

What Have Economists Been Doing for the Last 50 Year?
A Text Analysis of Published Academic
Research from 1960-2010

Lea-Rachel Kosnik¹

UMSL Department of Economics Working Paper #1004

September 2014

Department of Economics
408 SSB
University of Missouri – St. Louis
1 University Blvd
St. Louis, MO 63121
<https://www.umsl.edu/econ/>

¹ University of Missouri-St. Louis, Department of Economics Email: kosnikl@umsl.edu

What Have Economists Been Doing for the Last 50 Years?
A Text Analysis of Published Academic Research from 1960-2010.

Lea Kosnik

Department of Economics
University of Missouri-St. Louis
St. Louis, MO 63121-4499
Tel: 1-314-516-5564
Fax: 1-314-516-5352
kosnikl@umsl.edu

Working Paper

Abstract

This paper presents the results of a text based exploratory study of over 20,000 academic articles published in seven top research journals from 1960-2010. The goal is to investigate the general research foci of economists over the last fifty years, how (if at all) they have changed over time, and what trends (if any) can be discerned from a broad body of the top academic research in the field. Of the 19 JEL-code based fields studied in the literature, most have retained a constant level of attention over the time period of this study, however, a notable exception is that of macroeconomics which has undergone a significantly diminishing level of research attention in the last couple of decades, across all the journals under study; at the same time, the “microfoundations” of macroeconomic papers appears to be increasing. Other results are also presented.

JEL Code: A11; B4

Keywords: text analysis, economics research, research diversity, topic analysis

Acknowledgements: Grateful acknowledgement is given for particularly insightful assistance from Harald Uhlig, Dan Hamermesh, David Laband, Adair Morse, Stefano DellaVigna, David Card, Bob Tollison, Allen Bellas, Ian Lange, Garth Heutel, John Whitehead, and various seminar participants at regional economic meetings, and the University of Missouri-St. Louis.

“The ideas of economists and political philosophers, both when they are right and when they are wrong are more powerful than is commonly understood. Indeed, the world is ruled by little else. Practical men, who believe themselves to be quite exempt from any intellectual influences, are usually slaves of some defunct economist.”

-- John Maynard Keynes

Introduction

If the economics profession holds as much sway as Keynes famously attributed to it, it would be worthwhile now and again to self-reflect and analyze what topics of research academic economists have spent their valuable time investigating. This paper presents the results of a text based exploratory study of over 20,000 academic articles published in seven top research journals from 1960-2010. The goal is to investigate the general research foci of economists over the last fifty years, how (if at all) they have changed over time, and what trends (if any) can be discerned from a broad body of the top academic research in the field. It is worth noting that no attempt is made, in this paper, to investigate the relative importance or quality of the research efforts of academic economists over this time span, only what topics they have in fact been studying.

Textual analysis (sometimes also called ‘content analysis’ or ‘computational linguistics’) involves the accumulation of a large amount of text (research articles, digitized books, online message boards, or twitter feeds, for example), cleaning and parsing the text with unique algorithms, and then turning the text into a database where the words themselves are statistically analyzed for trends and correlative patterns (Grimmer and King 2010; Michel et al. 2011; Evans and Foster 2011). Interesting social science examples of recent textual analyses include an investigation of culture from Top Ten song lyrics (DeWall et al. 2011), gender identification in

literary styles (Koppel et al. 2011), media slant in newspapers (Gentzkow and Shapiro 2010), and bargaining power in US-American Indian treaties (Spirling 2010).

This project utilizes all full-length monographs published in seven top journals in the field of economics from 1960-2010. The text is organized in a relational database, mapped to various characteristics of each article (author, year, journal, etc.) and the entire corpus, as well as cuts of the data by the specific characteristics, is analyzed. The main result is that, similar to results found in Card and DellaVigna (2013), the majority of fields in economics have maintained a relatively constant level of attention over the years. A major exception, however, is macroeconomics which, as in Kim et al. (2006) has shown a decreased level of attention over the past few decades, across all the major journals; at the same time, more refined analysis finds that the microfoundations of published macroeconomic papers have increased. Other interesting results include evidence for an increasing level of mathematization of economics over the decades, and for Agricultural and Natural Resource Economics to be an outlier as compared to the other fields, with significantly higher co-authorship versus solo authorship levels.

Literature Review

There is a rich history of self-reflection in the academic economics literature. Numerous authors have tried to analyze what researchers do, what topics they tend to focus on, and what (if any) practical impact economics research has had (Scroggs 1975; Granger 1994; Medema and Samuels 1996; Cropper 2000; Fuchs 2002; Pardey and Smith 2004; Sen 2007; Hamermesh 2013). A related, similarly navel-gazing theme in the economics literature involves studies of academic departments (Colander 1989), academic journals (Hawkins et al. 1973; Eagly 1975; Kagann and Leeson 1978; Laband et al. 2002; Card and DellaVigna 2013; Stern 2013), co-

authorship rates (Laband and Tollison 2000; Goyal et al. 2006; Hamermesh 2014), and other measures of intellectual collaboration and dissemination in the field that sometimes touch on topical analysis and what economists actually research (Kim et al. 2006).

These discussions, rankings, and lists have been around for decades, all offering differing views on the relevance of economics research and its trending topics. The ability to come up with robust, empirical-based conclusions from them, however, has been limited. Most of the evidence presented takes the form of simple counts of research articles, classified into broad categories determined by the researcher (Hamermesh 2013). Due to the laborious nature of this categorization process, where an individual has to read and categorize each and every article, such empirical evidence is often select and composed of relatively small sample sizes. Other forms of empirical evidence on the research trends in academic economics include counts of JEL codes (Card and DellaVigna 2013), citation based analysis (Kim et al. 2006) surveys of professionals in the field, and readings of Nobel prize acceptance speeches (Smith et al. 2004). None of this evidence is based on the text of the actual research itself; instead, it is all broad categorization and summarization. To date, there doesn't appear to have been any attempts to create quantifiable variables, for example on frequency of topical keywords over time, derived from something objectively calculated in the literature itself (from the full-length monographs, or even from just the article titles or abstracts). With the development of computational linguistic analysis, however, the possibility now exists to empirically summarize and test scores of research articles for topical themes in consistent, objective ways. One of the contributions of this paper, therefore, is its unique methodological take (i.e. text analysis) on a historically popular topic.

Some of the earliest research involving computer-aided¹ textual analysis was done in the fields of psychology (Sexton et al. 1999) and communications (Stephen 1999). Over the past decade it has grown to include interesting studies in other fields including political science (Gentzkow and Shapiro 2010; Spriling 2010), literature (Koppel et al. 2011), and even religious studies (Dershowitz et al. 2011). But the prevalence of textual analysis in the economics literature is slim (Kosnik 2014). There have been a few notable finance papers, such as on the ability of stock message boards to predict the Dow Jones Industrial Average (Antweiler and Frank 2004) and the effect of negative words in the *Wall Street Journal* on stock market returns (Tetlock 2007), but textual analysis in the economics literature is still in its infancy.

Those textual analyses that have been published in the social sciences literatures appear to take one of two forms: exploratory studies, or analytical investigations. Exploratory studies do not purport to prove specific hypotheses stated a priori, instead, they involve frequency and pattern analysis in order to objectively analyze a range of text in the hopes of uncovering intriguing results that may *then* lead to analytical investigations with appropriate hypotheses. For example, textual analysis of the top academic journals in the field of communications (Stephen 1999) was able to highlight which subtopics within the field received the most published attention, in which years, and from which specific journals.² A subsequent analytical investigation (Stephen 2000) of these research articles focused more specifically on the question of whether there was gender bias in published academic research in the communications field. An explanatory approach is taken with this project where we begin without any stated a prior

¹ Non-computer-aided text analysis has a much longer history. William Gladstone, for example, used it in the late 1800s to predict that the ancient Greeks were color blind (Dedrick 1998); he did this by tallying the color words found in works by Homer and noted that particular colors never appeared.

² Similar sorts of exploratory studies of the literature have been done in other fields, including psychology (Ellis et al. 1988), and health studies (Duncan 1991).

hypotheses, and proceed with frequency analysis towards a few tentative, perhaps suggestive conclusions from the data.³

Data

The data for this project constitutes 20,321 articles published in the following seven top-tier academic journals from 1960-2010: *American Economic Review*, *Econometrica*, *Journal of Economic Literature*, *Journal of Economic Perspectives*, *Journal of Political Economy*, *Quarterly Journal of Economics*, and *Review of Economic Studies*. The journal list was chosen after considering a number of different rankings, including Engemann and Wall (2009), Kalaitzidakis et al. (2001), and a variety of online listings. Previous research investigating trends in economics has also concentrated on the journals in this list (Laband and Tollison 2000; Laband et al. 2002; Card and DellaVigna 2013; Hammermesh 2013). The journals chosen are general interest journals and not field journals; a useful extension of this research in future years will be to extend the textual analysis to similar investigations and questions within subfields of economics.

We are aware that the *Journal of Economic Perspectives* (*JEP*) and the *Journal of Economic Literature* (*JEL*) are fundamentally distinct from the other five journals on the list. The articles in *JEP* and *JEL* are typically unrefereed and the topics and articles chosen are heavily editor influenced. We could have left *JEP* and *JEL* out of the analysis, but we left them in because they continuously rank highly in all available lists of academic journal rankings. The goal of this paper is to discern top foci of academic economists' attention – not necessarily its quality, importance, or optimality – and as such what these journals produce seems relevant. In

³ Future research efforts with this dataset are planned that will investigate hypotheses on research quality, journal impact, policy relevance, and other important issues that textual-based analysis should have the ability to fruitfully explore.

addition, because in the empirical section we break most of the results down by journal, leaving *JEP* and *JEL* in doesn't affect any of the disaggregated results.

All of the articles published in the seven journals studied, for the years 1960-2010, is in the database. The corpus includes everything research-oriented that has been published in English,⁴ including full-length monographs, full-length book reviews, and comments and replies. Entries not included in the dataset include editor's notes, conference announcements and programs, auditor's reports, and other similar non-research focused entries. Special symposium articles are included.⁵ Given these criteria the corpus includes 20,321 articles, some descriptive information for which can be found in Tables 1-3.

One of the criticisms of earlier research on publication trends in the economics literature is that the limited sample sizes they are usually based upon is so small; one benefit of this research is that this is not the case. This dataset is extensive enough that it can be fruitfully analyzed from a number of different angles, including by year, by journal, by monograph type,⁶ or by degree of co-authorship for example. It may be that interesting trends emerge from an analysis of the entire corpus, but it may also be that particular years or time periods also exhibit distinctive trends. A large originating dataset will be useful for analyzing specific cuts of the dataset. A second contribution of this paper, therefore, is the uniquely long time span comprehensively studied.

⁴ Some of these journals, especially in earlier years, included the occasional article in French or German.

⁵ It is worth noting, however, that the *American Economic Review's* annual *Papers and Proceedings* issue is not included.

⁶ Articles less than 5 pages in length are generally comments and replies, while articles greater than 5 pages in length are more often full-length research papers or book reviews.

Article, Abstract, or Title?

Previous research investigating academic trends from published research articles (Stephen 2000; Stephen 1999; Ellis 1988) didn't use the entire published monograph, but focused solely on the abstracts, or sometimes even just the titles. Doing so leads to a much smaller, more manageable text database, as well as faster computer processing times, but there is a fear that focusing solely on abstracts or titles might miss the larger picture of what a research article is about. In this analysis, therefore, the entire corpus of text is analyzed, as well as just the abstracts, and just the titles for comparison purposes.⁷ Each analysis has distinct advantages and disadvantages.

For example, analyzing the entire research articles may give a complete picture of concepts and foci covered, however, it gives shorter shrift to articles with a substantial amount of mathematical notation. Mathematical notation is simply skipped over and ignored by the algorithms utilized in this analysis, so if an article highlights a concept through mathematical notation, that emphasis is missed. For an article that instead gets all its points across in pure narrative, such a problem does not occur.

Focusing on an analysis of abstracts alone is one way around this problem. Abstracts tend not to have any mathematical notation in them whatsoever, and by definition they outline the important points of a research article. An analysis of abstracts alone may give a more balanced overview of research foci in economics. However, one drawback to focusing on abstracts alone is that some articles are not published with abstracts, particularly smaller articles,

⁷ All reported analyses followed standard text analysis techniques, including cleaning and parsing (where relevant) of the data, and the deletion of common, non-descriptive words such as "a," "the," "of," and "and."

replies, comments, and notes.⁸ It should be kept in mind, therefore, that the results from an analysis of abstracts alone is weighted towards the larger, more in-depth research articles.

