (0.084) | 0.461*** (0.166) | | | 0.025 (0.155) |
| US Estimate | | | 0.088 (0.092) | | | |
| Mean Early Pred. Ideology (jel1) | | | | 0.408** (0.169) | | |
| Mean Pred. Ideology (jel1) Sample 2-IV | | | | | 0.413* (0.231) | |
| Meta-Analysis FE | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.13 | 0.05 | 0.06 | 0.05 | 0.04 | 0.00 |
| Observations | 198 | 198 | 179 | 193 | 198 | 229 |
| Ideology Range | 1.33 | 1.33 | 1.33 | 1.33 | 1.33 | 1.03 |
| F-stat | | | | | 865.28 | |

Robust Standard Errors, clustered by author combination. Outcome variable is coefficient rank within category. Ideology is calculated as the mean ideology of the authors.

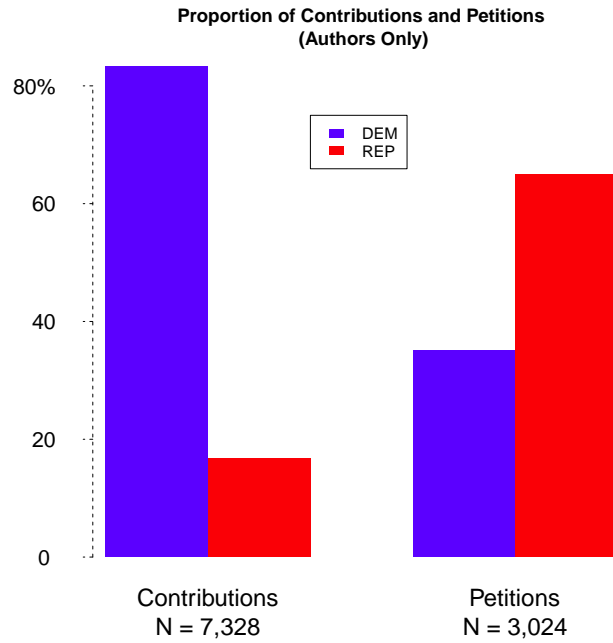
Table 11: Correlation Between Author Ideology and Policy-Relevant Elasticity Coefficient Rank-Robustness

| Panel A: Ideology (No Topic) | | | | | | |
|---|-------------------------|---------------------|--------------------|--------------------|---------------------|-------------------|
| | (1) Cat. X 5-Year FE | (2) Unweighted | (3) US control | (4) Early Pred. | (5) IV | (6) Placebo |
| Mean Pred. Ideology (notopic) | 0.156*** (0.058) | 0.166*** (0.055) | 0.121** (0.056) | | | 0.066 (0.074) |
| US Estimate | | | 0.099* (0.055) | | | |
| Mean Early Pred. Ideology (notopic) | | | | 0.121* (0.066) | | |
| Mean Pred. Ideology (notopic) Sample 2-IV | | | | | 0.175 (0.107) | |
| Meta-Analysis FE | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.09 | 0.05 | 0.06 | 0.03 | 0.02 | 0.01 |
| Observations | 198 | 198 | 179 | 193 | 198 | 229 |
| Ideology Range | 1.94 | 1.94 | 1.94 | 1.94 | 1.94 | 1.50 |
| F-stat | | | | | 426.37 | |
| Panel B: Ideology (CTM100) | | | | | | |
| | (1) Cat. X 5-Year FE | (2) Unweighted | (3) US control | (4) Early Pred. | (5) IV | (6) Placebo |
| Mean Pred. Ideology (lda100) | 0.241** (0.121) | 0.256** (0.117) | 0.189 (0.121) | | | -0.035 (0.120) |
| US Estimate | | | 0.100* (0.055) | | | |
| Mean Early Pred. Ideology (lda100) | | | | 0.213* (0.116) | | |
| Mean Pred. Ideology (lda100) Sample 2-IV | | | | | 0.363*** (0.120) | |
| Meta-Analysis FE | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.08 | 0.04 | 0.05 | 0.03 | 0.04 | 0.01 |
| Observations | 197 | 197 | 178 | 192 | 197 | 227 |
| Ideology Range | 0.97 | 0.97 | 0.89 | 0.97 | 0.97 | 1.20 |
| F-stat | | | | | 967.83 | |
| Panel C: Ideology (JEL1) | | | | | | |
| | (1) Cat. X 5-Year FE | (2) Unweighted | (3) US control | (4) Early Pred. | (5) IV | (6) Placebo |
| Mean Pred. Ideology (jel1) | 0.219** (0.089) | 0.228*** (0.084) | 0.150* (0.090) | | | 0.022 (0.101) |
| US Estimate | | | 0.101* (0.055) | | | |
| Mean Early Pred. Ideology (jel1) | | | | 0.193** (0.091) | | |
| Mean Pred. Ideology (jel1) Sample 2-IV | | | | | 0.168 (0.126) | |
| Meta-Analysis FE | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.08 | 0.05 | 0.05 | 0.03 | 0.03 | 0.00 |
| Observations | 198 | 198 | 179 | 193 | 198 | 229 |
| Ideology Range | 1.33 | 1.33 | 1.33 | 1.33 | 1.33 | 1.03 |
| F-stat | | | | | 865.28 | |

Robust Standard Errors, clustered by author combination. Outcome variable is coefficient rank within category. Ideology is calculated as the mean ideology of the authors.

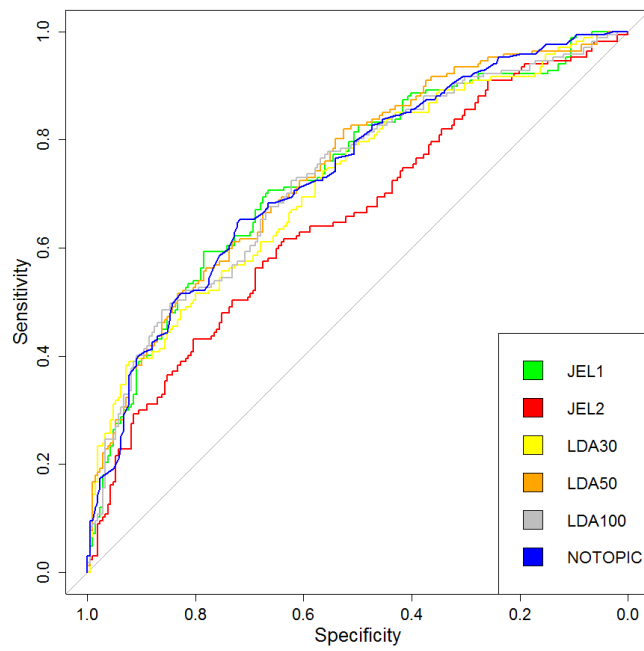
8 Figures

Figure 1: Patterns of Economist Political Behavior



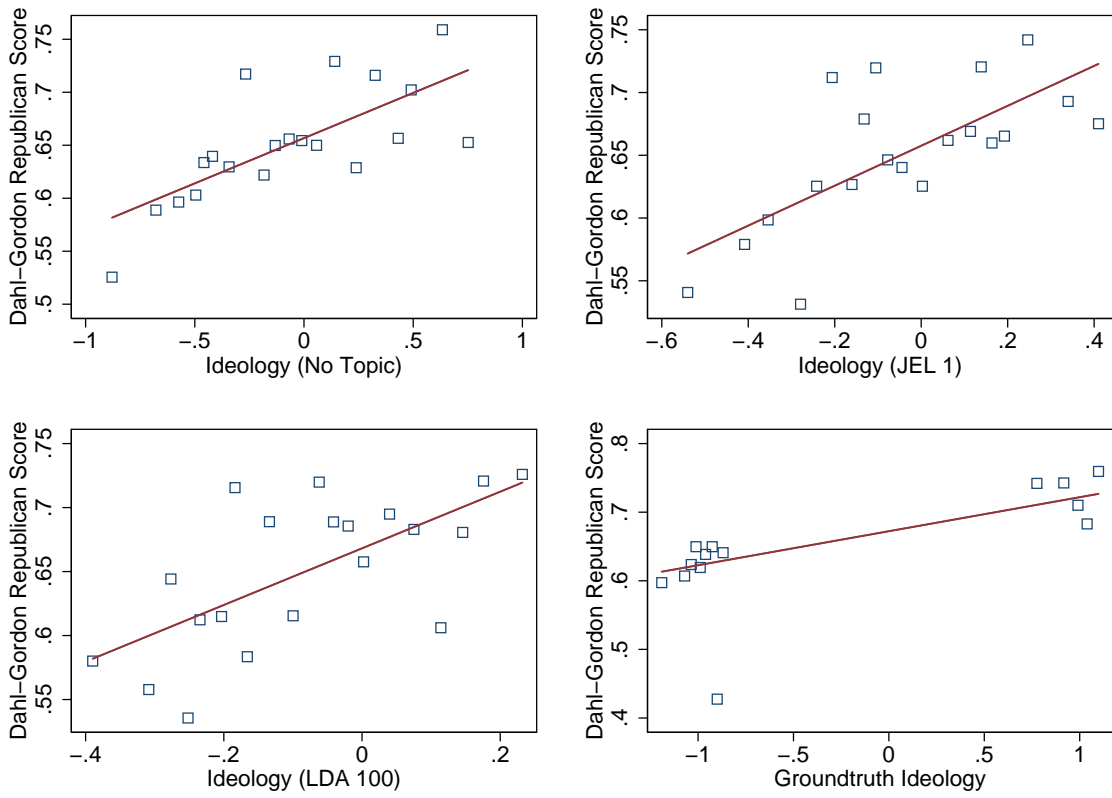
The proportion of campaign contributions to each party is shown on the left and the proportion of signatures on left- and right-leaning petitions is on the right. There were 1,101 authors making contributions and 1,456 signing petitions.

Figure 2: Receiver Operating Curves.



Plots of the true negative rates (specificity) against the true positive rates (sensitivity) for various topic mappings.

Figure 3: Partial Binned Scatterplots of IGM Responses on Ideology Measures.



Figures plot mean IGM conservative answers by vintiles of predicted author ideology, conditional on question fixed effects.

