Overview of the Dataset on Occupational Characteristics

Enghin Atalay, Phai Phongthiengtham, Sebastian Sotelo, Daniel Tannenbaum[1]

Abstract

This document provides an overview of the dataset we have constructed on the text content of job ads which appeared in the Boston Globe, New York Times, and Wall Street Journal. We describe the underlying data, outline our algorithm to turn the raw text into a structured dataset, and then explore some of the characteristics of the resulting dataset.

To construct a dataset on trends in occupational characteristics, we transform the unstructured text of newspaper job ads into a structured database linking job titles — alternatively Standard Occupational Classification (SOC) codes — to occupational characteristics, including work styles, skill and knowledge requirements, commonly performed tasks, and technology usage. This transformation of unstructured text is a multi-step process. First, we re-format the raw text, removing any superfluous markup, and correcting to the extent possible the spelling mistakes which were generated by ProQuest in their effort to transcribe text from the original data sources. Second, we discard all non-job-related advertisements. At this point, our individual text files contain pages of job ads, with multiple ads on a single page. In a third step, we identify the boundaries between individual ads and, concurrently, establish the vacancy posting's job title. Fourth, we elicit occupation-related information (the work styles, skill and knowledge requirements, and so on) from the text of each individual job ad. In the other documents on this website, we detail our efforts in performing these four steps.

In this document, we discuss the dataset which results from our four-step procedure. We discuss the number of job ads we are able to process for each year and newspaper, trends in the properties of the ads we process, and summary statistics for the variables we construct.

Figure 1 plots the number of cleaned job ads in our dataset for each year between 1940 and 2000. In total, there are 9.4 million job ads which were posted in these years. The

[1]Atalay and Phongthiengtham: Department of Economics, University of Wisconsin-Madison. Sotelo: Department of Economics, University of Michigan-Ann Arbor. Tannenbaum: Department of Economics, University of Nebraska-Lincoln. We acknowledge financial support from the Washington Center for Equitable Growth.

Figure 1: Number of Job Ads Per Year

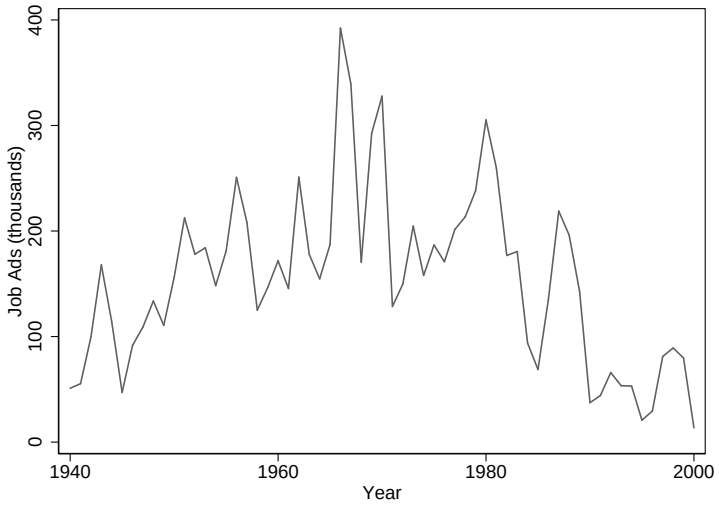
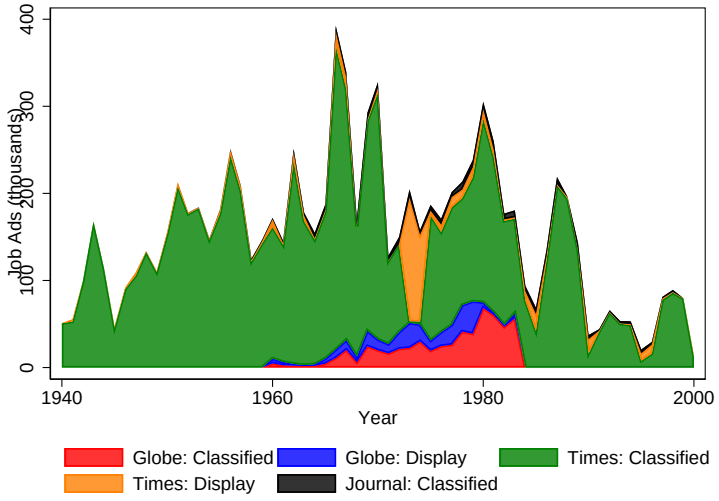