Finally, an analysis of titles alone allows, in some sense, an equal weighting across all the articles in the dataset. All articles have a title, and all titles are roughly the same length. Some are notoriously short (for example the infamous “Elephants” article by Kremer and Morcom (2000)), but for the most part, analyzing titles alone is a way to dispense with mathematical notation as well as equally weight every observation in the dataset.

Figures 1-3 present frequency analyses on the database of text from the three methods of analysis described above (Articles, Abstracts, and Titles). Frequency distributions of most texts seem to follow a power law, whereby there is a long tail of words (to the right of the graph) that appear very few times, and a few words that dominate (the left side of the graph), and this appears to be true for these text databases as well.⁹ What this tells us is that all inquiries into the three text databases are dominated by key words, and that the majority of the words in any given body of text are actually used rather infrequently. This is helpful with regards to the textual analysis as it allows one to focus on a smaller body of words - the ones that occur with more frequency – in the analysis. In Figure 1 for example, on the entire corpus of text, there are really only about 1,000 words (out of around 16,000) that occur with a significant degree of repetition across the cases. In Figure 2, it is approximately 500 words, and in Figure 3, the Titles database, it is only about 250 words. Note that the most common word in the Articles corpus is “model” and it appears 439,646 times. The most common word in the Abstracts and Titles corpora is also “model,” appearing 10,806 and 1,586 times respectively. This is one indicator that whichever

⁸ In addition, abstracts tend to be used more today, than in the early years of the 1960s and 1970s. For longer research articles from those years that were published without abstracts, the first 1-2 paragraphs of the articles themselves were labeled and coded as if it were an abstract.

⁹ Frequency distributions on portions of each text database, for example text limited by year or by journal, all also exhibit distinct power law distributions.

way we analyze the research, through the entire body of the articles, the abstracts alone, or just the titles, we find some common results. Other comparisons were also computed as a check on the comparability of text analysis method, including the percentages of top keywords and key phrases¹⁰ in common and levels of keyword and key phrase case correlation (i.e. commonality across cases and not just in terms of total frequencies). There were a few notable differences, such as that Titles had a higher prevalence of “comment,” “note,” and “reply” in them, and that Articles had a higher prevalence of proper names in them, but otherwise many of the top frequencies were common across the method of analysis.

We feel comfortable, therefore, in the analysis which follows concentrating on the results from the Articles corpus. Everything is analyzed across the three corpora, but because there were few differences of note, for brevity’s sake the results displayed come from the Articles database alone.

Methodology & Results

We begin the analysis into research foci of published academic research by creating topic dictionaries whose lists of keywords and key phrases are considered representative of well-defined fields in economics.¹¹ These “bag-of-word” model dictionaries were created through a complete compilation of the disparate keyword lists assigned to field categories in the Journal of

¹⁰ Frequency analysis by key phrase (i.e. by “n-gram”) was done with $2 \leq n \leq 5$. For example, 2-grams are two keyword phrases such as “public good,” “interest rate,” “utility function,” or “monetary policy.” 3-grams are three keyword phrases such as “rate of return,” “real interest rate,” “necessary and sufficient,” “supply and demand,” “World War II,” “marginal tax rate,” and “maximum likelihood estimator.” 4-grams include “rates of time preference,” “marginal product of labor,” and “price elasticity of demand,” and sensical 5-grams include “pure theory of international trade,” “credit risk and credit rationing,” “public provision of private good,” and “general method of moment estimator.”

¹¹ Analysis of thematic content by topic dictionaries is a well-accepted practice in the field of text analysis (Weiss et al. 2005).

Economic Literature (JEL) Classification System.¹² The JEL system is composed of twenty distinct field categories, however one category (Y - Miscellaneous) contains no keywords so it was dropped from the analysis. The remaining 19 categories contained a total of 4,800 keywords/phrases, of which Table 4 gives a per category breakdown.¹³ The keywords themselves can be found at the JEL Classification System website.

These lists, or bag-of-word subject dictionaries, were applied to the text database to arrive at composite frequencies of use.¹⁴ Figure 4 shows category frequency of use over the entire 50 years of the dataset. It tells us that Microeconomics (D) is the most prevalent research category published in these journals, with Macroeconomics (E) and Labor (J) a more distant second and third, respectively. The dominance of Microeconomics is significant, with a 38% greater frequency of term use relative to Macroeconomics. Law and Economics (K), Special Topics (Z), Teaching (A), and History of Economic Thought (B) receive considerably less attention in the literature. Perhaps this is because they have their own specialized field journals, but so do Microeconomics, Macroeconomics, and Labor research categories, and yet they are still represented quite highly in these top general interest journals.

Figure 5 shows what happens to these category frequencies over time. For most of the 19 subject categories, their relative share of attention in the literature over the past five decades has not appreciably changed. This accords with similar results found in Card and DellaVigna (2013) and Kim et al. (2006), whereby the relative share of publications in specific disciplines has held steady over long time spans.¹⁵ However, there are a few exceptions, the most noteworthy of

¹² <http://www.aeaweb.org/econlit/jelCodes.php>. Lists downloaded as of June, 2014.

¹³ Note that the 4,800 keywords/phrases do contain some overlaps across the categories.

¹⁴ See Appendix for procedural detail.

¹⁵ Note that neither Card and DellaVigna (2013) – which base their results on JEL code analysis – nor Kim et al. (2006) – which base theirs on citation analysis – breaks down the categories of study exactly as we do. Card and DellaVigna (2013) identify 14 distinct subfields, while Kim et al. (2006) look only at 11.

which is Macroeconomics (E), which beginning in the early 1970s suffered a steady and appreciable decline in research attention.¹⁶ The decline for Macroeconomics (E) in Figure 5 is significant across the decades, at the 1% level. The economics profession has been roundly criticized in the media and lay literature for failing to foresee and predict the 2007 recession and associated financial crises; the fact that research attention in the field of Macroeconomics had appreciably declined in the years before these crises may be noteworthy.¹⁷

A bit of further exploration of this topic uncovers another interesting trend. Of the published articles containing macroeconomics content, a check on the simultaneous level of microeconomics content reveals that while macroeconomics research overall has been on the decline, macroeconomics papers with “microfoundation” content (as measured by JEL category D keyword analysis) appears to be on the rise. Figure 6 illustrates this. Detail on what types of microfoundational research on macroeconomics topics is being done is not explored here, and is left for a future research paper. For now, what is revealed in the text analysis is that interesting trends have occurred over the past five decades both in and within Macroeconomics research.

Journals:

Next we investigate research foci across the specific journals over time. Figure 7 shows graphs for each journal of all the 19 topic categories, but with Macroeconomics (E) again bolded for easy discernibility.¹⁸ This figure shows that the general decline in Macroeconomics published research is common across all the journals under study, and isn’t the result of one or

¹⁶ Kim et al. (2006) did find a similar effect for Macroeconomics, but they found a declining effect for Microeconomics too, which we did not.

¹⁷ Note that Financial Economics (G) also shows a statistically significant decline in research attention across three of the four decades (from the 1960s to the 1970s, the 1970s to the 1980s, and the 1990s to 2000s).

¹⁸ For Figures 7-11, the vertical axis continues to be “% Total Word Count.”

two specific journals greatly changing focus. For whatever reason, Macroeconomics has been losing publishing space to other fields across the top academic general interest journals.

Figure 8 explores the microfoundations of macroeconomics research, this time by journal. An interesting distinction emerges. The overall trend we found earlier (Figure 6) holds as well for all five of the refereed journals under study, but it doesn't for *JEP* and *JEL*. The heavily editor influenced, often non-refereed articles published in *JEP* and *JEL* do not show the same trend in this area, as the other journals do.

Moving on from Macroeconomics, there are too many categories (19) to repeat a figure similar to Figure 7 for each of the distinct categories, so instead we highlight just a few other interesting trends. Figure 9, for example, is a graph of each journal, this time of Mathematical Methods (C) alone so its trend can be easily discerned. The graphs illustrate an increasing level of mathematization of economics over the decades, for the majority of the journals under study. AER, for examples, sees a 56% increase in mathematical keyword and key phrase use from 1960 to 2010, JEL sees a 66% increase, JEP a 53% increase, JPE a 49% increase, QJE an 82% increase, and RES a whopping 200% increase. Only the journal *Econometrica* shows any decline in mathematization over the time period under study, and likely this is because it had a high level of mathematization to begin with, relative to the other journals. Comprehensively, significance tests over the decades show an increase in mathematization at the 1% level from the 1970s to the 1980s, and from the 1980s to the 1990s (there was also an increase from the 1960s to the 1970s, but it was not statistically significant). Whether this increase in mathematization is a good or bad development is left for another debate, but the published research record does appear to confirm the trend.

Taking advantage of the unique methodology utilized in this paper, Figure 10 illustrates some of the specific keywords in JEL category C that had the most gain (and loss) over the time span studied. From these results it appears that mathematical methods as applied to game theory led the gain in JEL category C's increasing research attention, while input-output models, IO, and linear programming applications experienced significant declines.

Figure 11 illustrates what has been happening in Microeconomics (D), the most prevalent research category in the set of research articles overall. The graphs confirm a trend that is common for most of the other nineteen categories, that its share of research in the top general interest journals has, for the most part, been constant; not just in total, but across the distinct journals as well. This begs the question as to *why* the specific categories of research are being given the consistent levels of attention that they are – is it a direct result of the types of research articles initially submitted for review and publication? Or, is it a result of editor tastes, tastes which seem to be consistent across the decades as well as across the journals? Where *is* this division of research attention coming from, and why?

Page Length:

Next we consider research type by page length. In our dataset there are articles as short as a single page (generally a note or comment), and longer monographs up to as many as ninety-nine pages. In looking at page length we make the implicit assumption that longer articles are more in-depth articles. In this section, therefore, we investigate whether certain subjects receive more in-depth (as proxied by a page length greater than five) research attention than others, and how this may have changed over time.

Figure 12 is a boxplot of the research categories by page length. For the most part, there is not much of a difference in category emphasis between shorter and longer research articles,

and indeed, the relative sizes of the boxes across research categories for both types of articles mimics the category influence of research attention overall (Figure 4).

There are a few categories where the research attention appears to be less in-depth. Teaching (A) and History of Economic Thought (B) have nearly 50% more frequency of term use in shorter articles than in longer articles.¹⁹ These are also two of the categories with the least research attention overall. At the same time, some of the categories with the most research attention (Microeconomics (D) and Labor (J), in particular) also have relatively more frequency of term use within in-depth articles. This seems to indicate that those categories with a dominance of shorter articles are also often the categories that receive less attention overall in the literature. The superstar research categories (such as Microeconomics and Labor) receive not just the most research attention, but also the most in-depth research attention.

Over time, the frequency of term use across shorter and longer research articles, per research category, shows no significant differences. Figure 13 provides the graphs for the research categories we focused on earlier (Macroeconomics (E), Mathematical Methods (C), and Microeconomics (D)). Notably, the attention paid to Macroeconomics declines nearly in tandem for both shorter and longer research articles. Mathematical Methods (C) and Microeconomics (D) also show no discernible differences in research attention between shorter and longer research articles.

The graphs for most of the rest of the research categories are similar to those found in Figure 12 in that there is no discernible difference in frequency of term use between shorter and longer articles. Two exceptions, however, are for Teaching (A) and History of Economic Thought (B). Figure 14 shows the graphs for these categories over time, and they display a distinctly unique trend where the level of shorter articles steadily declines, while the frequency

¹⁹ The differences between shorter and longer articles for the other research categories averages 11%.

of term use in longer articles increases. They both cross in the early 1980s. These two categories (A and B), alone among all the research categories, appear to be undergoing changes in research attention such that in the last few decades they are getting more attention in in-depth articles, and notably less in shorter comment and notice articles.

Number of Authors:

Finally, we consider research category by co-authorship level. The purpose is to investigate whether groups of authors investigate different topics significantly more or less than solo authors. Are there research topics that tend to lend themselves more to co-authorship? Perhaps as a result of social networking effects in particular fields?