Figure 4: Regression Coefficients On Economist Characteristics

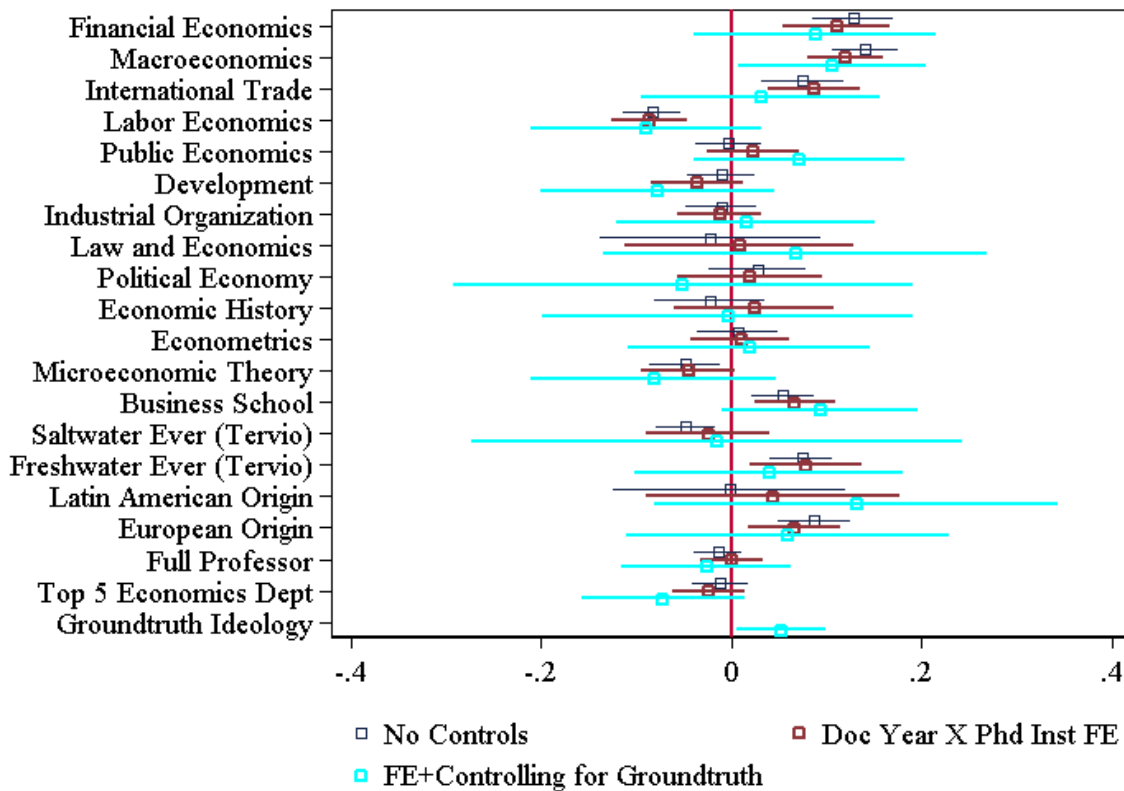
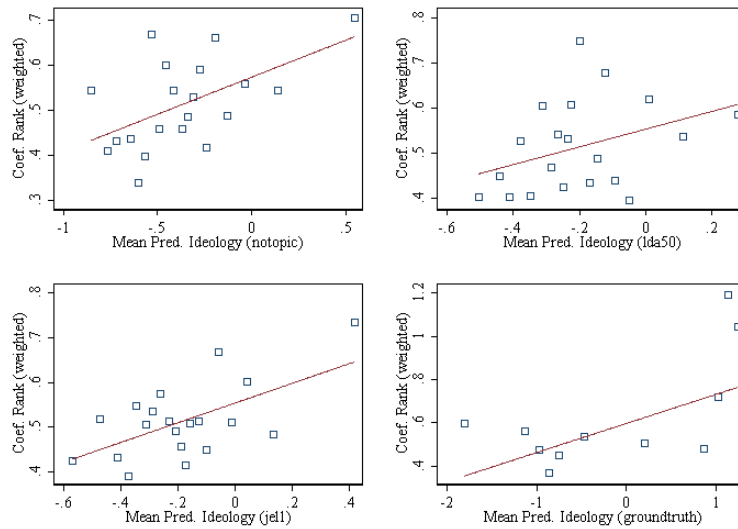


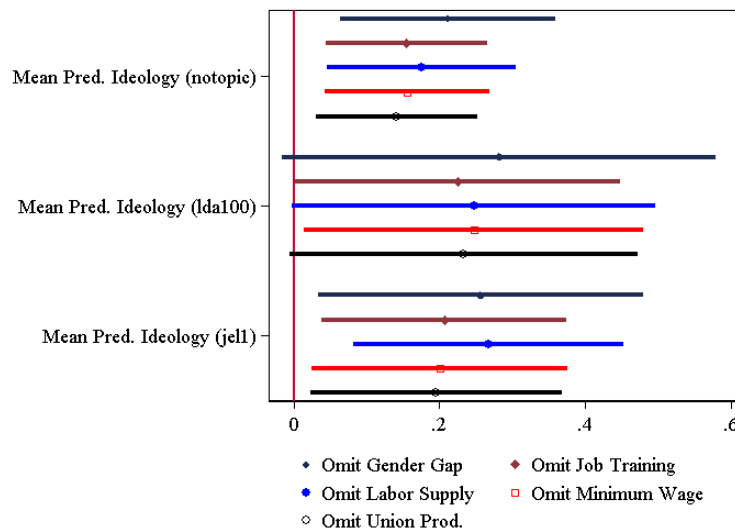
Figure plots coefficients and 95% confidence bands for coefficients on covariates from three regressions. The top set of coefficients include no other controls, the middle set of coefficients controls for 5-year interval when Ph.D. was obtained interacted with Ph.D. institution fixed effects. The last set of coefficients includes the groundtruth control as a covariate, as a test for prediction error correlated with observables.

Figure 5: Binned Scatterplots of Coefficient Rank Against Predicted Ideology (FKP elasticities).



Figures plot mean elasticity rank (within category) by vintiles of predicted author ideology, conditional on meta-analysis fixed effects.

Figure 6: Correlation of Coefficient Rank and Ideology Omitting Each Category of Elasticity



Each estimate shows correlation between predicted ideology measure and coefficient rank, omitting a category of elasticity.

9 Appendix

A.1 Sorting, Peer-review and Partisan Science

In this section we provide a simple analytic framework to clarify what our methodology is estimating and under what assumptions it recovers individual ideology. We consider ideology to be a scalar variable indexing economists from left to right that captures any correlation between partisan political behavior and patterns in academic writing. The model also can be used to shed light on how the professional incentives of academic economists interact with personal ideology to generate ideology measured in academic articles. In our model, economists choose the ideology revealed in their academic papers in order to optimize a combination of ideological preferences, professional incentives, and a preference for being neutral or centrist.²⁵

Suppose individual economists are indexed by ideology θ_i distributed on $U[-1, 1]$. This index corresponds to our “Ground-Truth” measure of ideology, observed partisan political behavior, which we only observe for a small sample. Economists derive utility from publishing papers with ideology $\theta_{P(i)}$, that is close not only to their true ideology θ_i , but to political neutrality, which is given a weight $1 - \Phi$, with $\Phi \in (0, 1)$. A low Φ corresponds to researchers taking pride in being non-partisan experts, and they derive utility from being difficult to pigeonhole politically.

We will use the word “centrist” below to mean the political ideology score close to 0. We do not denote any particular view as “unbiased” or “ideology-free”. Our metric is the real line bounded by -1 and 1 with the center at 0 (or near 0 depending upon the sample or chosen field). This 0 could correspond to the median American voter, who theoretically should be indifferent between the two parties. The center is not necessarily the truth any more than left or right are “biased” and we consequently avoid the word “bias”. Other approaches may wish to pick different points in our ideological spectrum.

In addition, researchers derive utility not only from “researcher objectives” but they also care about professional or career concerns. If ideology (or neutrality) matters for publication, letters of recommendation, or future government and consulting opportunities, then economists may alter the tone and content of their research to be closer to one that is optimal for these pecuniary career outcomes. If academic and publication incentives are paramount, we might expect θ_C to reflect the ideology of editors, senior colleagues, and peer-reviewers.²⁶ We do not take a stand on which of these is most important, nor do we model how the market extracts information about ideology from written work, and instead simply represent the career-optimizing ideology as θ_C , which we weight by $1 - \lambda$ with $\lambda \in (0, 1)$. Combining these three forces, we have total utility given by:

$$V(\theta_{P(i)}, \theta_i) = -\lambda\Phi(\theta_{P(i)} - \theta_i)^2 - \lambda(1 - \Phi)\theta_{P(i)}^2 - (1 - \lambda)(\theta_{P(i)} - \theta_C)^2 \quad (5)$$

If $\lambda = 1$, then there are no professional incentives to be close to career-optimizing θ_C , and so a researcher’s revealed partisanship would be given by $\theta_{P(i)} = \Phi\theta_i$. If $\lambda = 1$ and $\theta_i = 0$ or if $\Phi = 0$, so that the researcher is politically neutral and has no incentives to express ideology, then the economist

²⁵We do not interpret “centrist” as “unbiased” or more accurate, however, as being non-partisan or centrist could in fact be another form of bias.

²⁶In Appendix A.1, we present an extension of this model where we allow economists to sort into fields. Fields are important because they are the source of randomly drawn peers for ones publications and promotion. This results in an equilibrium where fields are completely partitioned, where all left-wing economists are in one field and all right-wing economists are in another. Ideology in this case imperfectly correlates with field.

will choose $\theta_{P(i)} = 0$ in their writing. The difference between Φ and a θ_i captures the difference between being centrist ($\theta_i = 0$) versus wishing to be centrist in published academic work despite being non-centrist ($\Phi = 0, \theta_i \neq 0$), which are potentially two different motivations. If $\theta_C \neq 0$ then it implies that there is a level of partisanship that optimizes professional or career objectives.

$$\theta_{P(i)} = \lambda\Phi\theta_i + (1 - \lambda)\theta_C \quad (6)$$

Generally, if $0 < \lambda < 1$ and $\Phi > 0$, then the economist will choose the ideology of their paper $\theta_{P(i)}$ as a point between their personal ideology and their career maximizing ideology. Equation 6 describes how the ideology observed in a paper is a function of own ideology, as well as the strength of preferences against partisanship (Φ) and career/pecuniary incentives λ . As Φ or λ approaches 0, $\theta_{P(i)}$ approaches θ_C , so that career concerns dominate own ideology, leading the economist to converge on the level of partisanship in their field, department, or other professionally important source. As λ approaches 1 publication ideology will reflect own preferred ideology, which could be 0 if either $\theta_i = 0$, so that the economist is actually centrist, or Φ small, in which case the economist cares about being politically neutral in their work despite having own ideology possibly different from 0. If $\theta_C = 0$ and λ is small, then the institutions are ‘‘Mertonian’’: substantial incentives are provided for even ideological economists to be centered.

Empirically, suppose publication ideology is given by:

$$\theta_{P(i)} = X_{P(i)}\beta + \epsilon \quad (7)$$

, where $X_{P(i)}$ is a vector of text features of publications $P(i)$ written by author i and β is an unknown coefficient vector. Then we have the true model:

$$\theta_i = X_{P(i)}\frac{\beta}{\lambda\Phi} - \frac{1 - \lambda}{\lambda\Phi}\theta_C \quad (8)$$

We do not observe θ_C , so we need an assumption to recover an unbiased predictor of θ_i as a function of $X_{P(i)}$ alone. The first assumption we could make is that θ_C is uncorrelated with $X_{P(i)}$, so we can estimate equation (8) consistently. However, even if this assumption fails, there are slightly weaker assumptions under which we can still proceed to recover a valid predictor of θ_i without necessarily having to identify separately the correct structural coefficients β, λ, Φ .