Figure 2: Number of Job Ads Per Year, by Subsample



number of job ads per year was highest in the middle decades of our sample, 228 thousand job ads per year in the 1960s and 198 thousand per year in the 1970s, than it was in the later decades (178 thousand job ads per year in the 1980s and 52 thousand per year in the 1990s) or earlier decades (98 thousand job ads per year in the 1940s and 179 thousand job ads per year in the 1950s). There are two reasons for this non-monotonic relationship in the number of ads per year in our sample. First, our dataset contains data from the Boston Globe only between 1960 and 1983. Second, there seems to be have been a secular decline in the number of ads which were posted in the New York Times (the largest component of our sample) beginning in the 1950s. In Figure 2, we plot the number ads for the three newspapers in our sample, separately plotting the number of Display Ads and Classified Ads.[2] Most of the ads in our sample, 8.6 million out of 9.4 million, are Classified rather than Display Ads. Moreover, 88 percent of the ads in our sample were published in the New York Times, 9 percent in the Boston Globe, and 2 percent in the Wall Street Journal.

In Figure 3, we plot the number of correctly spelled words for each of the subsamples of text. This figure depicts first that the number of words per ad is somewhat higher in Display Ads (61 words per ad) than Classified Ads (53 words per ad), and second that the number of words per ad is greater for more recent years: 58 words per ad in the 1990s, compared to 52 words per ad in the 1960s.[3] Given these trends, it will be necessary (as we do in the paper) to control for changes in ad length when characterizing the evolution of occupational characteristics.

Figure 5 plots the fraction of ads in our dataset for which we have not been able to identify the job's Standard Occupation Classification (SOC) code. We are unable to classify ads' SOC codes primarily in cases in which the job title is unique, where the job title is mentioned only once in our sample. Examples of these types of job titles include "bilingual spanish counselor," "costume jewelry bookkeeper," and "sales rep printing ink." On average, 33 percent of of the ads in our dataset have a missing SOC code, with a moderate increase trend in the fraction of ads for which an SOC code is missing: Within the 1960-2000 period, the sample period for *The Evolving U.S. Occupational Structure,* 36 percent of the ads in our dataset have a missing SOC code.

In Table 1, we compare the SOC composition of ads, for the top 20 occupations, across the five different subsamples in the dataset. The shares of vacancies, across occupations, is

[2]As we mention in the paper, "Compared to Classified Ads, Display Ads are less abundant, have larger font text, may contain photographs or other graphic art (something never included in Classified Ads), and are more expensive for firms to post. Since the text in Display Ads are larger, our code more easily identifies and processes the text in these ads. For this reason, we apply our processing code separately for these different types of ads."

[3]The average ad length was longest, ad 70 words per ad, in the 1940s.