The majority of articles in the dataset are solo authored (62%), but we do see co-authorship levels with as many as ten co-authors. In Figure 15 we group all papers written by two or more people as co-authored and graph a comparison of research category focus by solo and co-authorship levels. The results, as with Figure 12, do not show significant differences between the categories, and as well mimic overall research focus rates as illustrated in Figure 4. While there are many more categories that are dominated by solo-authored articles over co-authored ones (12 out of the 19 categories), that is likely a reflection of the fact that solo authored articles simply dominate the dataset. Overall, the conclusion appears to be that solo authors and co-authors appear to investigate economics research topics at similar levels; there is no dominance for co-authorship in particular fields.

We do note, however, that there is one rather unique outlier: Agricultural and Natural Resource Economics (Q), which has a rate of co-authorship 50% higher than solo authorship. The reason for this anomaly is not immediately obvious. Agricultural Economics has been a thriving field for a long time, but Natural Resource Economics is a relatively young subfield,

with seminal papers having been written as recently as the 1970s. The dominance of co-authorship levels in this category may be a result of the fact that more papers have been written in it later in the time span under study, and in recent decades co-authorship levels overall have risen (Hamermesh 2014).

When we look at frequencies of term use in solo authored and co-authored articles over time, for nearly all of the research categories under study term usage has remained relatively stable in solo authored articles, however, it has tended to increase in co-authored articles. Figure 16 provides a flavor, with graphs of the three particular research categories we have been following throughout this paper: Macroeconomics (E), Mathematical Methods (C), and Microeconomics (D). The bottom two graphs (Mathematical Methods and Microeconomics) show a relatively stable level of term use in the solo authored articles, but an increasing rate of term use in co-authored articles; it appears as though co-authored articles have increased their density of key jargon over time. Macroeconomics (E), ever the outlier, does not show any similar levels of increasing frequency, and instead displays a relatively constant (if volatile) level of term use across authorship type.

Conclusions

This research provides a review of the academic economics literature over the fifty years from 1960-2010. Articles published in this time span from seven top journals in the field were computationally analyzed with standard text analysis techniques for frequency patterns and thematic trends. The broadly optimistic goal of this research was to advance our knowledge and understanding of the economics profession by shedding light on what economists have been focusing on over the last number of decades.

A few conclusions can be drawn. First, that Microeconomics dominates the research attention in the economics profession, and by a significant margin. The next most researched field is Labor Economics, and after that Macroeconomics. The rest of the 19 fields studied receive less attention than these three.

A second conclusion is that, while Macroeconomics may be one of the top three most researched fields in economics from 1960-2010, its share of research attention has been steadily declining. The height of Macroeconomics research in the top general interest journals in academia was in the late 1960s, early 1970s; since then, the amount of published research attention devoted to the subject of Macroeconomics has been on a steady decline. After the recent 2007 recession and economic meltdown, the economics profession received a lot of criticism for not investigating some of the macroeconomic trends that we now know led to the crises; this may be a result of a declining lack of interest in the field by researchers.

At the same time, the level of research attention paid to Mathematical Methods in academic research has increased. It has long been whispered that the economics profession has become increasingly “mathematized” since the 1960s; the text analysis presented here provides empirical evidence for this trend.

All of the other research categories studied in this paper have maintained relatively stable levels of research attention over the years. This holds true not just across time, but across the individual journals studied, across an investigation into shorter and longer monograph types, and across solo versus co-authorship levels. It begs the question as to why? Why has there been such a steady division of research attention across the specific fields, over time and across journals? Are economists really such a staid bunch that our broad general interests never seem to change (even if, within these broader fields, newer developments may have occurred)? Is this

support for the stability of our preferences? Even if individual researchers' broad preferences have not changed much over time, why haven't the journals developed more of a comparative advantage by individualizing themselves with respect to research focus? What does it say about the competitive structure of the journal publishing field that none of these journals appears to have developed a comparative advantage in specific research foci? Or is it that the journals have developed specializations, just in something other than research foci (maybe publishing times)? A worthwhile future research agenda would be to investigate not just the levels of research attention in particular journals, but the optimality of such divisions. Why *do* we study what we study, and should that change?

Another important area for future research attention would be to investigate not just the amount of research attention dedicated to distinct fields, but also the relative importance of that attention. Extending the current dataset by coupling it with citation information, or journal impact factors over time, would allow an investigation not just into levels of research attention, but into levels of impact as well.

Table 1 - Article Counts per Decade

Journal	1960s	1970s	1980s	1990s	2000s	Totals
American Economic Review	693	1,184	1,194	888	1,092	5,051
Econometrica	896	1,105	841	570	671	4,083
Journal of Economic Literature	14	180	142	201	216	753
Journal of Economic Perspectives	0	0	144	615	620	1,379
Journal of Political Economy	1,271	1,181	814	563	460	4,289
Quarterly Journal of Economics	496	539	564	462	457	2,518
Review of Economic Studies	315	534	525	394	480	2,248
Totals	3,685	4,723	4,224	3,693	3,996	20,321

Table 2 - Article Counts by Page Length

Journal	0-5	5+	Totals
American Economic Review	1,179	3,872	5,051
Econometrica	1,037	3,046	4,083
Journal of Economic Literature	110	643	753
Journal of Economic Perspectives	161	1,218	1,379
Journal of Political Economy	1,447	2,842	4,289
Quarterly Journal of Economics	316	2,202	2,518
Review of Economic Studies	257	1,991	2,248
Totals	4,507	15,814	20,321

Table 3 – Article Counts by Numbers of Authors

Journal	1	2	3	4	5+	Totals
American Economic Review	2,778	1,788	415	60	10	5,051
Econometrica	2,551	1,218	270	42	2	4,083
Journal of Economic Literature	557	152	34	7	3	753
Journal of Economic Perspectives	885	365	94	16	19	1,379
Journal of Political Economy	3,053	1,003	205	26	2	4,289
Quarterly Journal of Economics	1,465	782	223	35	13	2,518
Review of Economic Studies	1,328	735	161	21	3	2,248
Totals	12,617	6,043	1,402	207	52	20,321

Figure 1 - Article Frequency Distribution

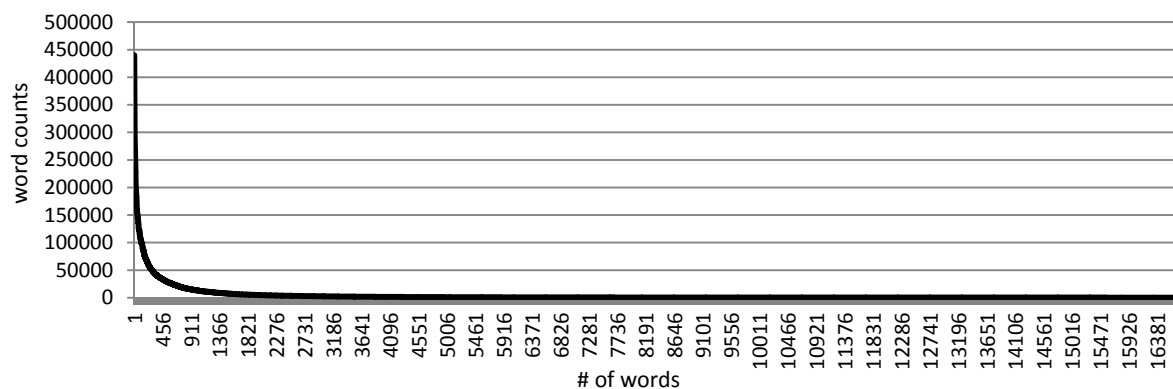


Figure 2 - Abstract Frequency Distribution

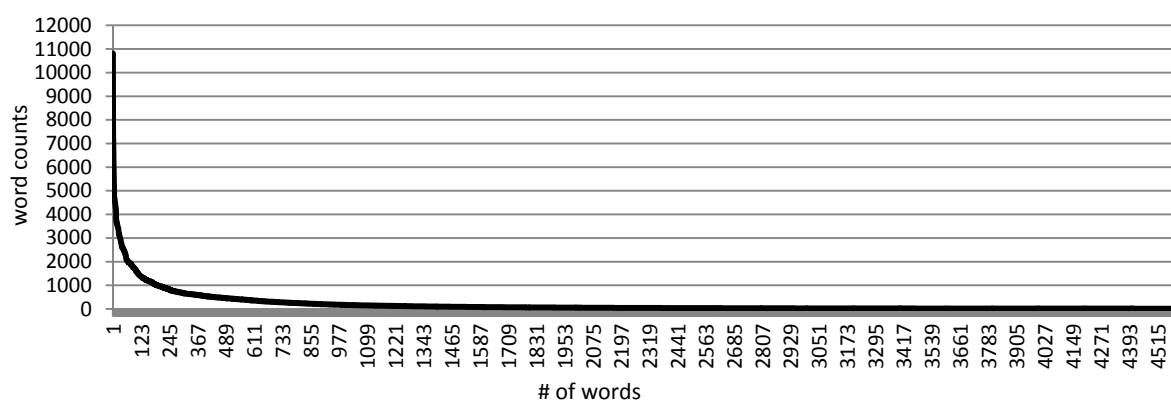


Figure 3 - Title Frequency Distribution

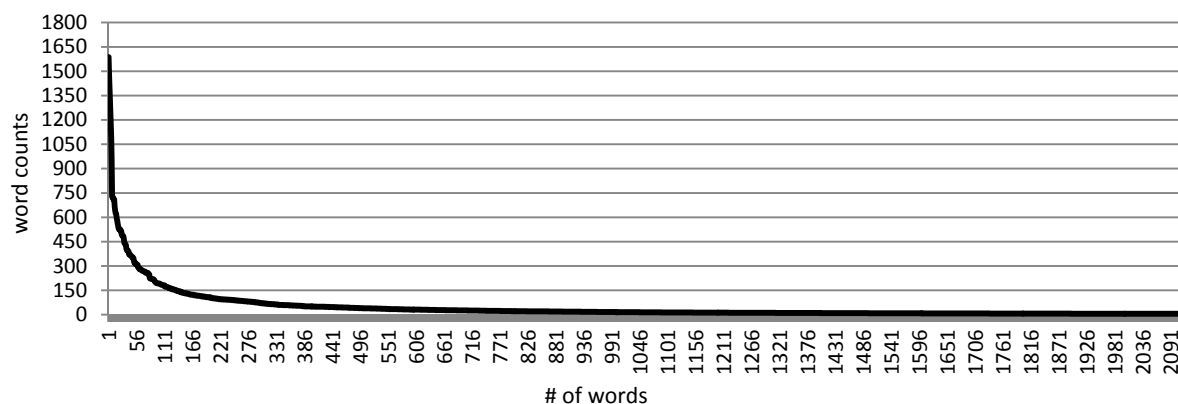


Table 4 – JEL Classification Categories and Keyword Counts

	Category	Counts
A	General Economics and Teaching	49
B	History of Economic Thought, Methodology, and Heterodox Approaches	107
C	Mathematical and Quantitative Methods	346
D	Microeconomics	504
E	Macroeconomics and Monetary Economics	368
F	International Economics	344
G	Financial Economics	264
H	Public Economics	222
I	Health, Education, and Welfare	147
J	Labor and Demographic Economics	459
K	Law and Economics	137
L	Industrial Organization	540
M	Business Administration and Business Economics, Marketing, and Accounting	142
N	Economic History	171
O	Economic Development, Technological Change, and Growth	349
P	Economic Systems	150
Q	Agricultural and Natural Resource Economics, Environmental and Ecological Economics	305
R	Urban, Rural, Regional, Real Estate, and Transportation Economics	169
Z	Other Special Topics	27

Source: <http://www.aeaweb.org/econlit/jelCodes.php>
 Keyword counts as of June, 2014.