The second assumption is that career-maximizing ideology is a linear combination of text features of publications and own ideology. Formally, this can be written:

$$\theta_C = X_{P(i)}\beta_C + \alpha_C\theta_i + \nu \quad (9)$$

This may be a strong assumption if there are unobserved characteristics of an economist that predict ideological incentives independent of own ideology that are not revealed in their writing. However, if we include a rich enough set of features of text, which in practice will be topic-specific phrase frequencies, it may be plausible to assume that we absorb the field-specific ideology. The assumption expressed in (9) says that there are enough professional niches so that economists sort into fields, methodologies, or departments according to closeness of ideology, and any remaining ideology that is due to constraints and not preferences are captured by $X_{P(i)}$. Then, using (8) and (9) we can estimate

the following reduced form equation:

$$\theta_i = X_{P(i)}\gamma + \eta \quad (10)$$

Where $\gamma = \frac{\beta - (1-\lambda)\beta_C}{\phi\lambda + (1-\lambda)\alpha_C}$, and a linear regression would recover the best unbiased linear predictor $\hat{\gamma}$. Under the assumption of a valid estimate of γ , we can then forecast $\hat{\theta}_j$, given a document represented by a vector of text features $Z_{P(j)}$. This will be the core of our empirical approach. The main technical issue is that the length of the vector X is larger than the number of observations i , which rules out OLS and requires us to use a dimension-reduction methodology such as partial least squares. We will also use the IGM subsample of economists for whom we observe rich demographic covariates to check whether omission of demographic and professional characteristics introduces important biases in our predicted ideology.

What are possible determinants of θ_C ? We can use this framework to examine how peer-review and sorting may generate a correlation between fields and methodologies and political preferences. Peer-review provides a natural mechanism. If peers act as gatekeepers for publication and promotion within a field or methodology, and peers have ideological preferences, then economists will sort into those fields and methodologies where peers are ideologically sympathetic.

To fix ideas suppose there are two fields F that partition the set of economists, P_L and P_M . Researchers can choose a field prior to publishing a paper. Editors invite peer reviewers at random from the set of economists who have chosen that field. We assume that when peers referee a paper they reject papers that are too far from the ideological mean of researchers in that field. So formally this yields for $F \in \{L, M\}$:

$$\theta_F = E[\theta_i | i \in F] \quad (11)$$

This is a reduced-form way of capturing the pressure towards conformity with the other researchers in a field that peer-review induces. Referees are anonymous, and generally sampled from the population of scholars who have previously worked in that field.

We further assume that the career concerns of researchers are purely determined by field, so that $\theta_C = \theta_F$. An equilibrium in this model is a partition of $-1, 1$ into L and F such that no researcher wishes to change fields. Clearly, from equation 1, each researcher would like to sort into the field that is closest to them in ideology, which is not identical to own ideology only to the extent there is a taste for political neutrality or non-partisanship, i.e. $\Phi \approx 0$. This results in the following proposition.

Proposition: If $\Phi \neq \frac{1}{2}$, there are two classes of equilibria in this model:

1. Degenerate equilibria: ideologies are evenly distributed within each field so both fields have mean ideology 0.
2. Full Sorting equilibria: One field has all economists with ideology < 0 , and so the mean ideology of the field is $-\frac{1}{2}$, while the other field has all economists with ideology > 0 and so has mean ideology $\frac{1}{2}$.

Proof: We first show that each of these is an equilibrium.

Suppose there is a partition P_L, P_M such that $P_M \cap P_L = 0$ and $P_M \cup P_L = [-1, 1]$ and $E[\theta_i | i \in P_j] = 0$. Then every researcher gets the same utility in each field, and so is indifferent between fields. Thus no researcher wishes to switch fields and this is an equilibrium.

Now suppose there is a partition P_L, P_M such that $E[\theta_i | i \in P_M] = \frac{1}{2}$ and $E[\theta_i | i \in P_L] = \frac{-1}{2}$. Then researchers with ideology $\theta_i < 0$ will choose whichever is close to $\Phi\theta_i$, which is L and researchers with ideology $\theta > 0$ will similarly choose M . For all $\theta_i \in M$ we have $\Phi\theta_i \in M$ and $\theta_i \in L$ implies $\Phi\theta_i \in L$. Thus $L = [-1, 0)$ and $F = (0, 1]$ and the partition is an equilibrium.

We next show there can't be any other equilibria. Assume a partition P_M, P_L is an equilibrium where at least one partition P_s has $E[\theta | \theta \in P_s] \neq 0$. We first show that all such partitions must be a pair of intervals $[-1, x], (x, 1]$ (WLOG one closed and one open could be reversed) and then show that $x = 0$ is the only equilibrium. Suppose this equilibrium is not a pair of intervals. Then there is a set x, y, z , such that $x < y < z$, and $x, z \in P_M$ and $y \in P_L$. However, then $|\Phi x - E[\theta | P_M]| \leq |\Phi x - E[\theta | P_L]|$ and $|\Phi z - E[\theta | P_M]| \leq |\Phi z - E[\theta | P_L]|$, but $y \in P_M$ implies $|\Phi y - E[\theta | P_M]| \leq |\Phi y - E[\theta | P_L]|$. This implies that $x, z \leq \frac{\theta_M + \theta_L}{2\Phi}$ while $y \geq \frac{\theta_M + \theta_L}{2\Phi}$ which contradicts $x < y < z$.

Now suppose $[-1, x], (x, 1]$ is an equilibrium. If, WLOG, $x > 0$, then $\theta_L = \frac{x-1}{2}$ and $\theta_M = \frac{x+1}{2}$. Now, for all y such that $\Phi y \leq \frac{1}{2}(\theta_L - \theta_M) = \frac{x}{2}$, we will have $|\Phi y - \theta_L| \leq |\Phi y - \theta_M|$, and so all such y will choose P_L . Similarly y such that $\Phi y \geq \frac{x}{2}$ will choose P_M .

Since $\Phi \neq \frac{1}{2}$ then either $\Phi x < \frac{x}{2}$ and there exists an ϵ such that $\frac{x}{2} > \Phi(x + \epsilon) > 0$ and thus $x + \epsilon$ would choose P_L . Similarly if $\Phi x > \frac{x}{2}$ there is an ϵ such that $\Phi(x - \epsilon) > \frac{x}{2}$ and so $x - \epsilon$ would choose P_M . Thus this cannot be an equilibrium, and so $x \leq 0$. A similar argument shows that $x < 0$ cannot be an equilibrium and hence the only equilibrium partitions are $[-1, 0), [0, 1]$ or $[-1, 0], (0, 1]$.

This model implies that revealed ideology $\theta_{P(i)}$ will in fact be a mix of own ideology θ_i and field ideology θ_L or θ_M . Sorting implies different fields will have distinct political preferences. In this model, while there is sorting, it is not perfect. This motivates including topic-adjusted frequencies in $X_{P(i)}$ as it allows us to use within-field differences in language as predictors for θ_i . Since self-reported fields do not correspond perfectly to paper topics, we can still estimate effects of fields on ideology recovered from within-topic predictions of ideology. While not explicitly in our model, sorting additionally implies that ideology does not change much over the career, and that changes in ideology are not predicted by field.

"Field" in this model could easily be replaced with "Methodology", as long as the peer-review process remains the same. This is of course plausible, as editors will choose referees also on the basis of shared methodology. This is how empirical work, while estimating the same parameter, could still have ideological sorting. If there is selection into methodology that is fine enough (e.g. structural vs reduced-form, micro versus macro estimates), then even estimates of the same parameter could be vulnerable to the same forces of sorting that lead to ideology being correlated with field. A message of this very simple model is that peer-review, together with sorting, may in fact make academic institutions less-"Mertonian".

A.2 Linking Economists to FEC Data

Fuzzy string matching is computationally expensive, so we take the common practical step of creating a candidate set of FEC contributors for each AEA member. We define the candidate set for an AEA member as those FEC contributions where the contributor's last name starts with the same three characters as that of the AEA member.

For each AEA member and his candidate set of FEC contributions, we compute a similarity score between the following variables that appear in both datasets: name, occupation, and employer.²⁷ We

²⁷We use Python's *diffib* module that incorporates a version of the Ratcliff-Obershelp pattern matching algorithm (Ratcliff

map zip codes to latitude-longitude points and compute the distance from the AEA member's location to each candidate FEC contribution. To reduce the likelihood of a match for people with common names, we compute an additional predictor variable which captures the probability that a person's name is unique (Perito et al., 2011). If a name is more likely to appear in the general population, then its predictive ability in determining whether a match exists is reduced.

We model the likelihood that an AEA-FEC pair is a match as a function of the constructed variables from above. We select 1,300 pairs and manually verify if a match exists. We sample 900 of these pairs and estimate the coefficients to a logistic regression model. We repeat this process with new samples one thousand times and for each sample determine the predictive accuracy of the model on the held out set of 400 AEA-FEC pairs. On average, we make a correct prediction 96.5% (s.e. 0.015) of the time. We take the mean values of the parameter sets generated from the regressions and predict matches for the entire dataset. Using this procedure, we are able to identify 21,409 contributions made by 2,884 AEA members. We drop transactions amounts which are less than zero, leaving us with 21,229 contributions from 2,882 members.

The FEC data indicates if a candidate or committee is associated with a particular party. Of the contributions that could be mapped directly to a party, 97% went to either Democrats or Republicans, so we only keep track of three types of recipients: Democrats, Republicans, and Others. Besides parties that are neither Democrat or Republican, the Other category includes cases where the party affiliation is blank or listed as unknown or none. According to this assignment, AEA members made 12,508 contributions to Democrats, 4,420 to Republicans, and 4,301 to Others between 1979 and 2012.

Examining the list of committees in the Others category, it is apparent that a subset of the recipients have known political affiliations. For example, 659 contributions went to ActBlue, which funds Democrats, and 236 contributions were made to Club for Growth, a conservative fundraiser.²⁸ To assign parties to these types of committees in the Others category, we tallied their contributions in a similar manner as above. Our decision rule was that if the committee gave more than 80% to Democrats (Republicans), then we classify its party affiliation as Democrat (Republican). After this step we counted 13,892 contributions to Democrats, 4,670 to Republicans, and 2,667 to Others.

Of these contributions, 7,631 were made by economists who have written a paper in our dataset while 13,595 were made by other AEA members. Based on a visual inspection of the employee and occupation fields in the FEC data, it appears that many of the members in the latter group are either from the government or private industry. Table A.1 provides summary statistics on both author and non-author contributors. At the contribution level, 80.0% go to left-leaning PACs while 16.1% go to right leaning ones. For non-authors these figures are 61.6% and 27.0%, respectively. Of the contributors who have written a paper in our dataset, 11.6% gave to both left-leaning and right-leaning committees compared with 20.3% for non-authors.

and Metzner, 1988) The algorithm works by finding the number of matching characters in the longest common subsequence between two strings. This number is multiplied by two and divided by the total number of characters in the two strings. For example, the distance between 'abcdef' and 'adbecf' is $\frac{2}{3}$ since the longest common subsequence is 'abcf'.