Figure 3: Average Ad Length, by Subsample

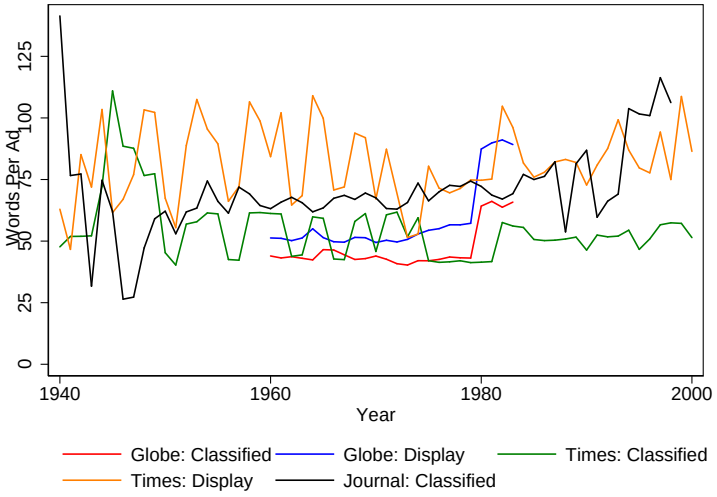


Figure 4: Number of Job Ad Words Per Year, by Subsample

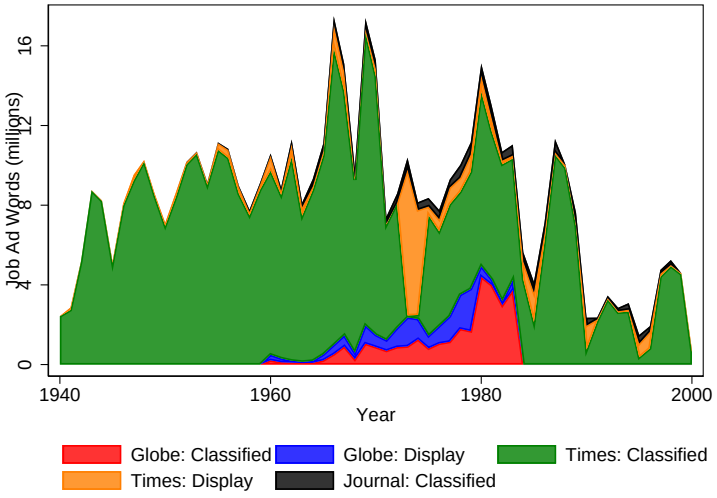
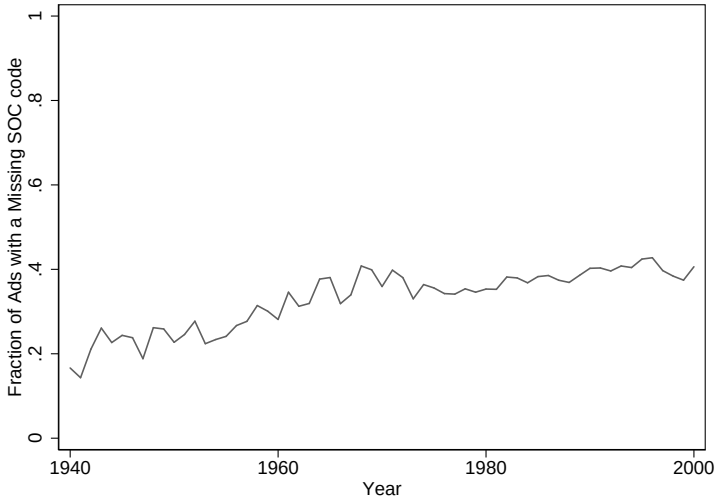


Figure 5: Fraction of Job Ads without a Matched SOC Code



broadly similar across the five subsamples, with somewhat larger shares in managerial and financial occupations in Display Ads, and larger shares in office and secretarial occupations in Classified Ads.

Tables 2 to 9 present summary statistics for the different groups of occupational characteristics. The first four tables, Tables 2 to 5, present the average frequency of mentions of different O*NET Elements. For the average O*NET Element, there are approximately 0.088 mentions of words related to the particular occupational characteristic per job ad. Our procedure recovers a somewhat larger number of words, 0.110 per job ad, for Knowledge Requirements than for Work Styles (0.040 mentions per job ad). Finally, Tables 6 through 9 give summary statistics for job ads' mentions of technology usage, of experience and education requirements, of work activities (using Spitz-Oener, 2006,'s definitions), and worker skills (using Deming and Kahn, 2017,'s definitions).

Finally, Figure 6 depicts the heterogeneity (both in the level and in the trends) of keyword mentions across 2-digit SOC codes. Each of the five panels of this figure presents the keyword frequencies of one of the five Spitz-Oener task measures. According to this figure, for example, managerial occupations are both relatively intensive in non-routine analytic and non-routine interactive tasks and have faster than average growth rates throughout the sixty year sample period. Mentions of non-routine analytic and non-routine interactive tasks are broadly increasing within most 2-digit SOC occupations. The bottom panels depict similarly broad within-occupation trends in the frequency of mentions of words related to routine manual and routine cognitive tasks.