Figure 4 – JEL Categories Frequency of Use, 1960-2010

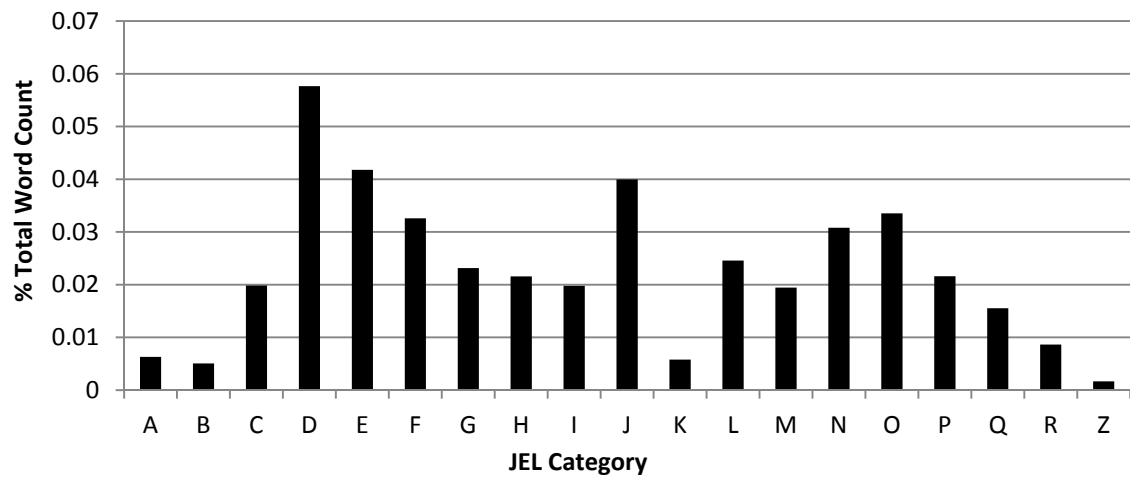


Figure 5 – JEL Categories Frequency of Use, per Year

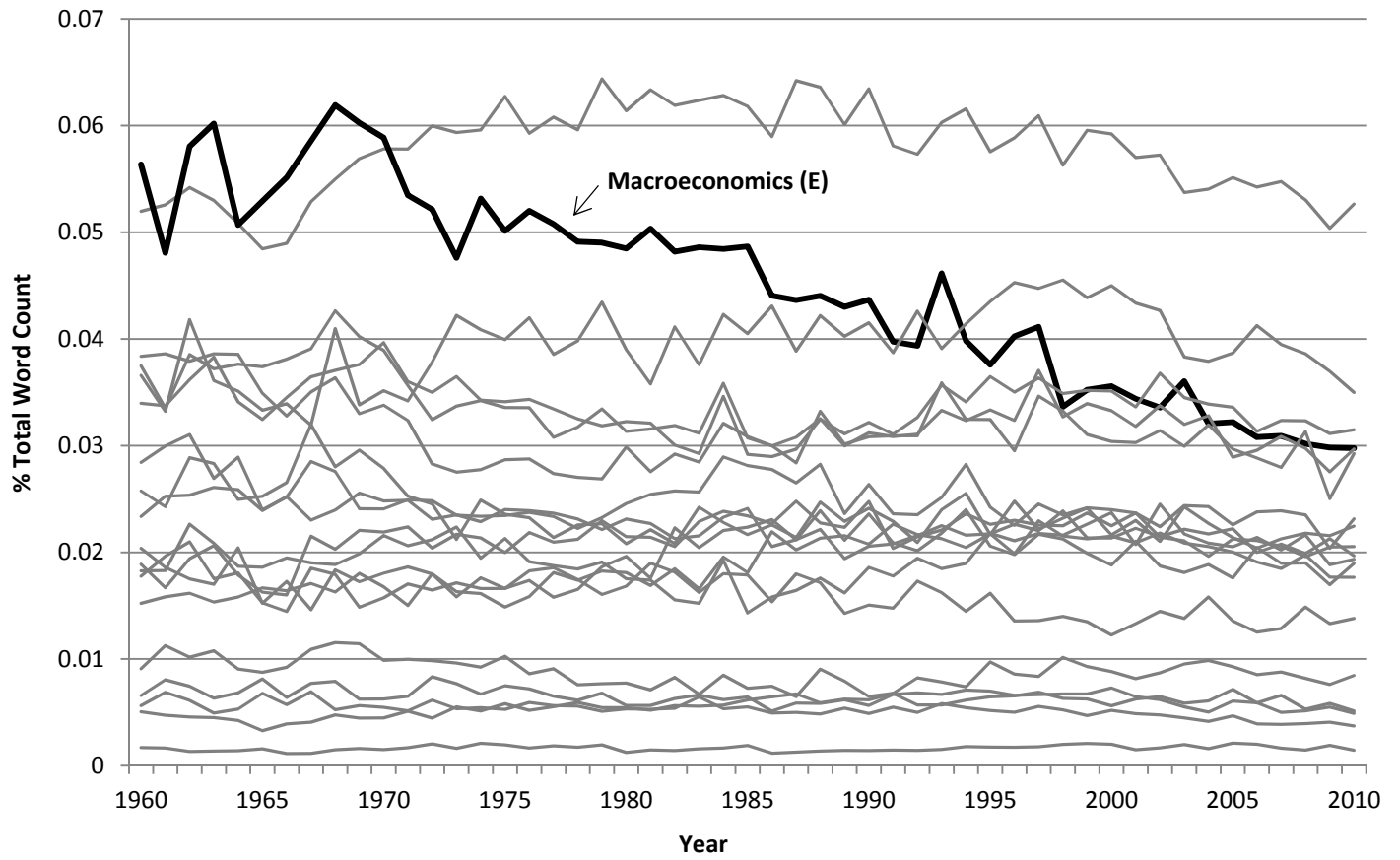


Figure 6 – The Microfoundations of Macroeconomics Research, Over Time

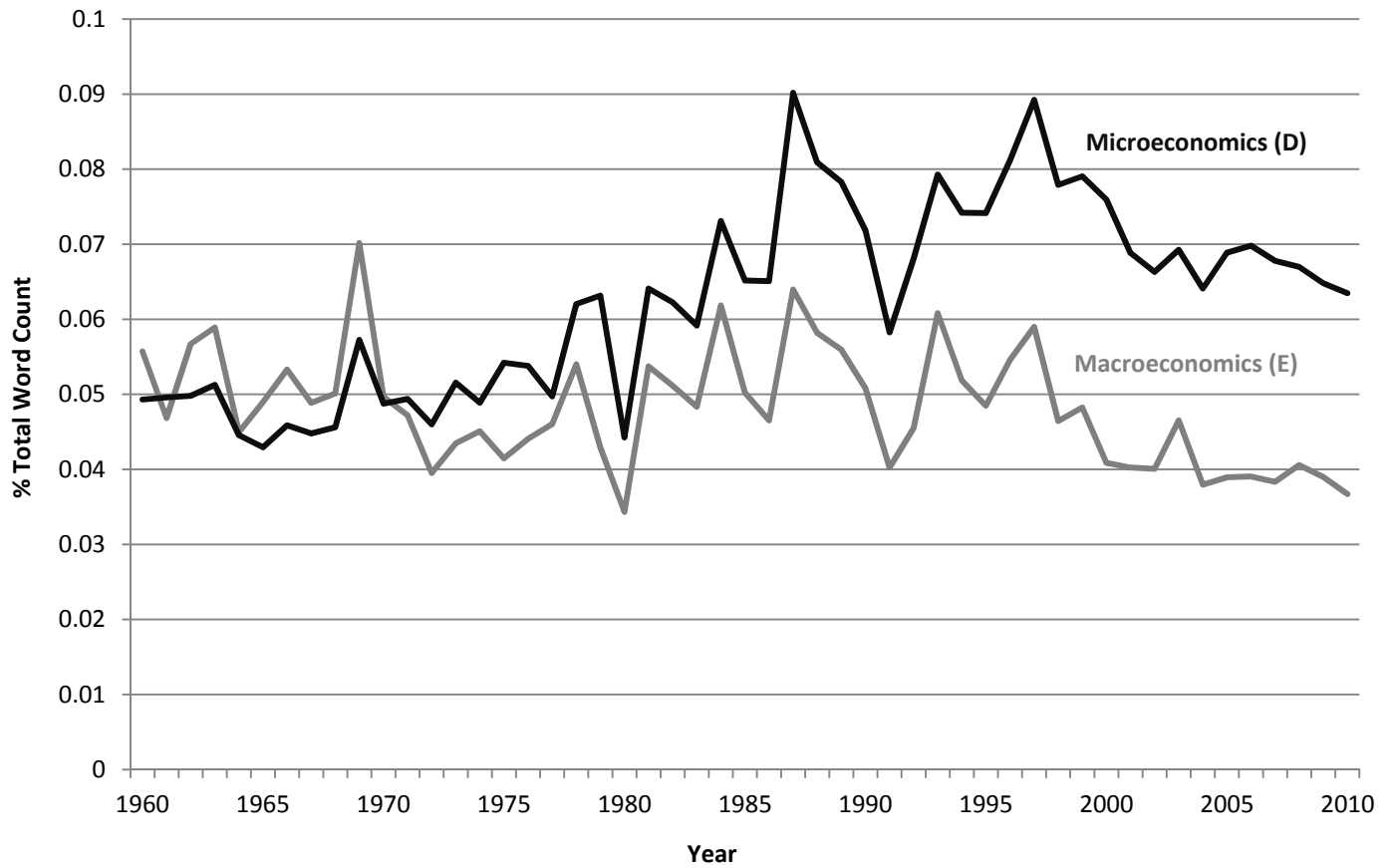
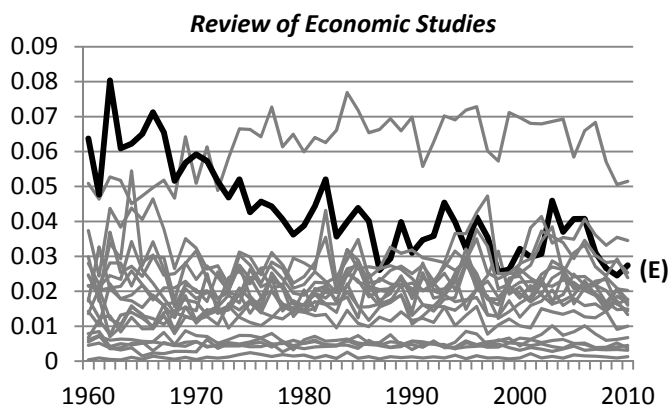
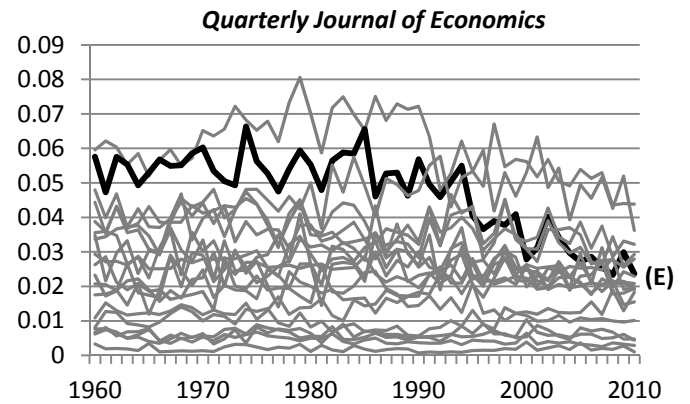
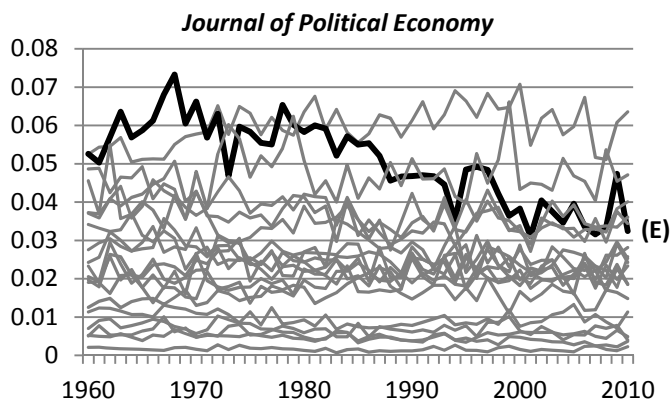
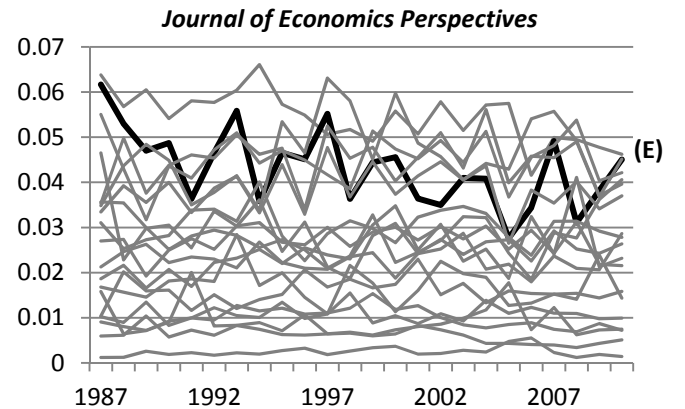
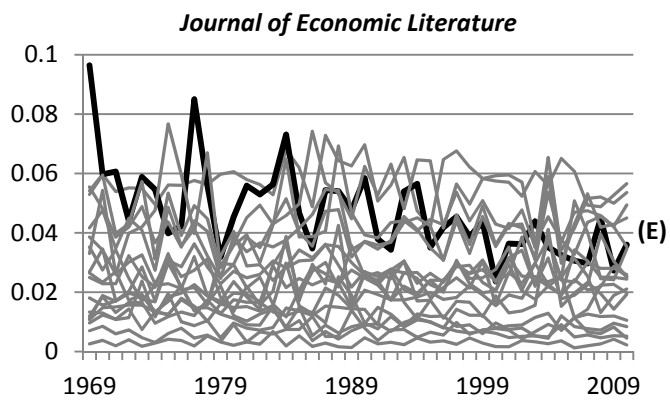
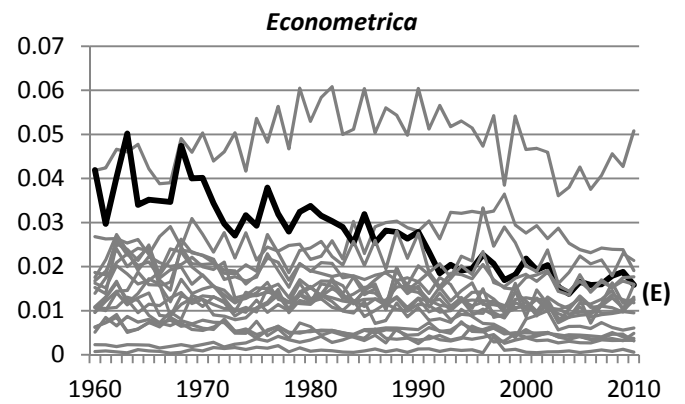
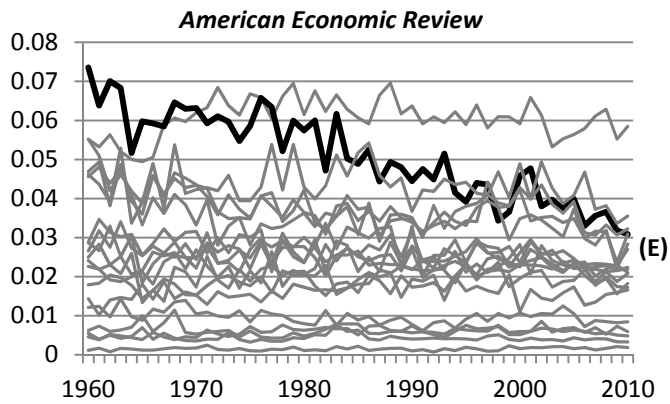