²⁸See <http://www.opensecrets.org/orgs/summary.php?id=D000021806> and <http://www.opensecrets.org/orgs/summary.php?id=D000000763>

A.3 Measuring JEL Topic Prediction Accuracy

We assess our methodology for predicting JEL codes through three standard accuracy measures from information retrieval. For a given paper, let $jel_{econlit}$ be the set of JEL codes from EconLit and $jel_{predict}$ the set of predicted codes. The recall of the Econlit set is defined as

$$\text{recall}_{econlit} = \frac{|jel_{econlit} \cap jel_{predict}|}{|jel_{econlit}|}$$

which is the fraction of EconLit JELs that appear in the prediction set. The precision of the prediction set is defined as

$$\text{precision}_{predict} = \frac{|jel_{econlit} \cap jel_{predict}|}{|jel_{predict}|}$$

which is the fraction of the predicted JELs that appear in the EconLit set. A high recall can be obtained by including more JELs in $jel_{predict}$, but this may decrease the precision if many of the items in $jel_{predict}$ are not in $jel_{econlit}$. For example, if we predicted that each paper belongs to all JELs, then our recall would be perfect but our precision would be quite low. Similarly, it is possible to obtain high precision by including fewer items in $jel_{predict}$, but this may decrease recall if many items in $jel_{econlit}$ are not in $jel_{predict}$. To balance these countervailing forces, precision and recall can be combined into the F -measure which is the weighted harmonic mean of the two:

$$F_{\beta} = (1 + \beta^2) \frac{\text{precision} \cdot \text{recall}}{\beta^2 \cdot \text{precision} + \text{recall}}$$

Setting β to one weights precision and recall equally while a higher (lower) β gives more importance to precision (recall). We set β equal to one for our comparisons. We compute the precision, recall, and $F1$ score for each pair of EconLit and predicted JELs and take the average. As another comparison to our methodology, we exploit the fact that 1,256 of the papers in the NBER database are working paper versions of published works in the JSTOR set for which we have EconLit-assigned codes. Many of these NBER papers have JEL codes that are author-assigned and do not perfectly match up with the EconLit-assignments. We treat the author-assigned codes as predictions for the final codes. We compare how well our predictions overlapped with the EconLit-assigned codes versus the overlap between NBER and EconLit codes. Table A.4 shows that our methodology is better aligned than NBER author self-assignments with EconLit JEL codes.

A.4 Predicting Editorial Authorship

We compare the performance of our algorithm to TAM2/12 (Paul and Girju, 2010) and mview-LDA (Ahmed and Xing, 2010) on the bitterlemons corpus which consists of editorials written by Palestinian and Israeli authors. Both of these are unsupervised topic models which jointly model topic and ideology. We also show results from predicting using a support vector machine, which is a supervised learning model similar to ours but does not account for topics. For more information on the bitterlemons dataset see Lin et al. (2008). We sample 80% of the editorials for the training set and produce predictions with and without topic adjustments (*CTM10*, a topic mapping with 10 topics and *No Topics*). Table A.5 shows how well our model predicts if a test set author is Israeli or Palestinian compared

with results from TAM2/12 and mview-LDA.²⁹ Our algorithm performs on par. Additionally, it does not appear that accounting for topic is necessarily helpful in improving ideology prediction accuracy. We discuss the performance of our prediction algorithm more in our companion paper (Jelveh, Kogut, and Naidu 2014).

²⁹The TAM2/12 and mview-LDA values are taken from the respective papers. In the case of mview-LDA, the value is an estimate from a figure in the paper.

Table A.1: Campaign Contribution Data: Panel A provides summary statistics on AEA member campaign contributions at the contribution level and Panel B provides summary statistics at the member level.

| Panel A: Contribution-Level | | | | | | | |
|-----------------------------|--------|------------|------------|--------------|-------------------------|----------------------|----------------------|
| | N | Dem. Share | Rep. Share | Total Amount | Amount per Contribution | Amount Share to Dem. | Amount Share to Rep. |
| Authors | 7,631 | 80.0% | 16.1% | \$6,151,074 | \$806 | 76.3% | 19.4% |
| Non-Author | 13,595 | 61.6% | 27.0% | \$11,657,804 | \$858 | 64.5% | 26.6% |

| Panel B: Individual-Level | | | | | | | |
|---------------------------|-------|---------------------|-----------------------------|-----------------------------|-------------------|-----------------|-----------------|
| | N | Contrib. per Person | Contrib. per Person to Dem. | Contrib. per Person to Rep. | Amount per Person | Dem. per Person | Rep. per Person |
| Authors | 1,125 | 6.78 | 5.42 | 1.09 | \$5,468 | \$4,172 | \$1,059 |
| Non-Author | 1,757 | 7.74 | 4.77 | 2.08 | \$6,635 | \$4,277 | \$1,761 |

Table A.2: The 35 petitions from Hedengren et al. (2010). The last column is the categorization applied in this paper.

| Petition | Date | Organizer or Sponsor | Category | Signatures | Authors | Political Category |
|---|----------|--|----------|------------|---------|--------------------|
| Support Market Oriented Health Care Reform 1994 | 03/16/94 | The Independent Institute | Augm | 637 | 224 | Rep |
| Oppose Antitrust Protectionism | 06/02/99 | The Independent Institute | Augm | 240 | 101 | Rep |
| Support Market Oriented Health Care Reform 2000 | 03/01/00 | The Independent Institute | Augm | 538 | 226 | Rep |
| Economists for Sweatshops | 07/29/00 | Academic Consortium on International Trade | Augm | 252 | 80 | Other |
| Oppose Death Tax | 05/21/01 | National Taxpayers Union | Augm | 279 | 119 | Rep |
| Scholars Against Sweatshop Labor | 10/22/01 | Political Economy Research Institute | Reduc | 435 | 98 | Other |
| Oppose Bush Tax Cuts | 02/01/03 | Economic Policy Institute | Reduc | 464 | 273 | Dem |
| Oppose Tax Increase | 01/14/04 | National Taxpayers Union | Augm | 116 | 10 | Rep |

| | | | | | | |
|---|----------|---|-------|------|-----|-------|
| Endorse John Kerry for President | 08/25/04 | John Kerry Campaign (Not Sure) | Other | 10 | 10 | Dem |
| Oppose John Kerry for President | 10/13/04 | George W. Bush Campaign (Not Sure) | Other | 367 | 148 | Rep |
| Warning Future of Social Security Increase | 05/11/05 | Cato Institute | Augm | 454 | 155 | Rep |
| Immigration | 06/19/06 | The Independent Institute | Augm | 523 | 183 | Other |
| Support Raising the Minimum Wage | 09/27/06 | The Economic Policy Institute | Reduc | 659 | 317 | Dem |
| Oppose Marijuana Prohibition | 11/30/06 | Marijuana Policy Project | Augm | 554 | 108 | Other |
| Oppose Government Regulation of Internet ("Network Neutrality") | 03/28/07 | AEI-Brookings Joint Center | Augm | 17 | 10 | Rep |
| Statement on Prediction Markets | 05/01/07 | AEI-Brookings Joint Center | Augm | 25 | 10 | Other |
| Economists Against Protectionism | 08/01/07 | The Club for Growth | Augm | 1028 | 320 | Other |
| Oppose "Windfall Taxes" | 10/17/07 | National Taxpayers Union | Augm | 234 | 82 | Rep |
| Support John McCain Economic Plan | 05/11/08 | John McCain Campaign (Not Sure) | Other | 326 | 132 | Rep |
| Raising Some Concerns about Government Bail Out for Mortgages | 09/24/08 | John Cochrane | Other | 230 | 124 | Rep |
| Support Government Bail Out for Mortgages | 10/01/08 | Unknown | Reduc | 76 | 47 | Dem |
| Concerned about Climate Change | 10/07/08 | Nancy Olewiler | Reduc | 254 | 112 | Dem |
| Support Federal Recovery Act | 11/19/08 | Center for Economic and Policy Research | Reduc | 387 | 138 | Dem |
| Oppose Federal Recovery Act | 01/27/09 | Cato Institute | Augm | 203 | 105 | Rep |
| Oppose Budget Reduction in Washington State | 02/19/09 | Washington State Budget & Policy Center | Reduc | 7 | 4 | Dem |
| Support Employee Free Choice Act | 02/24/09 | The Economic Policy Institute | Reduc | 40 | 34 | Dem |
| Support Cap and Trade | 03/04/09 | Southern Alliance for Clean Energy | Reduc | 601 | 142 | Other |
| Replace Federal Income Tax with FairTax | 03/29/09 | FairTax.org | Other | 80 | 24 | Rep |

| | | | | | | |
|--|----------|---|-------|------|-----|-------|
| Support Using Procurement Auctions Over Grant Submissions | 04/13/09 | Paul Milgrom | Other | 64 | 24 | Other |
| Support Government Intervention to Promote Biofuels | 04/21/09 | Union of Concerned Scientists | Reduc | 16 | 11 | Dem |
| Oppose Green Protectionism | 05/08/09 | Atlas Global Initiative for Free Trade Peace and Prosperity | Augm | 1215 | 230 | Rep |
| Fed Independence Petition | 07/15/09 | Wall Street Journal | Other | 183 | 62 | Other |
| Support Tax Increase on Corporations and High Income Persons | 10/07/09 | Oregon Center for Public Policy | Reduc | 36 | 10 | Dem |
| Government Oriented Health Care Reform 2009 | 11/17/09 | Unknown | Reduc | 23 | 19 | Dem |
| Support for a Financial Transactions Tax | 12/03/09 | Center for Economic and Policy Research | Reduc | 204 | 73 | Dem |

Table A.3: Correlation Between Indicator for Groundtruth Sample and Author Characteristics

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|----------------------|-----------------------|-----------------------|-----------------------|----------------------|---------------------|-----------------------|
| Financial Economics | 0.184*** (0.0287) | 0.187*** (0.0299) | 0.185*** (0.0369) | 0.191*** (0.0315) | 0.164*** (0.0289) | | |
| Labor Economics | 0.0453 (0.0343) | 0.0675** (0.0318) | 0.0610 (0.0386) | 0.0597 (0.0379) | 0.0426 (0.0361) | | |
| Macroeconomics | 0.0672** (0.0265) | 0.0686*** (0.0248) | 0.0749*** (0.0279) | 0.0769*** (0.0280) | 0.0805** (0.0405) | | |
| Public Economics | 0.127*** (0.0281) | 0.143*** (0.0317) | 0.148*** (0.0277) | 0.148*** (0.0296) | 0.120*** (0.0344) | | |
| Business School | 0.0269 (0.0338) | 0.0227 (0.0350) | | -0.0230 (0.0435) | -0.0320 (0.0379) | -0.0375 (0.0353) | -0.0196 (0.0477) |
| Saltwater Ever (Tervio) | 0.137*** (0.0349) | 0.129*** (0.0340) | 0.0907* (0.0462) | 0.104*** (0.0363) | 0.0663** (0.0317) | 0.0355 (0.0366) | 0.0434 (0.0408) |
| Freshwater Ever (Tervio) | 0.0274 (0.0405) | 0.0187 (0.0407) | 0.0124 (0.0635) | 0.0504 (0.0401) | 0.0375 (0.0408) | 0.0136 (0.0496) | 0.00505 (0.0515) |
| Doctoral Degree Year | -0.939*** (0.110) | -0.946*** (0.147) | -0.967*** (0.112) | -0.957*** (0.109) | | | |
| Years Between Undergrad and PhD Degrees | | -0.00948 (0.00651) | | | | | -0.00632 (0.00809) |
| Latin American Origin | | -0.0439 (0.102) | | | | | -0.0951 (0.143) |
| European Origin | | 0.000699 (0.0551) | | | | | 0.0268 (0.0636) |
| Full Professor | | -0.00438 (0.0317) | | | | | 0.0275 (0.0405) |
| Department FE | No | No | Yes | No | No | No | No |
| University FE | No | No | No | Yes | Yes | Yes | Yes |
| Field FE | No | No | No | No | No | Yes | Yes |
| PhD Year FE | No | No | No | No | Yes | Yes | Yes |
| Observations | 903 | 855 | 903 | 903 | 977 | 977 | 855 |

Robust standard errors, clustered at the primary field level. Finance, Labor, Macro, and Public all refer to self-reported primary fields on CVs. Saltwater and Freshwater are constructed from Tervio and are described in the text.