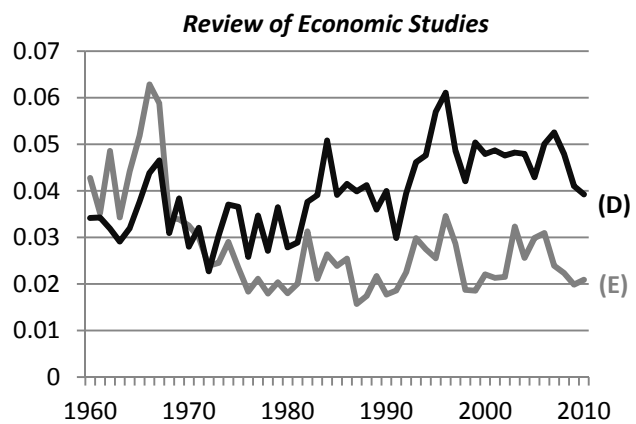
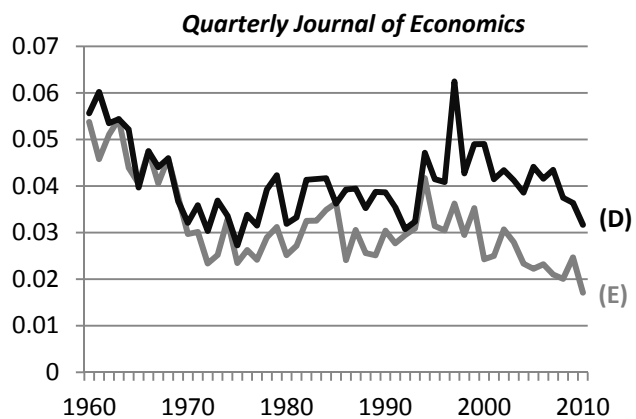
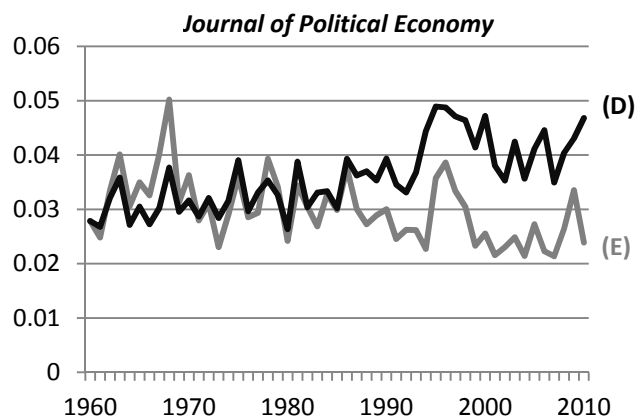
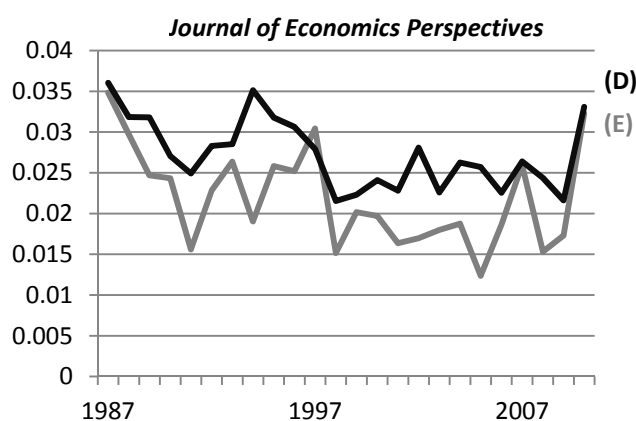
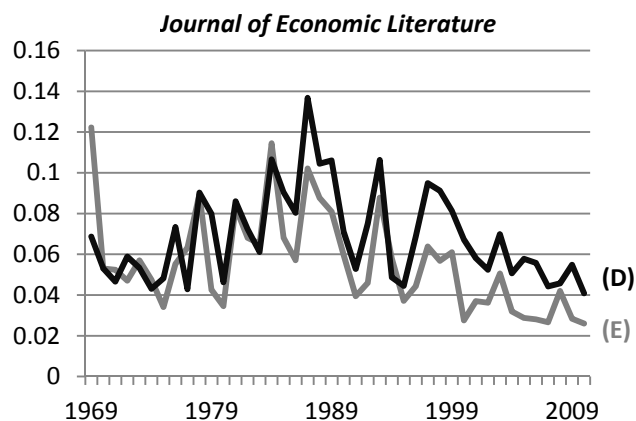
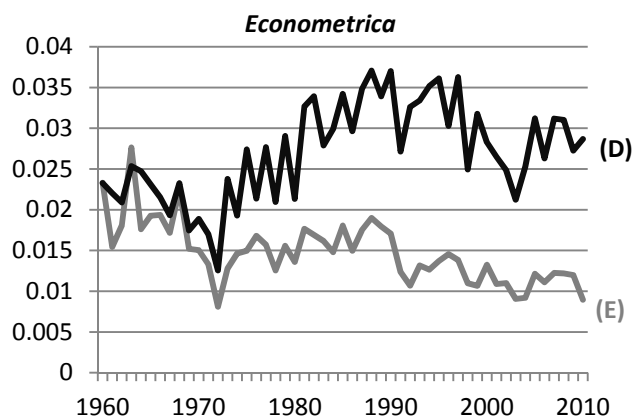
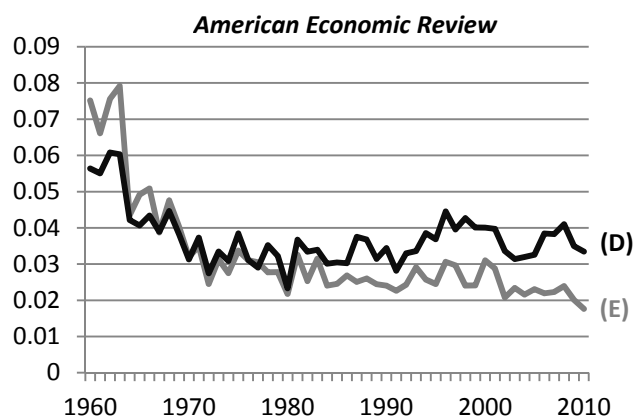


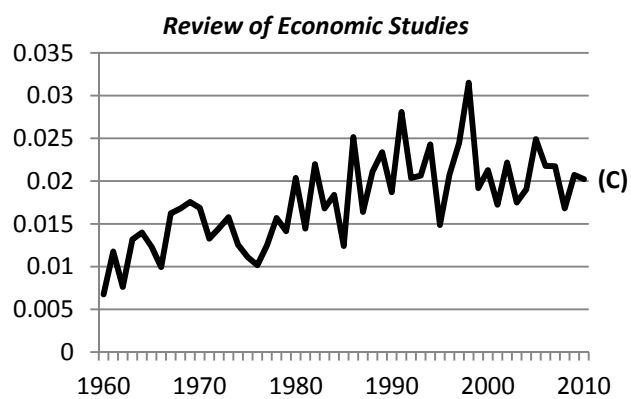
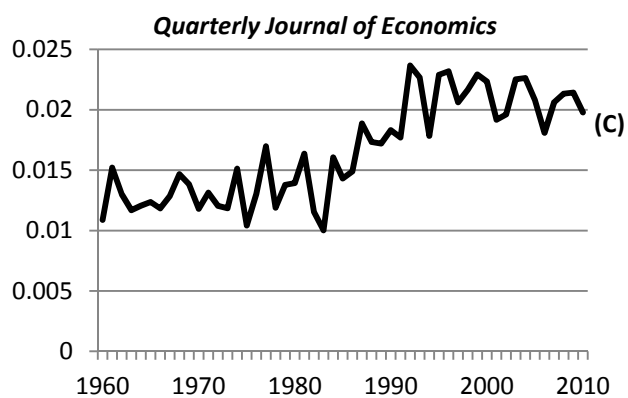
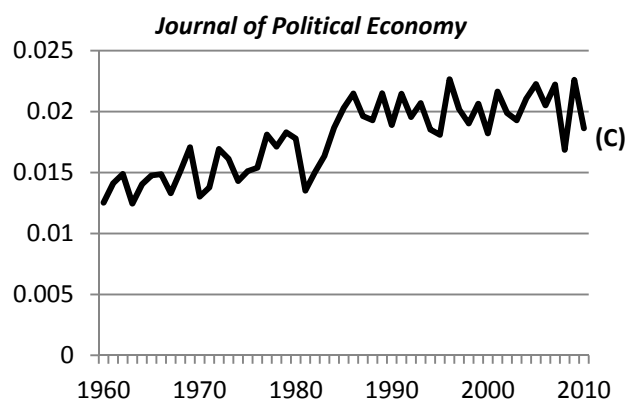
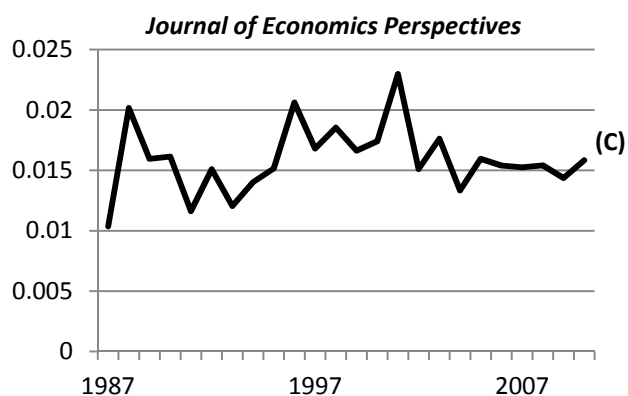
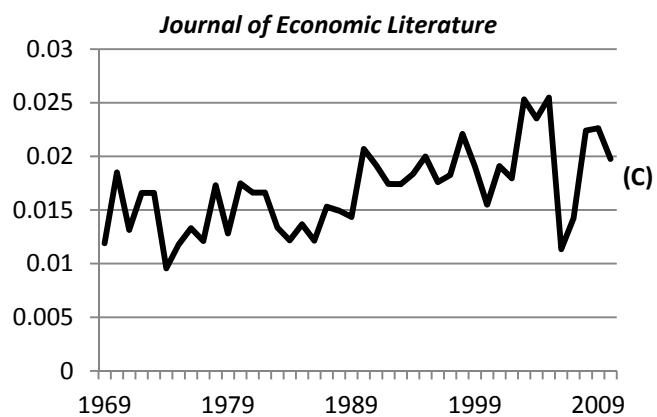
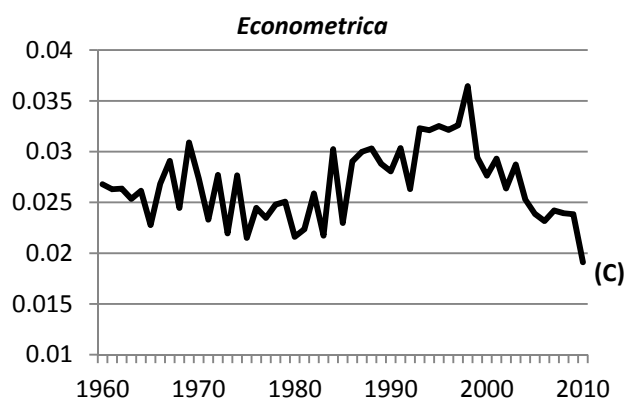
Figure 7 – JEL Categories Frequency of Use, per Journal, per Year
Macroeconomics (E) Highlighted