Table A.4: Metrics for assessing the accuracy of predicted EconLit-assigned JEL codes.

| Panel A: First-Level JELS | | |
|---------------------------|--------------|------------|
| | NBER Authors | Our Method |
| F_1 | 0.752 | 0.769 |
| Recall | 0.673 | 0.709 |
| Precision | 0.788 | 0.816 |

| Panel B: Second-Level JELS | | |
|----------------------------|--------------|------------|
| | NBER Authors | Our Method |
| F_1 | 0.630 | 0.669 |
| Recall | 0.477 | 0.479 |
| Precision | 0.589 | 0.768 |

Table A.5: Predictive performance (%) on bitterlemons dataset using 80% of the data for training.

| Model | No. Topics | Accuracy (%) |
|-----------|------------|--------------|
| TAM2/12 | 12 | 87.5 |
| mview-LDA | 10 | 94 |
| SVM | 0 | 94 |
| LDA10 | 10 | 91.0 |
| No Topics | 0 | 94.4 |

A.1 Other Descriptive Results

A.1 Robustness of Patterns of Individual Ideology

In this section we document professional patterns of predicted ideology of economists. We are interested in how our predicted measure of political ideology is correlated with field of research, employment and educational history, and national origin. We hypothesize that sorting, perhaps driven by the process of peer-review and evaluation, results in politically similar economists systematically located in particular fields and departments.

As noted in section *, we collected CV data. We reported in Table A.6 that ideological scores show that there is sorting by fields and institutions. These results support the reasoning to adjust the ideological scores by field in the empirical analysis, including that of the influence of ideology on elasticities.

It is possible that these results from the CV reflect nevertheless how we constructed the topic-adjusted ideology variables. Consequently, along with the demographic and professional variables we include here a control for whether or not the author appears in the ground truth sample. We do this to assess the importance of selection into the ground truth sample. Our estimates rely on the assumption that the relationship between n-grams and ideology is the same among the population that contributes and the population that doesn't contribute. Previous work evaluating the problem of non-representative samples in large prediction exercises like ours have generally found the bias to be quite low, on the order of 2%. If the parameters on other coefficients change substantially when this control is included, it would suggest that our sample of campaign contributors and petition signers are different in terms of their relationship between ideology predicted from text and individual characteristics. However, our coefficients are quite similar regardless of whether we control for the ground truth sample indicator or not, and the ground truth variable is always insignificant. While we do not show them here, none of our ideology predictions are

correlated with being in the ground truth sample in simple bivariate regressions either. While the ground truth sample is different on observables from the other economists in our CV sample, as shown in Appendix Table A.3, differential selection of economists into the ground truth sample does not seem to be an important for predicting political preferences.³⁰

A natural concern is that our predicted ideology scores are in fact driven by reverse causality, where research findings are driving support for particular political ideologies. If this were true, then we might expect predicted ideology to change over an economists' career, and ideological differences to narrow over time. We first check the stability of predicted ideology within a given individual economist. To do this, we split each economist's published work into the first and second 50% of their published words and look at the stability of ideology over time. The raw correlation is 0.55. Appendix A.1 shows that ideology is highly correlated over time, even conditional on all the professional variables we have used in this section. Secondly, we check if the distributions of ideologies look different for young, old, tenured, and untenured faculty, and find that the distributions are remarkably similar between old and young faculty (both faculty above and below median age in our sample have variance of 0.21 in the LDA-50 ideology prediction) as well as untenured and tenured faculty (0.21 and 0.22 variance, respectively, also with LDA50 ideology). Thus we believe our ideology scores are not reflecting transient political behavior that responds to available evidence, but instead a more durable characteristic of an individual.

While we have focused on predicting the ideology of individual economists, one might worry how much our results are affected by selection into publication. We can examine this by predicting the ideology of individual articles, and running a parallel analysis at the article level, controlling for journal and year of published article fixed effects. All of our results remain in this specification, and are omitted for space reasons. We can also decompose predicted ideology into journal, year, and individual economist components. We find that authors explain 30% of the total variation, while journals explain 6% and year explains 0.2%. In addition the journal and year variation is largely orthogonal to the author variation. Together, these results suggest that selection by authors of articles into journal or timing of publication is not important for the patterns of predicted individual political preferences we document here.

Table A.6: Department Level CV Summary Statistics

| | mean | sd | count |
|---------------------------|--------|------|-------|
| <hr/> | | | |
| boston_college | | | |
| Ideology (LDA 50), strong | -0.14 | 0.21 | 12 |
| Ideology (JEL 1), strong | -0.15 | 0.25 | 12 |
| <hr/> | | | |
| boston_u | | | |
| Ideology (LDA 50), strong | -0.21 | 0.23 | 26 |
| Ideology (JEL 1), strong | -0.18 | 0.20 | 26 |
| <hr/> | | | |
| brown | | | |
| Ideology (LDA 50), strong | -0.25 | 0.12 | 6 |
| Ideology (JEL 1), strong | -0.12 | 0.23 | 6 |
| <hr/> | | | |
| caltech | | | |
| Ideology (LDA 50), strong | -0.14 | 0.20 | 12 |
| Ideology (JEL 1), strong | -0.13 | 0.25 | 12 |
| <hr/> | | | |
| carnegie_mellon | | | |
| Ideology (LDA 50), strong | -0.16 | 0.22 | 8 |
| Ideology (JEL 1), strong | -0.097 | 0.20 | 8 |
| <hr/> | | | |
| carnegie_mellon_tepper | | | |
| Ideology (LDA 50), strong | 0.12 | 0.18 | 10 |
| Ideology (JEL 1), strong | 0.18 | 0.35 | 10 |
| <hr/> | | | |
| columbia | | | |
| Ideology (LDA 50), strong | -0.22 | 0.19 | 44 |
| Ideology (JEL 1), strong | -0.18 | 0.21 | 44 |
| <hr/> | | | |
| columbia_business_school | | | |
| Ideology (LDA 50), strong | -0.091 | 0.21 | 10 |

³⁰We also found little robust evidence of peer-effects at either the Ph.D. cohort-grad school or cohort-grad school-field level. We did find some evidence of advisor-advisee correlation in ideology, but this was sensitive to the inclusion of advisor field controls.

| | | | |
|---------------------------|--------|-------|----|
| Ideology (JEL 1), strong | -0.033 | 0.22 | 10 |
| cornell | | | |
| Ideology (LDA 50), strong | -0.13 | 0.24 | 24 |
| Ideology (JEL 1), strong | -0.17 | 0.27 | 24 |
| cornell_johnson | | | |
| Ideology (LDA 50), strong | -0.32 | 0.045 | 3 |
| Ideology (JEL 1), strong | -0.45 | 0.21 | 3 |
| dartmouth | | | |
| Ideology (LDA 50), strong | -0.23 | 0.17 | 20 |
| Ideology (JEL 1), strong | -0.26 | 0.18 | 20 |
| dartmouth_tuck | | | |
| Ideology (LDA 50), strong | -0.040 | 0.29 | 5 |
| Ideology (JEL 1), strong | 0.025 | 0.33 | 5 |
| duke | | | |
| Ideology (LDA 50), strong | -0.16 | 0.24 | 40 |
| Ideology (JEL 1), strong | -0.10 | 0.31 | 40 |
| duke_fuqua | | | |
| Ideology (LDA 50), strong | -0.072 | 0.11 | 5 |
| Ideology (JEL 1), strong | 0.0092 | 0.20 | 5 |
| harvard | | | |
| Ideology (LDA 50), strong | -0.20 | 0.17 | 40 |
| Ideology (JEL 1), strong | -0.17 | 0.19 | 40 |
| harvard_business_school | | | |
| Ideology (LDA 50), strong | -0.058 | 0.26 | 26 |
| Ideology (JEL 1), strong | -0.021 | 0.23 | 26 |
| lse | | | |
| Ideology (LDA 50), strong | -0.21 | 0.17 | 20 |
| Ideology (JEL 1), strong | -0.15 | 0.25 | 20 |
| mit | | | |
| Ideology (LDA 50), strong | -0.21 | 0.19 | 29 |
| Ideology (JEL 1), strong | -0.17 | 0.20 | 29 |
| mit_sloan | | | |
| Ideology (LDA 50), strong | -0.062 | 0.15 | 16 |
| Ideology (JEL 1), strong | -0.016 | 0.23 | 17 |
| northwestern | | | |
| Ideology (LDA 50), strong | -0.22 | 0.19 | 29 |
| Ideology (JEL 1), strong | -0.20 | 0.29 | 29 |
| northwestern_kellogg | | | |
| Ideology (LDA 50), strong | 0.13 | 0.10 | 6 |
| Ideology (JEL 1), strong | -0.078 | 0.13 | 6 |
| nyu | | | |
| Ideology (LDA 50), strong | -0.14 | 0.18 | 25 |
| Ideology (JEL 1), strong | -0.094 | 0.17 | 25 |
| nyu_stern | | | |
| Ideology (LDA 50), strong | -0.073 | 0.18 | 35 |
| Ideology (JEL 1), strong | -0.054 | 0.25 | 35 |
| ohio_state | | | |
| Ideology (LDA 50), strong | -0.064 | 0.21 | 20 |
| Ideology (JEL 1), strong | -0.082 | 0.27 | 20 |
| princeton | | | |
| Ideology (LDA 50), strong | -0.14 | 0.21 | 30 |
| Ideology (JEL 1), strong | -0.10 | 0.18 | 30 |
| stanford | | | |