**Figure 8 – The Microfoundations of Macroeconomics Research,
per Journal, per Year**

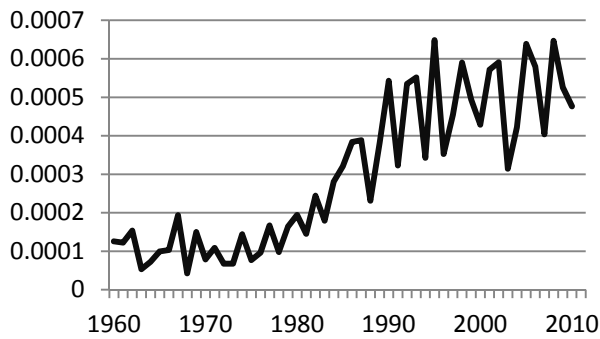


**Figure 9 – JEL Categories Frequency of Use, per Journal, per Year
Mathematical Methods (C) Highlighted**

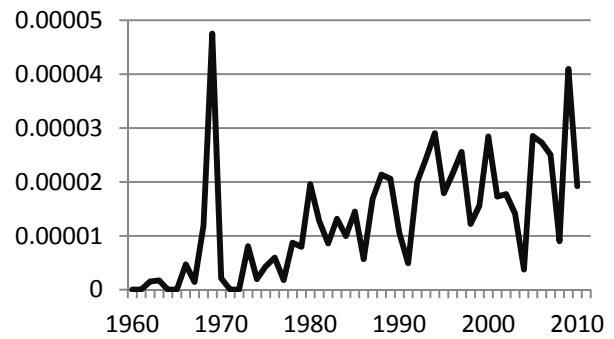


**Figure 10 – Mathematical Methods (C),
Change in Specific Keywords**

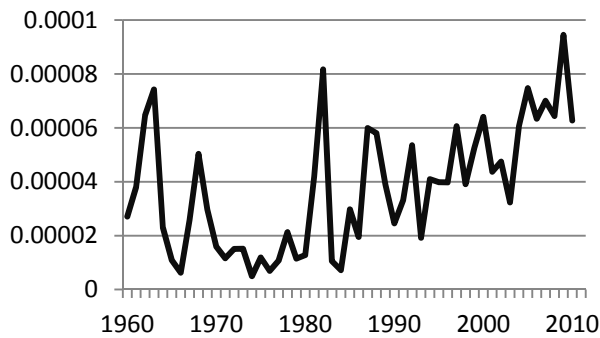
"GAMES"



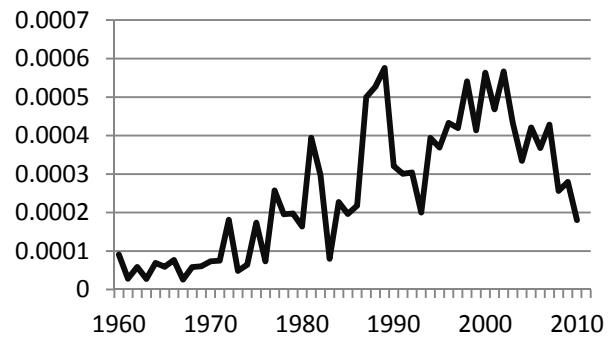
"PRISONER'S_DILEMMA"



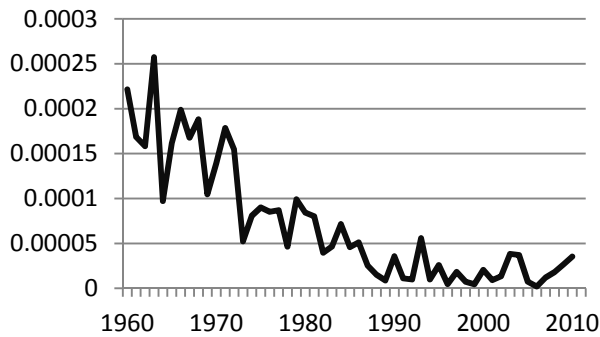
"LABORATORY"



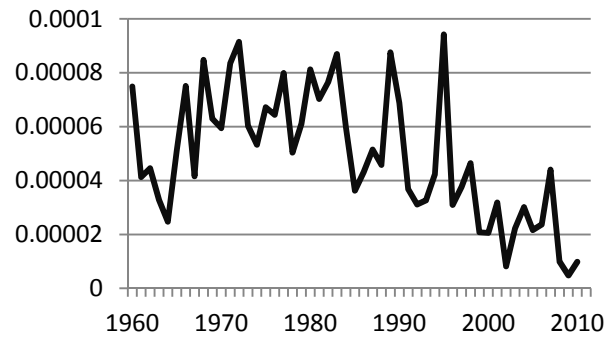
"BID"



"INPUT_OUTPUT"



"IO"



"LINEAR_PROGRAMMING"

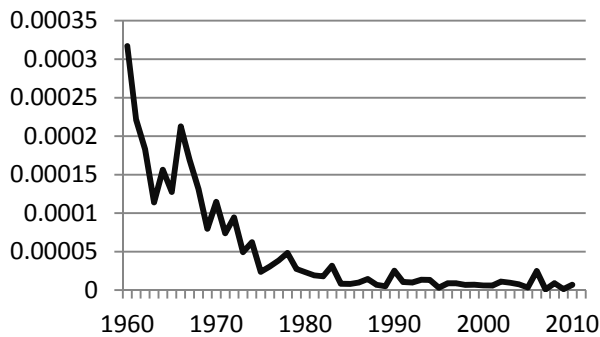


Figure 11 – JEL Categories Frequency of Use, per Journal, per Year
Microeconomics (D) Highlighted

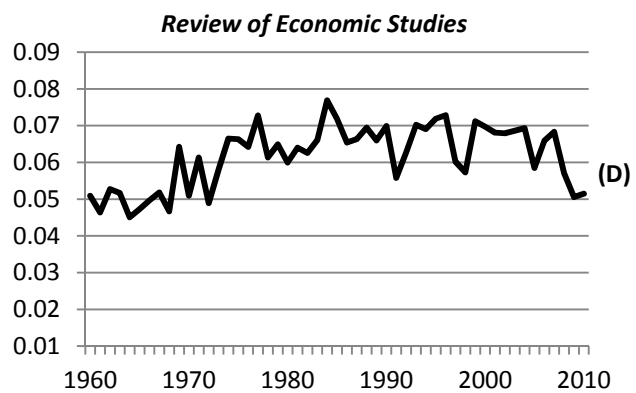
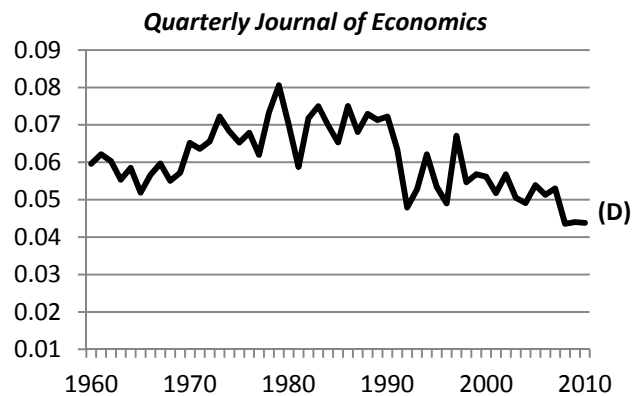
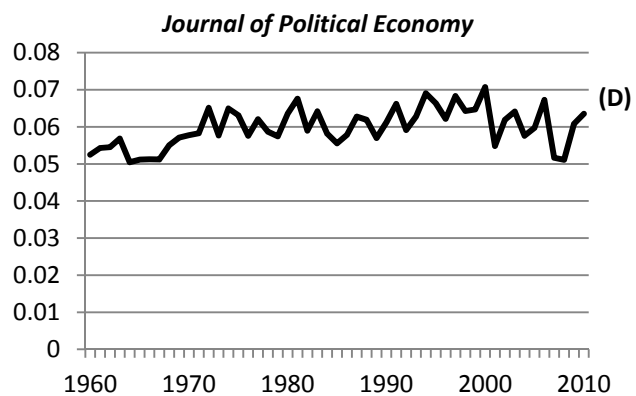
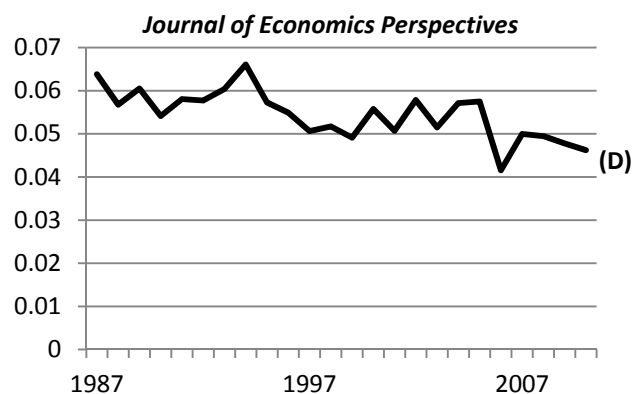
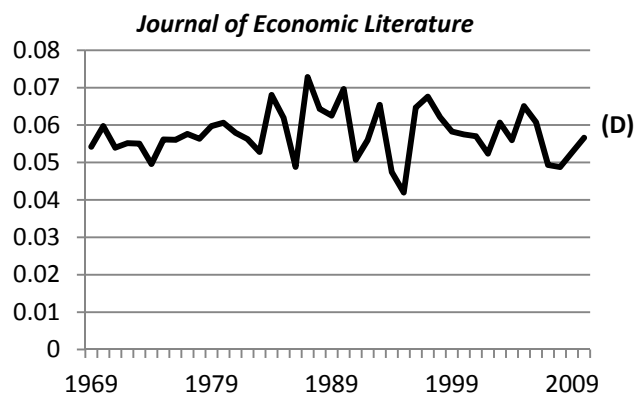
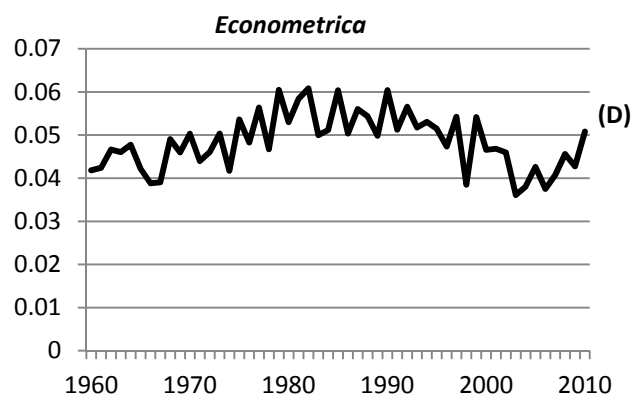