| | | | |
|---------------------------|--------|-------|----|
| Ideology (LDA 50), strong | -0.20 | 0.14 | 31 |
| Ideology (JEL 1), strong | -0.20 | 0.17 | 31 |
| stanford_gsb | | | |
| Ideology (LDA 50), strong | -0.093 | 0.27 | 22 |
| Ideology (JEL 1), strong | -0.082 | 0.30 | 22 |
| u_chicago | | | |
| Ideology (LDA 50), strong | 0.038 | 0.41 | 2 |
| Ideology (JEL 1), strong | -0.081 | 0.23 | 2 |
| u_chicago_booth | | | |
| Ideology (LDA 50), strong | -0.035 | 0.23 | 39 |
| Ideology (JEL 1), strong | 0.048 | 0.25 | 39 |
| u_illinois | | | |
| Ideology (LDA 50), strong | -0.11 | 0.18 | 13 |
| Ideology (JEL 1), strong | -0.100 | 0.30 | 13 |
| u_maryland | | | |
| Ideology (LDA 50), strong | -0.10 | 0.23 | 25 |
| Ideology (JEL 1), strong | -0.084 | 0.24 | 25 |
| u_michigan | | | |
| Ideology (LDA 50), strong | -0.23 | 0.17 | 36 |
| Ideology (JEL 1), strong | -0.26 | 0.17 | 36 |
| u_michigan_ross | | | |
| Ideology (LDA 50), strong | -0.10 | 0.27 | 6 |
| Ideology (JEL 1), strong | -0.097 | 0.23 | 6 |
| u_minnesota | | | |
| Ideology (LDA 50), strong | -0.17 | 0.11 | 16 |
| Ideology (JEL 1), strong | -0.13 | 0.18 | 16 |
| u_penn | | | |
| Ideology (LDA 50), strong | -0.19 | 0.14 | 19 |
| Ideology (JEL 1), strong | -0.18 | 0.22 | 19 |
| u_penn_wharton | | | |
| Ideology (LDA 50), strong | 0.0035 | 0.23 | 22 |
| Ideology (JEL 1), strong | -0.024 | 0.28 | 22 |
| u_rochester | | | |
| Ideology (LDA 50), strong | -0.31 | 0.091 | 6 |
| Ideology (JEL 1), strong | -0.29 | 0.24 | 6 |
| u_southern_california | | | |
| Ideology (LDA 50), strong | -0.17 | 0.18 | 8 |
| Ideology (JEL 1), strong | -0.15 | 0.21 | 8 |
| u_wisconsin | | | |
| Ideology (LDA 50), strong | -0.18 | 0.19 | 19 |
| Ideology (JEL 1), strong | -0.19 | 0.21 | 19 |
| uc_berkeley | | | |
| Ideology (LDA 50), strong | -0.22 | 0.18 | 49 |
| Ideology (JEL 1), strong | -0.18 | 0.23 | 50 |
| uc_berkeley_haas | | | |
| Ideology (LDA 50), strong | -0.069 | 0.23 | 18 |
| Ideology (JEL 1), strong | -0.034 | 0.33 | 18 |
| uc_davis | | | |
| Ideology (LDA 50), strong | -0.22 | 0.24 | 18 |
| Ideology (JEL 1), strong | -0.23 | 0.27 | 18 |
| uc_san_diego | | | |
| Ideology (LDA 50), strong | -0.19 | 0.21 | 26 |

| | | | |
|---------------------------|--------|------|-----|
| Ideology (JEL 1), strong | -0.19 | 0.24 | 26 |
| ucla | | | |
| Ideology (LDA 50), strong | -0.20 | 0.20 | 27 |
| Ideology (JEL 1), strong | -0.18 | 0.23 | 27 |
| vanderbilt | | | |
| Ideology (LDA 50), strong | -0.049 | 0.22 | 23 |
| Ideology (JEL 1), strong | -0.032 | 0.39 | 23 |
| washington_u_st_louis | | | |
| Ideology (LDA 50), strong | -0.057 | 0.18 | 14 |
| Ideology (JEL 1), strong | -0.15 | 0.15 | 14 |
| yale | | | |
| Ideology (LDA 50), strong | -0.22 | 0.15 | 19 |
| Ideology (JEL 1), strong | -0.19 | 0.22 | 19 |
| yale_school_of_management | | | |
| Ideology (LDA 50), strong | -0.068 | . | 1 |
| Ideology (JEL 1), strong | 0.16 | . | 1 |
| Total | | | |
| Ideology (LDA 50), strong | -0.15 | 0.21 | 960 |
| Ideology (JEL 1), strong | -0.13 | 0.25 | 962 |
| Observations | 962 | | |

Table A.7: Correlation Between Author Ideology (jel1) and CV Characteristics

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|--------------------------|-----------------------|------------------------|----------------------|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Financial Economics | 0.199*** (0.0359) | 0.172*** (0.0369) | | | 0.164*** (0.0361) | 0.178*** (0.0354) | 0.173*** (0.0368) | 0.193*** (0.0727) |
| Macroeconomics | 0.167*** (0.0255) | 0.144*** (0.0274) | | | 0.160*** (0.0258) | 0.146*** (0.0258) | 0.163*** (0.0256) | 0.0874 (0.0579) |
| International Trade | 0.116*** (0.0285) | 0.123*** (0.0301) | | | 0.113*** (0.0284) | 0.119*** (0.0285) | 0.114*** (0.0280) | 0.139* (0.0739) |
| Labor Economics | -0.142*** (0.0258) | -0.164*** (0.0266) | | | -0.139*** (0.0255) | -0.149*** (0.0251) | -0.136*** (0.0255) | -0.203*** (0.0697) |
| Public Economics | -0.00920 (0.0300) | -0.00377 (0.0305) | | | -0.0164 (0.0298) | -0.00981 (0.0302) | -0.0109 (0.0294) | 0.0637 (0.0671) |
| Development | -0.0278 (0.0291) | -0.0195 (0.0293) | | | -0.0261 (0.0284) | -0.0242 (0.0288) | -0.0253 (0.0283) | -0.0271 (0.0652) |
| Econometrics | 0.0839** (0.0367) | 0.0808** (0.0365) | | | 0.0829** (0.0363) | 0.0756** (0.0363) | 0.0808** (0.0359) | -0.0115 (0.0941) |
| Microeconomic Theory | -0.0612** (0.0300) | -0.0822*** (0.0299) | | | -0.0653** (0.0299) | -0.0721** (0.0295) | -0.0678** (0.0297) | -0.116 (0.0797) |
| Business School | | | 0.133*** (0.0349) | 0.0365 (0.0338) | 0.0842*** (0.0262) | 0.0589** (0.0296) | 0.0856*** (0.0264) | 0.0611 (0.0649) |
| Saltwater Ever (Tervio) | | | 0.0210 (0.0577) | -0.0149 (0.0509) | -0.0244 (0.0400) | -0.00977 (0.0460) | -0.0199 (0.0400) | 0.136 (0.0933) |
| Freshwater Ever (Tervio) | | | 0.0939* (0.0517) | 0.0813* (0.0459) | 0.0700* (0.0394) | 0.0416 (0.0456) | 0.0697* (0.0386) | -0.0567 (0.107) |
| Latin American Origin | | | 0.173** (0.0693) | 0.0901 (0.0861) | 0.0721 (0.0840) | 0.0620 (0.0797) | 0.0680 (0.0830) | 0.116 (0.120) |
| European Origin | | | 0.133*** (0.0355) | 0.107*** (0.0310) | 0.0842*** (0.0289) | 0.0701** (0.0289) | 0.0858*** (0.0289) | 0.0788 (0.0949) |
| Full Professor | | | -0.0289 (0.0301) | 0.0192 (0.0293) | 0.0115 (0.0224) | 0.00196 (0.0252) | 0.0106 (0.0223) | 0.0317 (0.0505) |
| Top 5 Economics Dept | | | -0.0149 (0.0781) | 0.0323 (0.0653) | -0.0383 (0.0243) | -0.00861 (0.0642) | -0.0376 (0.0242) | -0.120** (0.0512) |
| Groundtruth Ideology | | | | | | | | 0.0746*** (0.0284) |
| Department FE | No | Yes | No | No | No | No | No | No |
| University FE | No | No | Yes | Yes | No | Yes | No | No |
| Field FE | No | No | No | Yes | No | No | No | No |
| PhD Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 895 | 895 | 895 | 895 | 895 | 895 | 895 | 262 |

Robust standard errors in parenthesis. Fields refer to self-reported fields on CVs.

Saltwater and Freshwater are constructed from Tervio and are described in the text.

Table A.8: Correlation Between Author Ideology (notopic) and CV Characteristics

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|--------------------------|-----------------------|-----------------------|-----------------------|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| Financial Economics | 0.335*** (0.0519) | 0.299*** (0.0540) | | | 0.279*** (0.0543) | 0.299*** (0.0522) | 0.293*** (0.0553) | 0.366*** (0.112) |
| Macroeconomics | 0.267*** (0.0381) | 0.240*** (0.0413) | | | 0.255*** (0.0378) | 0.230*** (0.0386) | 0.260*** (0.0377) | 0.214** (0.0874) |
| International Trade | 0.166*** (0.0449) | 0.159*** (0.0484) | | | 0.155*** (0.0451) | 0.162*** (0.0456) | 0.157*** (0.0445) | 0.146 (0.119) |
| Labor Economics | -0.314*** (0.0373) | -0.336*** (0.0383) | | | -0.305*** (0.0364) | -0.321*** (0.0367) | -0.300*** (0.0363) | -0.326*** (0.0996) |
| Public Economics | -0.0379 (0.0446) | -0.0335 (0.0450) | | | -0.0505 (0.0437) | -0.0405 (0.0440) | -0.0421 (0.0433) | 0.131 (0.0905) |
| Development | -0.0889* (0.0460) | -0.0830* (0.0468) | | | -0.0883* (0.0452) | -0.0887* (0.0462) | -0.0870* (0.0453) | -0.105 (0.101) |
| Econometrics | 0.151*** (0.0512) | 0.156*** (0.0508) | | | 0.151*** (0.0516) | 0.144*** (0.0513) | 0.148*** (0.0511) | 0.0468 (0.137) |
| Microeconomic Theory | -0.0223 (0.0459) | -0.0368 (0.0463) | | | -0.0298 (0.0457) | -0.0369 (0.0452) | -0.0335 (0.0455) | -0.0356 (0.126) |
| Business School | | | 0.222*** (0.0535) | 0.0619 (0.0516) | 0.124*** (0.0403) | 0.0935** (0.0448) | 0.126*** (0.0404) | 0.0703 (0.104) |
| Saltwater Ever (Tervio) | | | -0.0306 (0.0853) | -0.0772 (0.0735) | -0.0756 (0.0539) | -0.0940 (0.0620) | -0.0688 (0.0537) | 0.0788 (0.123) |
| Freshwater Ever (Tervio) | | | 0.165** (0.0750) | 0.0634 (0.0673) | 0.0994* (0.0550) | 0.0750 (0.0661) | 0.0989* (0.0537) | -0.0317 (0.160) |
| Latin American Origin | | | 0.407*** (0.102) | 0.253** (0.113) | 0.221** (0.112) | 0.206* (0.108) | 0.214* (0.110) | 0.398*** (0.151) |
| European Origin | | | 0.205*** (0.0548) | 0.174*** (0.0500) | 0.110** (0.0449) | 0.0896** (0.0437) | 0.113** (0.0447) | -0.00784 (0.142) |
| Full Professor | | | -0.0927** (0.0471) | -0.0158 (0.0440) | -0.0364 (0.0329) | -0.0399 (0.0369) | -0.0377 (0.0329) | -0.0968 (0.0761) |
| Top 5 Economics Dept | | | 0.0408 (0.124) | 0.0892 (0.100) | -0.0273 (0.0376) | 0.0438 (0.0983) | -0.0263 (0.0373) | -0.140* (0.0731) |
| Groundtruth Ideology | | | | | | | | 0.135*** (0.0429) |
| Department FE | No | Yes | No | No | No | No | No | No |
| University FE | No | No | Yes | Yes | No | Yes | No | No |
| Field FE | No | No | No | Yes | No | No | No | No |
| PhD Cohort FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Observations | 895 | 895 | 895 | 895 | 895 | 895 | 895 | 262 |

Robust standard errors in parenthesis. Fields refer to self-reported fields on CVs.
Saltwater and Freshwater are constructed from Tervio and are described in the text.

Table A.9: Correlation Between Author Ideology (groundtruth) and CV Characteristics

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|----------------------|----------------------|---------------------|---------------------|---------------------|--------------------|----------------------|
| Financial Economics | 0.653*** (0.157) | 0.650*** (0.162) | 0.622*** (0.195) | 0.616*** (0.170) | 0.366* (0.209) | | |
| Labor Economics | -0.354** (0.146) | -0.360** (0.153) | -0.219 (0.154) | -0.340** (0.145) | -0.376** (0.170) | | |
| Macroeconomics | 0.286*** (0.0883) | 0.264** (0.118) | 0.237** (0.114) | 0.166* (0.0978) | 0.110 (0.127) | | |
| Public Economics | -0.0536 (0.131) | 0.0271 (0.126) | 0.103 (0.112) | -0.0380 (0.110) | -0.151 (0.166) | | |
| Business School | 0.138 (0.132) | 0.137 (0.152) | | -0.0476 (0.120) | 0.0613 (0.203) | -0.0870 (0.252) | -0.104 (0.260) |
| Saltwater Ever (Tervio) | -0.336 (0.209) | -0.318 (0.197) | -0.215 (0.207) | -0.318 (0.214) | -0.273 (0.218) | -0.406 (0.285) | -0.524* (0.285) |
| Freshwater Ever (Tervio) | 0.342** (0.152) | 0.331** (0.154) | 0.382 (0.280) | 0.301* (0.161) | 0.379** (0.166) | 0.520** (0.195) | 0.420** (0.177) |
| Doctoral Degree Year | 1.071 (0.730) | 0.885 (0.792) | 1.298* (0.699) | 1.202* (0.712) | | | |
| Years Between Undergrad and PhD Degrees | | -0.0173 (0.0190) | | | | | -0.0241 (0.0354) |
| Latin American Origin | | -1.228*** (0.178) | | | | | -1.729*** (0.574) |
| European Origin | | 0.455* (0.245) | | | | | 0.368 (0.433) |
| Full Professor | | 0.162 (0.0981) | | | | | 0.123 (0.188) |
| Department FE | No | No | Yes | No | No | No | No |
| University FE | No | No | No | Yes | Yes | Yes | Yes |
| Field FE | No | No | No | No | No | Yes | Yes |
| PhD Year FE | No | No | No | No | Yes | Yes | Yes |
| Observations | 257 | 246 | 257 | 257 | 279 | 279 | 246 |

Robust standard errors, clustered at the primary field level. Finance, Labor, Macro, and Public all refer to self-reported primary fields on CVs. Saltwater and Freshwater are constructed from Tervio and are described in the text.

A.1.1 Stability of Ideology Predictions

In this Appendix section we look at whether our ideology scores exhibit changes over the careers of economists. We proceed by forming two predictions of ideology: $Ideology_i^{Pre}$, from the first 50% of an economist i 's academic writing by words, and $Ideology_i^{Post}$ from the last 50%. We only show results for the $LDA50$ measure of ideology, as others are quite similar. Figure A.1 shows the scatterplot between $Ideology_i^{Pre}$ and $Ideology_i^{Post}$ for all the AEA economists in our sample. A.2 shows the scatterplot for the CV sample, with saltwater/freshwater, business school, Ph.D. completion year fixed effects, subfield fixed effects, rank, presence in groundtruth sample, region of origin, and years between undergrad and Ph.D. all partialled out. As discussed in the main text, the correlation is quite strong in both figures. Table A.10 shows that the correlation between early ideology and late ideology is robust to all of the controls mentioned above for all three of our predictions.

Figure A.3 shows the distribution of ideologies by young and old, and full professors and associate/assistant professors for each of our three predicted ideologies. The distributions are quite similar, suggesting that little narrowing or widening of views is taking place as economists get older or more professionally secure.

Figure A.1: Stability of Predicted Ideology for Full Sample

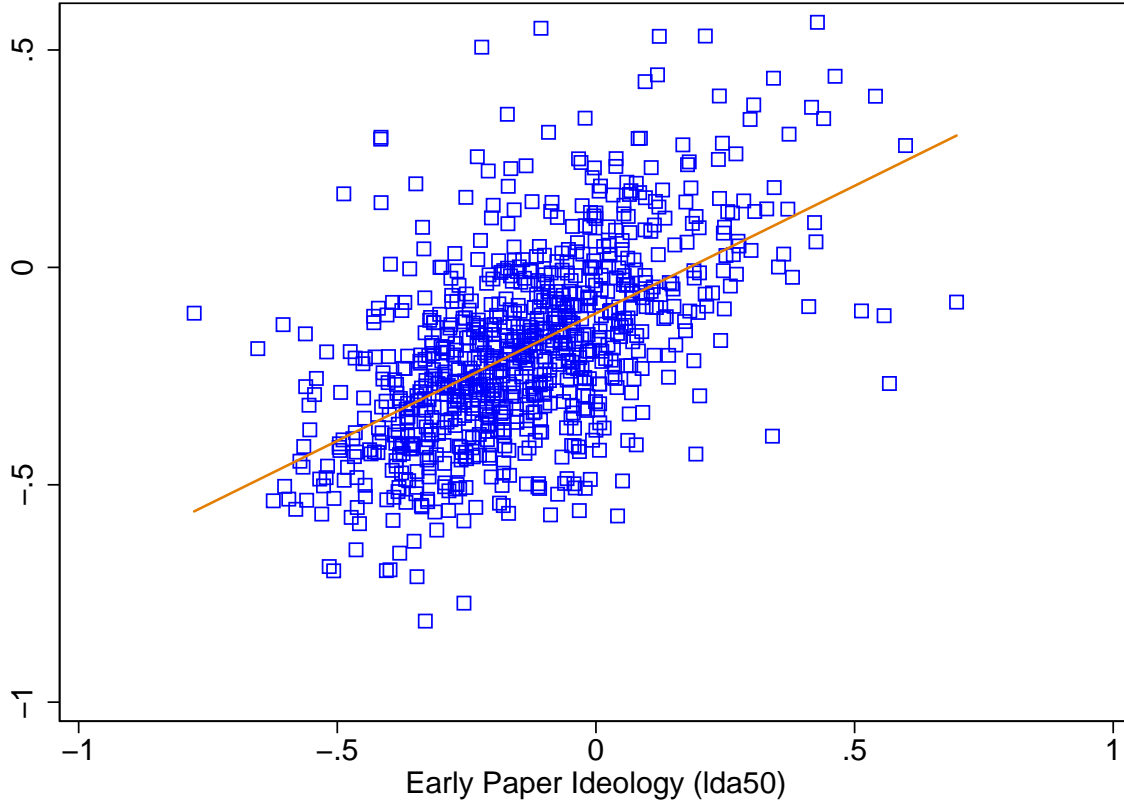
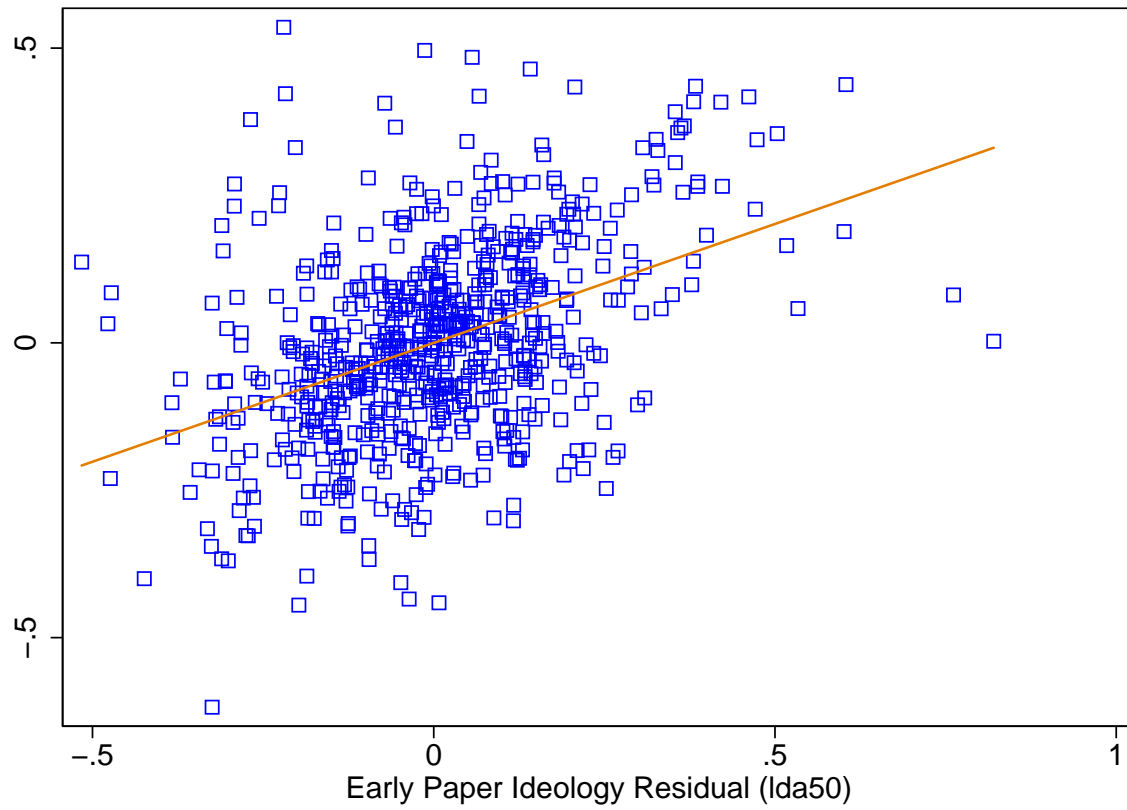


Table A.10: Correlation Between Late and Early Author Ideology

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|----------------------|----------------------|-----------------------|----------------------|----------------------|-----------------------|----------------------|----------------------|------------------------|
| Early Paper Ideology (lda50) | 0.586*** (0.0576) | 0.392*** (0.0540) | 0.403*** (0.0578) | | | | | | |
| Early Paper Ideology (jel1) | | | | 0.436*** (0.0547) | 0.282*** (0.0475) | 0.313*** (0.0391) | | | |
| Early Paper Ideology (notopic) | | | | | | | 0.723*** (0.0505) | 0.537*** (0.0323) | 0.544*** (0.0328) |
| Years Between Undergrad and PhD Degrees | | | 0.000373 (0.00402) | | | -0.00483 (0.00468) | | | -0.00838* (0.00489) |
| Latin American Origin | | | -0.00283 (0.106) | | | -0.0280 (0.105) | | | 0.0149 (0.132) |
| European Origin | | | 0.0480 (0.0351) | | | 0.0691** (0.0332) | | | 0.0701 (0.0466) |
| Full Professor | | | 0.000845 (0.0223) | | | -0.0357 (0.0286) | | | -0.00968 (0.0275) |
| Groundtruth Sample | | | -0.00650 (0.0155) | | | -0.0158 (0.0225) | | | -0.0244 (0.0213) |
| University FE | No | Yes | Yes | No | Yes | Yes | No | Yes | Yes |
| Field FE | No | Yes | Yes | No | Yes | Yes | No | Yes | Yes |
| PhD Year FE | No | Yes | Yes | No | Yes | Yes | No | Yes | Yes |
| Observations | 824 | 824 | 729 | 823 | 823 | 728 | 824 | 824 | 729 |

Robust standard errors, clustered at the primary field level.
 Saltwater and Freshwater are constructed from Tervio and are described in the text.