Figure 12 – JEL Categories Frequency of Use, by Page Length

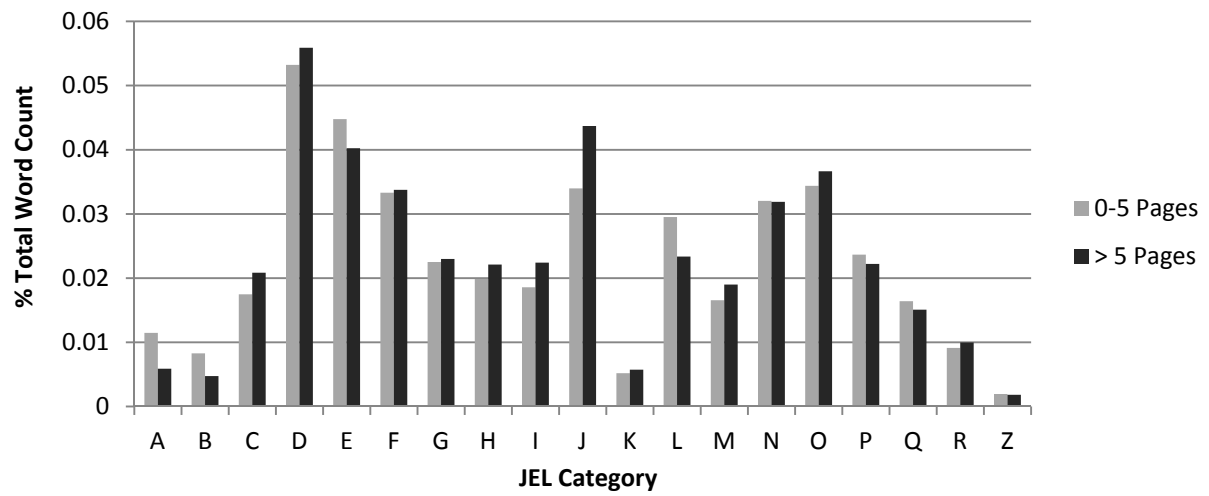
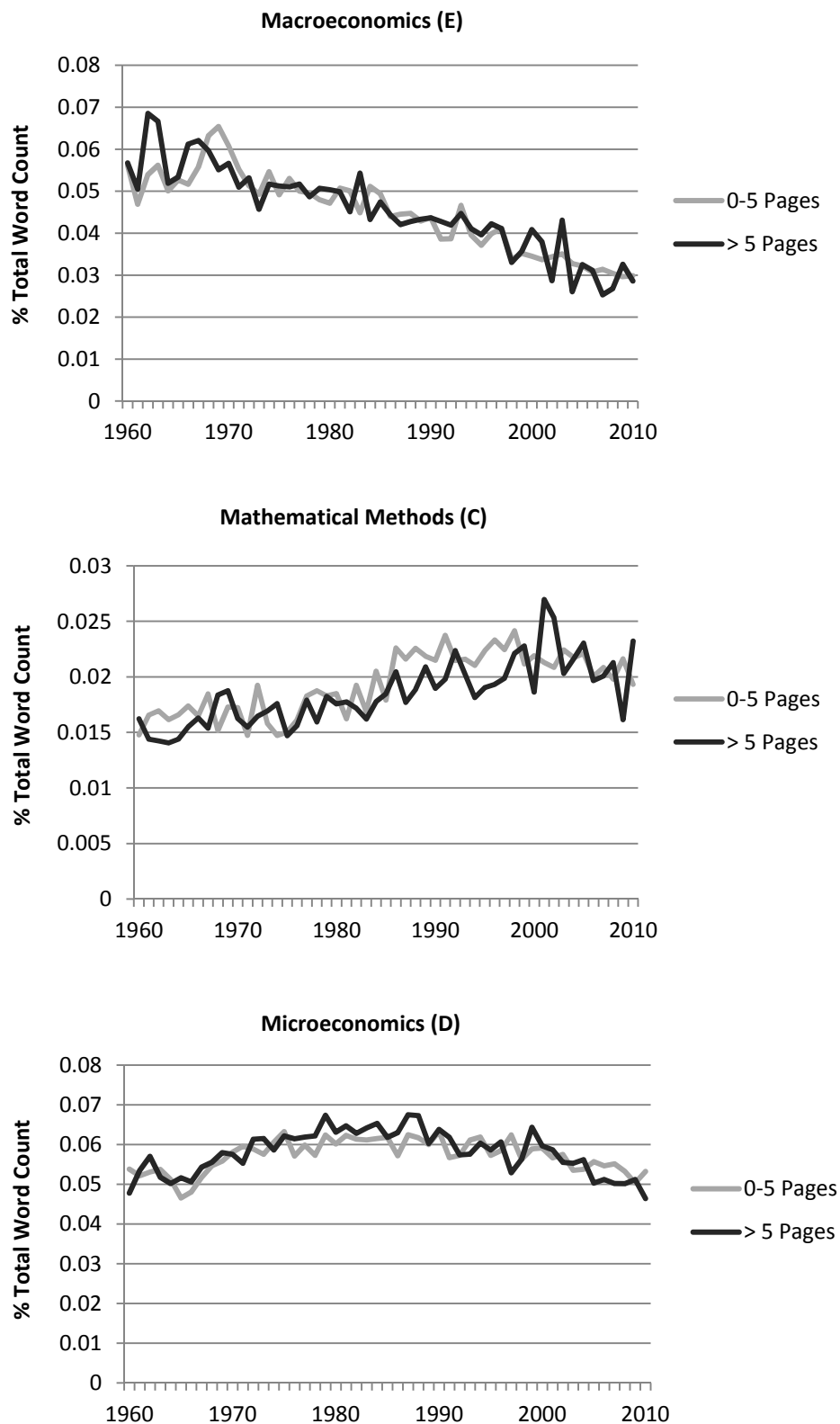


Figure 13 – JEL Categories Frequency of Use, per Page Length, per Year
Macroeconomics (E), Mathematical Methods (C), and Microeconomics (D) Highlighted



**Figure 14 – JEL Categories Frequency of Use, per Page Length, per Year
General Economics and Teaching (A) and History of Economic Thought (B) Highlighted**

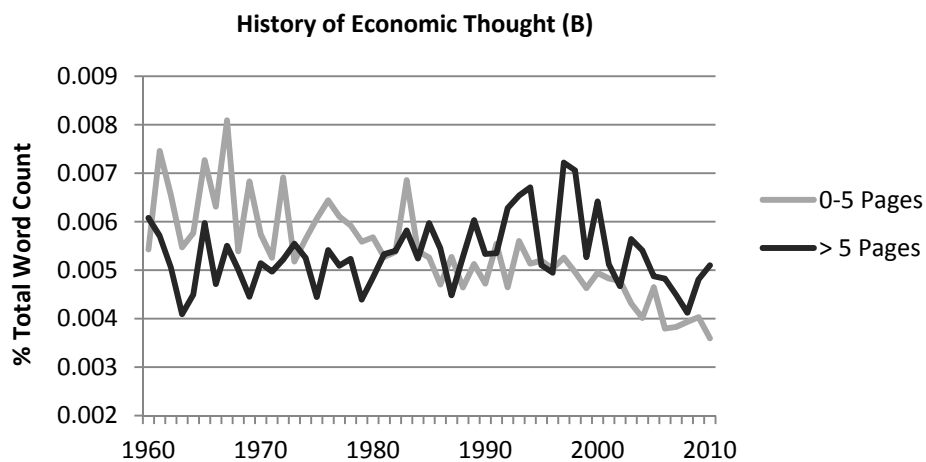
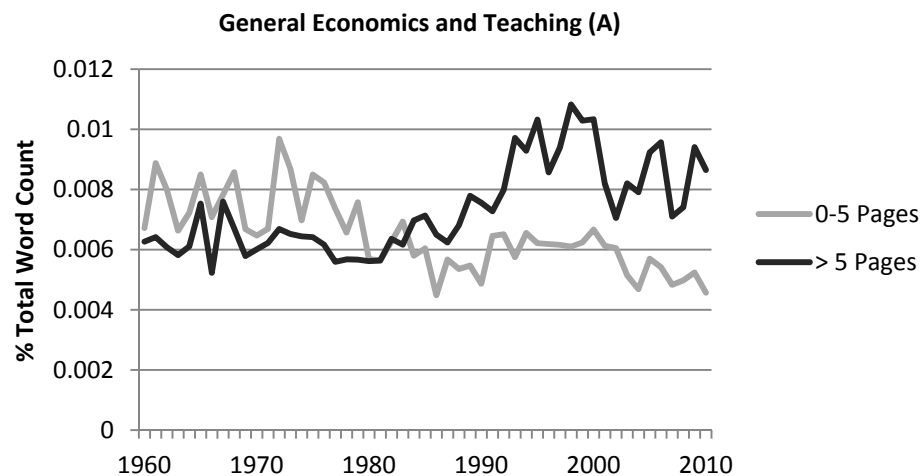
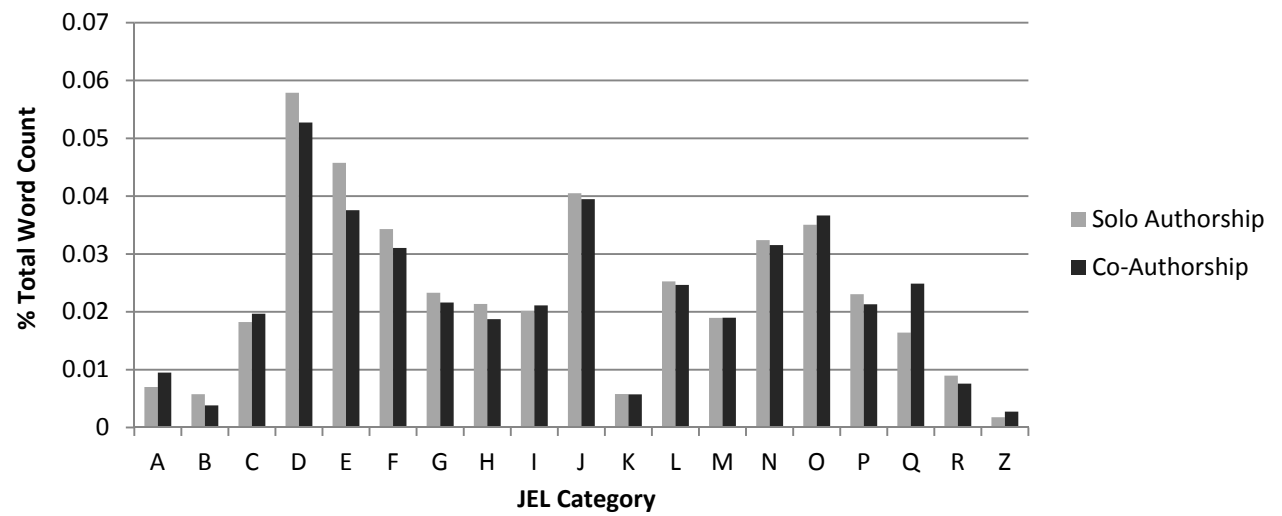
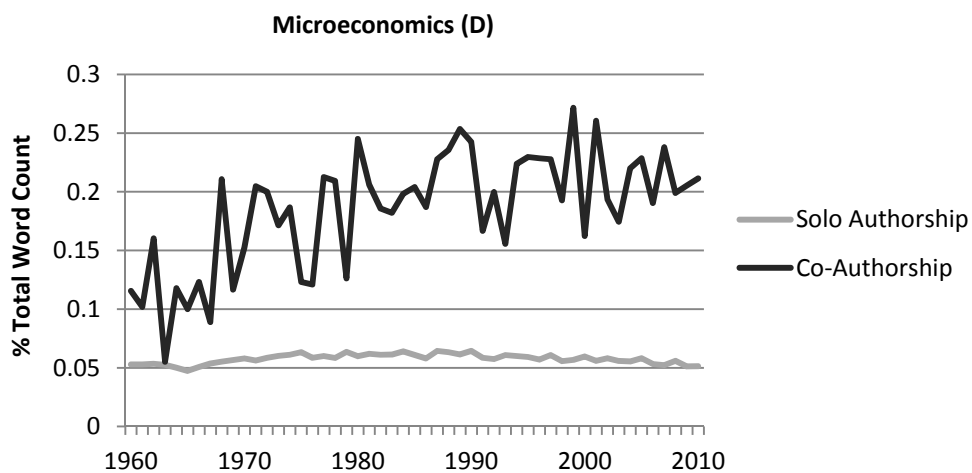
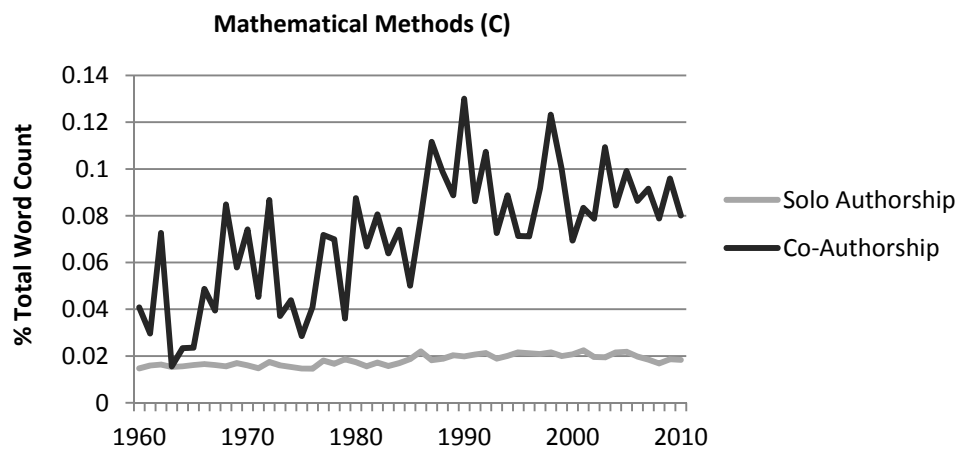
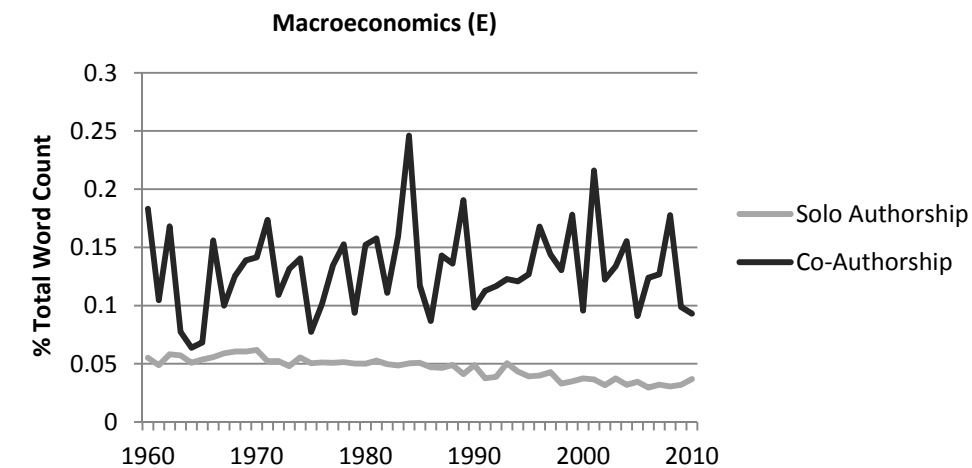