Figure A.2: Stability of Predicted Ideology Residuals for CV Sample



A.1.2 Journal Sample and Journal Ideology

Figure A.3: Distributions of Predicted Ideology by Age and Rank

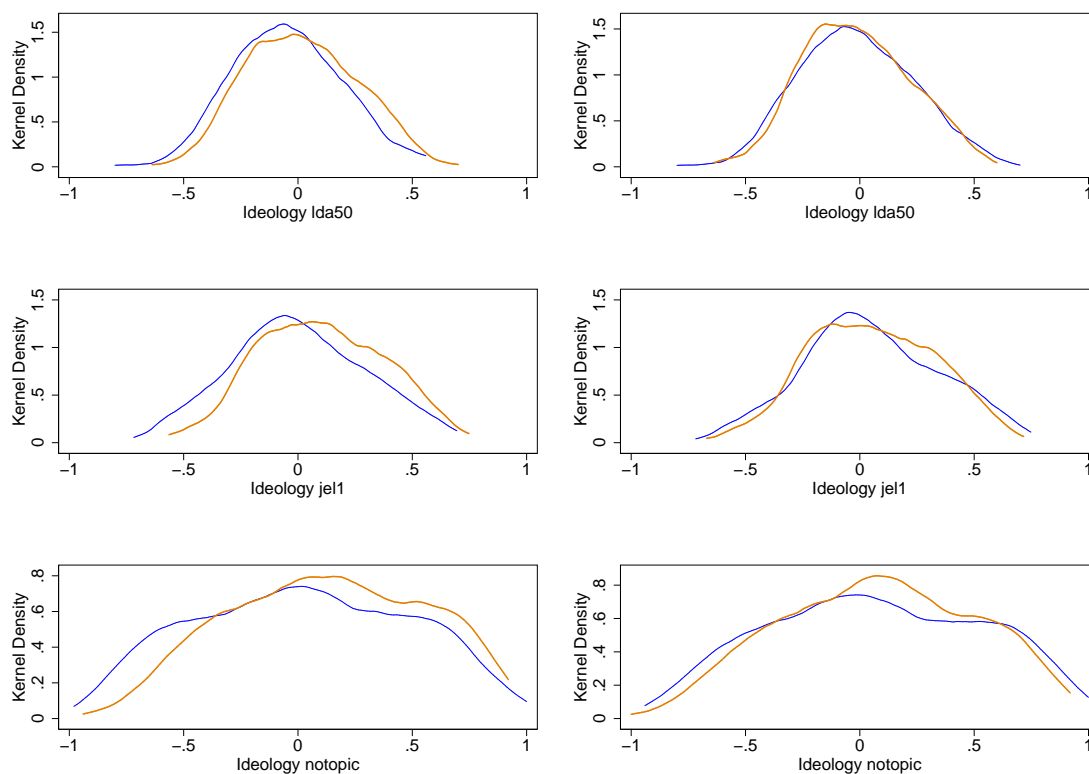


Table A.11: Journal Summary Statistics

| Journal | Count | LDA50 | JEL1 |
|--|-------|--------|--------|
| African Economic History | 104 | -0.038 | 0.021 |
| American Journal of Agricultural Economics | 2046 | 0.030 | -0.054 |
| Annales d'Economie et de Statistique | 596 | -0.345 | -0.169 |
| Annals of the American Academy of Political and Social Science | 1220 | -0.395 | -0.249 |
| Brookings-Wharton Papers on Urban Affairs | 42 | -0.212 | -0.154 |
| Brookings Papers on Economic Activity | 253 | -0.505 | -0.425 |
| Brookings Papers on Economic Activity. Microeconomics | 62 | -0.287 | -0.217 |
| Brookings Trade Forum | 81 | -0.331 | -0.296 |
| Canadian Journal of Political Science | 383 | -0.060 | 0.110 |
| Canadian Public Policy | 616 | -0.586 | -0.603 |
| Desarrollo Econ | 475 | -0.628 | -0.543 |
| Eastern Economic Journal | 544 | -0.055 | -0.070 |
| Eastern European Economics | 304 | -0.255 | -0.259 |
| Econom | 67 | -0.352 | -0.279 |
| Econometric Theory | 1045 | -0.073 | 0.110 |

Table A.11 (continued)

| | | | |
|--|-------|--------|--------|
| Econometrica | 1019 | -0.267 | -0.166 |
| Economic and Political Weekly | 15288 | -0.279 | -0.173 |
| Economic Development and Cultural Change | 503 | 0.103 | 0.275 |
| Economic Geography | 273 | -0.516 | -0.458 |
| Economic Policy | 170 | -0.145 | -0.159 |
| Economic Theory | 1188 | -0.181 | -0.250 |
| Economica | 452 | -0.102 | -0.101 |
| Emerging Markets Finance & Trade | 65 | -0.279 | -0.232 |
| Estudios Econ | 190 | -0.085 | -0.076 |
| Health Economics in Prevention and Care | 29 | -0.397 | -0.636 |
| IMF Staff Papers | 192 | -0.220 | -0.147 |
| Indian Journal of Industrial Relations | 449 | -0.301 | -0.263 |
| Industrial and Labor Relations Review | 599 | -0.246 | -0.195 |
| Innovation Policy and the Economy | 43 | -0.401 | -0.276 |
| International Economic Review | 729 | -0.189 | -0.181 |
| International Journal of Health Care Finance and Economics | 97 | -0.336 | -0.222 |
| Journal of Applied Econometrics | 511 | 0.043 | 0.037 |
| Journal of Business & Economic Statistics | 694 | 0.002 | 0.137 |
| Journal of Economic Growth | 163 | -0.158 | -0.137 |
| Journal of Economic Issues | 1043 | -0.546 | -0.554 |
| Journal of Economic Literature | 303 | -0.203 | -0.288 |
| Journal of Labor Economics | 446 | -0.202 | -0.037 |
| Journal of Law and Economics | 352 | 0.420 | 0.461 |
| Journal of Law, Economics, & Organization | 162 | 0.094 | 0.164 |
| Journal of Money, Credit and Banking | 874 | 0.281 | 0.407 |
| Journal of Population Economics | 471 | -0.259 | -0.144 |
| Journal of Post Keynesian Economics | 551 | -0.547 | -0.612 |
| Journal of the Economic and Social History of the Orient | 203 | -0.408 | 0.080 |
| Journal of the European Economic Association | 273 | -0.185 | -0.131 |
| Journal of Transport Economics and Policy | 332 | -0.096 | 0.004 |
| Land Economics | 659 | -0.092 | -0.208 |
| Marketing Letters | 444 | -0.144 | -0.049 |
| MIR: Management International Review | 256 | -0.236 | -0.116 |
| NBER International Seminar on Macroeconomics | 31 | -0.260 | -0.176 |
| NBER Macroeconomics Annual | 370 | -0.198 | -0.120 |
| New Labor Forum | 328 | -0.418 | -0.194 |
| Oxford Economic Papers | 296 | -0.168 | -0.118 |
| Public Choice | 1406 | 0.561 | 0.606 |
| Review (Fernand Braudel Center) | 250 | -0.384 | -0.344 |
| Review of African Political Economy | 894 | -0.275 | -0.253 |
| Review of Agricultural Economics | 545 | 0.132 | 0.184 |
| Review of International Political Economy | 343 | -0.327 | -0.065 |
| Review of World Economics | 128 | -0.306 | -0.213 |
| Revue | 1188 | -0.701 | -0.259 |
| Russian and East European Finance and Trade | 227 | -0.187 | -0.215 |
| Science & Society | 335 | -0.384 | -0.452 |
| Small Business Economics | 708 | -0.175 | -0.143 |
| Southern Economic Journal | 1111 | 0.095 | 0.102 |
| Soviet and Eastern European Foreign Trade | 19 | -0.104 | -0.442 |
| Staff Papers - International Monetary Fund | 256 | -0.078 | -0.031 |
| Supreme Court Economic Review | 61 | 0.200 | 0.238 |
| Tax Policy and the Economy | 94 | -0.281 | -0.290 |

Table A.11 (continued)

| | | | |
|---|------|--------|--------|
| The American Economic Review | 3031 | -0.167 | -0.155 |
| The American Economist | 320 | -0.020 | -0.022 |
| The American Journal of Economics and Sociology | 604 | 0.168 | 0.177 |
| The Brookings Review | 799 | -0.302 | -0.215 |
| The Business History Review | 198 | -0.348 | -0.303 |
| The Canadian Journal of Economics | 1065 | -0.178 | -0.112 |
| The Economic History Review | 415 | -0.175 | -0.224 |
| The Economic Journal | 1311 | -0.130 | -0.178 |
| The European Journal of Health Economics | 297 | -0.267 | -0.306 |
| The Journal of Developing Areas | 250 | -0.138 | -0.069 |
| The Journal of Economic Education | 520 | 0.090 | 0.011 |
| The Journal of Economic History | 617 | -0.089 | -0.113 |
| The Journal of Economic Perspectives | 921 | -0.230 | -0.202 |
| The Journal of Human Resources | 610 | -0.352 | -0.250 |
| The Journal of Industrial Economics | 379 | -0.116 | -0.144 |
| The Journal of Legal Studies | 353 | 0.132 | -0.082 |
| The Journal of Political Economy | 720 | 0.019 | 0.022 |
| The Journal of Risk and Insurance | 498 | 0.065 | -0.078 |
| The Quarterly Journal of Economics | 618 | -0.139 | -0.128 |
| The RAND Journal of Economics | 712 | -0.236 | -0.351 |
| The Review of Economic Studies | 643 | -0.246 | -0.203 |
| The Review of Economics and Statistics | 1108 | -0.045 | -0.011 |
| The Scandinavian Journal of Economics | 602 | -0.176 | -0.210 |
| The World Bank Economic Review | 259 | -0.297 | -0.233 |
| The World Bank Research Observer | 152 | -0.376 | -0.522 |
| Weltwirtschaftliches Archiv | 435 | -0.039 | 0.047 |

The list of 93 journals obtained from the JSTOR archive, the number of research articles in each journal, and the average JEL1 and LDA50 predicted ideologies across papers (the correlation between JEL1 and LDA50 is 0.862).

Figure A.4: The Number of NBER Working Papers by Year.

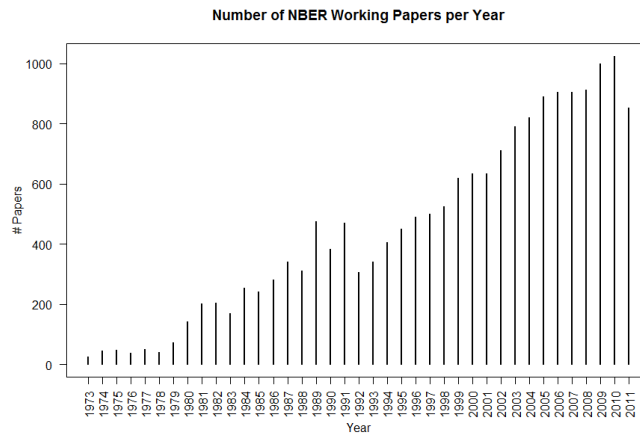


Figure A.5: The Number of Published Papers in JSTOR by Year.