Figure 15 – JEL Categories Frequency of Use, by Authorship Level



**Figure 16 – JEL Categories Frequency of Use, per Authorship Level, per Year
Macroeconomics (E), Mathematical Methods (C), and Microeconomics (D) Highlighted**



References

1. Antweiler, Werner, and Murray Z. Frank. 2004. "Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards." *The Journal of Finance* 59(3):1259-1294.
2. Card, David, and Stefano DellaVigna. 2013. "Nine Facts About Top Journals in Economics." *Journal of Economic Literature* 51(1):144-161.
3. Colander, David. 1989. "Research on the Economics Profession." *The Journal of Economic Perspectives* 3(4):137-148.
4. Cropper, Maureen L. 2000. "Has Economic Research Answered the Needs of Environmental Policy?" *Journal of Environmental Economics and Management* 39:328-350.
5. Dedrick, Don. 1998. *Naming the Rainbow: Color Language, Color Science, and Culture*. Kluwer Academic Publishers.
6. Dershowitz, Idan, Nachum Dershowitz, Moshe Koppel, and Navot Akiva. 2011. "Unsupervised Decomposition of a Document into Authorial Components." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* 1356-1364.
7. DeWall, C. Nathan, Richard S. Pond, Jr., W. Keith Campbell, and Jean M. Twenge. 2011. "Tuning in to Psychological Change: Linguistic Markers of Psychological Traits and Emotions Over Time in Popular U.S. Song Lyrics." *Psychology of Aesthetics, Creativity, and the Arts* 5(3):200-207.
8. Duncan, David F. 1991. "Health Psychology and Clinical Psychology: A Comparison Through Content Analysis." *Psychological Reports* 68:585-586.
9. Eagly, Robert V. 1975. "Economics Journals as a Communications Network." *Journal of Economic Literature* 13(3):878-888.
10. Ellis, Lee, Cora Miller, and Alan Widmayer. 1988. "Content Analysis of Biological Approaches in Psychology: 1894-1985." *Sociology and Social Research* 72(3):145-149.
11. Engemann, Kristie M., and Howard J. Wall. 2009. "A Journal Ranking for the Ambitious Economist." *Federal Reserve Bank of St. Louis Review* 91(3):127-139.
12. Evans, James A., and Jacob G. Foster. 2011. "Metaknowledge." *Science* 331:721-725.
13. Fuchs, Victor. 2002. "Reflections on Economic Research and Public Policy John R. Commons Award Lecture, 2002." *The American Economist* 46(1):3-9.

14. Gentzkow, Matthew, and Jesse M. Shapiro. 2010. "What Drives Media Slant? Evidence from U.S. Daily Newspapers." *Econometrica* 78(1):35-71.
15. Goyal, Sanjeev, Marco J. van der Leij, and Jose L. Moraga-Gonzalez. 2006. "Economics: An Emerging Small World." *Journal of Political Economy* 114(2):403-412.
16. Granger, Clive W.J. 1994. "Reducing Self-Interest and Improving the Relevance of Economic Research." In: *Logic, Methodology and Philosophy of Science IX: Proceedings of the Ninth International Congress of Logic, Methodology and Philosophy of Science*.
17. Grimmer, Justin, and Gary King. 2010. "General Purpose Computer-Assisted Clustering and Conceptualization." *Proceedings of the National Academy of Sciences* 1:1-8.
18. Hamermesh, Daniel S. 2014. "Age, Cohort, and Co-Authorship." *Working Paper*.
19. Hamermesh, Daniel S. 2013. "Six Decades of Top Economics Publishing: Who and How?" *Journal of Economic Literature* 51(1):162-172.
20. Hawkins, Robert G., Lawrence S. Ritter, and Ingo Walter. 1973. "What Economists Think of Their Journals." *Journal of Political Economy* 81(4):1017-1032.
21. Kagann, Stephen, and Kenneth W. Leeson. 1978. "Major Journals in Economics: A User Study." *Journal of Economic Literature* 16(3):979-1003.
22. Kalaitzidakis, Pantelis, Theofanis P. Mamuneas, and Thanasis Stengos. 2001. "Rankings of Academic Journals and Institutions in Economics." *Working Paper*.
23. Kim, E. Han, Adair Morse, and Luigi Zingales. 2006. "What Has Mattered to Economics Since 1970." *The Journal of Economic Perspectives* 20(4):189-202.
24. Koppel, Moshe, Shlomo Argamon, and Anat Rachel Shimoni. 2011. "Automatically Categorizing Written Texts by Author Gender." *Working Paper*.
25. Kosnik, Lea-Rachel. 2014. "Determinants of Contract Completeness: An Environmental Regulatory Application." *International Review of Law and Economics* 37:198-208.
26. Kremer, Michael, and Charles Morcom. 2000. "Elephants." *American Economic Review* 90(1):212-234.
27. Laband, David N., and Robert D. Tollison. 2000. "Intellectual Collaboration." *Journal of Political Economy* 108(3):632-662.
28. Laband, David N., Robert D. Tollison, and Gokhan Karahan. 2002. "Quality Control in Economics." *Kyklos* 55:315-334.

29. Laver, Michael, Kenneth Benoit, and John Garry. 2003. "Extracting Policy Positions from Political Texts Using Words as Data." *American Political Science Review* 97(2):311-331.
30. Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. "Computational Social Science." *Science* 323:721-723.
31. Medema, Steven G., and Samuels, Warren J., eds. 1996. *Foundations of Research in Economics: How Do Economists Do Economics?* Brookfield, VT: Edward Elgar.
32. Michel, Jean-Baptisti, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. 2011. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331:176-182.
33. Pardey, Philip G., and Vincent H. Smith. 2004. *What's Economics Worth? Valuing Policy Research*. Baltimore, MD: The Johns Hopkins University Press.
34. Scroggs, Claud L. 1975. "The Relevance of University Research and Extension Activities in Agricultural Economics to Agribusiness Firms." *American Journal of Agricultural Economics* 57(5):883-888.
35. Sen, Amartya. 2007. "The Discipline of Economics." *Economica* 75:617-628.
36. Sexton, J. Bryan, and Robert L. Helmreich. 1999. "Analyzing Cockpit Communication: The Links Between Language, Performance, Error, and Workload." In: *Proceedings of the Tenth International Symposium on Aviation Psychology* 689-695. Columbus, OH: The Ohio State University.
37. Smith, Vincent H., Philip G. Pardey, and Connie Chan-Kang. 2004. "The Economics Research Industry." In *What's Economics Worth? Valuing Policy Research*, eds. P.G. Pardey and V.H. Smith. Baltimore, MD: The Johns Hopkins University Press.
38. Spirling, Arthur. 2010. "Bargaining Power in Practice: U.S. Treaty-Making with American Indians 1784-1911." *Working Paper*.
39. Stephen, Timothy. 1999. "Computer-Assisted Concept Analysis of HCR's First 25 Years." *Human Communication Research* 25(4):498-513.
40. Stephen, Timothy. 2000. "Concept Analysis of Gender, Feminist, and Women's Studies Research in the Communication Literature." *Communication Monographs* 67(2):193-214.
41. Stern, David I. 2013. "Uncertainty Measures for Economics Journal Impact Factors." *Journal of Economic Literature* 51(1):173-189.

42. Tetlock, Paul C. 2007. "Giving Content to Investor Sentiment: The Role of Media in the Stock Market." *Journal of Finance* 62(3):1139-1168.
43. Weiss, Sholom M., Nitin Indurkha, Tong Zhang, and Fred J. Damerau. 2005. *Text Mining: Predictive Methods for Analyzing Unstructured Information*. New York, New York: Springer Science+Business Media, Inc.

Appendix

Procedural and Methodological Detail

Utilizing bag-of-word subject dictionaries to discern thematic content is a well accepted practice in computational text analysis (Weiss et al., 2005), however, the details in performing it in any given application are not necessarily straightforward. Textual analysis, while it offers unique benefits in terms of volume, objectivity, and speed is not, as Laver et al. (2003) succinctly put it, “a methodological free lunch.”

In this context, decisions had to be made about how to score the thematic dictionaries to arrive at the composite frequencies of use. Say a dictionary for category X is composed of ten words $(x_1, x_2, \dots, x_{10})$, if a corpora contains one instance of all ten words, is it given a score of 10? What about if it contains just one of the words, but ten times, does that garner a 10 rating as well?

There is no accepted standard here, and how one scores the thematic dictionary is generally up to the researcher. In this paper, since we are not trying to exclusively categorize each research paper into a single category (i.e. papers can have both macroeconomic and microeconomic content; indeed, many authors themselves frequently assign their papers multiple JEL categories across subjects), we chose a simplistic scoring method that is equitable across the bag-of-words. All words counted equally (i.e. we did not get into a subjective weighting of some keywords being more “macroeconomic” than others), and multiple uses of a word counted each time to the same extent. In other words, a cut of the corpora in our study (be it by journal, year, authorship level, etc.) is scored by simple counts of all the words in each of the 19 dictionaries. This means that most individual papers had counts in more than one subject dictionary. As most authors assign their own papers across JEL subject categories, we found this to be appropriate and acceptable in this research context. In other contexts, one may wish to uniquely categorize each paper to a single subject category, but in the study performed in this paper, we found that to be unnecessary, and indeed even counterintuitive.

We would also like to make mention here of the programs and procedures used to create the relational dataset of journal articles and associated characteristics. All of the journal articles were manually downloaded from JSTOR. A script could not be written to do this because of the Aaron Swartz controversy; JSTORs website is sophisticated enough now that it automatically interrupts any script from downloading too many articles from its site at one time. An irobot script, however, was written to scrape from the web all the characteristics of each of the papers, including title, year, author names, page numbers, etc. The text analysis frequencies were computed through the flexible WordStat software which, while performing the basic frequency calculations, allows for data export to perl and python for final data manipulation and organization.