Table 1: SOC Composition by Newspaper Subsample

	Boston Globe		New York Times		Wall Street Journal
SOC Code	Classified	Display	Classified	Display	Classified
414012	0.020	0.015	0.021	0.021	0.035
436012	0.028	0.020	0.027	0.020	0.001
132011	0.012	0.017	0.023	0.018	0.026
439022	0.018	0.013	0.038	0.016	0.000
172051	0.008	0.018	0.010	0.026	0.010
412031	0.013	0.010	0.027	0.012	0.009
113031	0.006	0.007	0.006	0.010	0.041
291141	0.018	0.019	0.013	0.014	0.001
111011	0.003	0.003	0.006	0.010	0.038
151131	0.010	0.022	0.006	0.014	0.005
411011	0.009	0.008	0.008	0.011	0.017
436014	0.011	0.006	0.021	0.010	0.003
231011	0.002	0.002	0.006	0.007	0.031
436011	0.008	0.007	0.011	0.013	0.007
411012	0.004	0.004	0.003	0.007	0.028
433021	0.010	0.005	0.019	0.007	0.001
172141	0.005	0.015	0.003	0.012	0.006
172199	0.005	0.014	0.002	0.011	0.007
433031	0.010	0.008	0.011	0.006	0.002
112021	0.002	0.003	0.001	0.005	0.022
999999	0.378	0.333	0.320	0.360	0.362

Notes: The correspondence between SOC codes and SOC titles are as follows: 414012: Sales Representatives, Wholesale and Manufacturing; 436012: Legal Secretaries; 132011: Accountants and Auditors; 439022: Word Processors and Typists; 172051: Civil Engineers; 412031: Retail Salespersons; 113031: Financial Managers; 291141: Registered Nurses; 111011: Chief Executives; 151131: Computer Programmers; 411011 First-Line Supervisors of Retail Sales Workers; 436014: Secretaries and Administrative Assistants, Except Legal, Medical, and Executive; 231011: Lawyers; 436011: Executive Secretaries and Executive Admin. Assistants; 411012: First-Line Supervisors of Non-Retail Sales Workers; 433021: Billing and Posting Clerks; 172141: Mechanical Engineers; 172199: Engineers, All Other; 433031: Bookkeeping, Accounting, and Auditing Clerks; 112021: Marketing Managers; 999999: Missing.

Table 2: Summary Statistics: Work Styles (Keyword Mentions Per Ad)

O*NET Work Styles	1940s	1950s	1960s	1970s	1980s	1990s	Avg.
1C1a Achievement/Effort	0.008	0.007	0.013	0.018	0.028	0.041	0.017
1C1b Persistence	0.003	0.004	0.006	0.007	0.009	0.009	0.006
1C1c Initiative	0.037	0.035	0.046	0.040	0.043	0.033	0.040
1C2b Leadership	0.047	0.046	0.072	0.100	0.140	0.180	0.090
1C3a Cooperation	0.007	0.005	0.004	0.007	0.010	0.015	0.007
1C3b Concern for Others	0.165	0.204	0.238	0.240	0.262	0.264	0.230
1C3c Social Orientation	0.022	0.019	0.015	0.013	0.024	0.038	0.019
1C4a Self Control	0.023	0.036	0.039	0.030	0.025	0.015	0.031
1C4b Stress Tolerance	0.008	0.009	0.012	0.012	0.011	0.012	0.011
1C4c Adaptability/Flexibility	0.022	0.015	0.015	0.015	0.026	0.028	0.019
1C5a Dependability	0.064	0.038	0.027	0.018	0.024	0.021	0.030
1C5b Attention to Detail	0.060	0.053	0.059	0.068	0.101	0.150	0.073
1C5c Integrity	0.009	0.007	0.009	0.007	0.007	0.008	0.008
1C6 Independence	0.002	0.002	0.004	0.006	0.007	0.008	0.005
1C7a Innovation	0.016	0.017	0.023	0.026	0.035	0.048	0.026
1C7b Analytical Thinking	0.007	0.008	0.017	0.021	0.034	0.062	0.021

Table 3: Summary Statistics: Skills (Keyword Mentions Per Ad)

O*NET Skill	1940s	1950s	1960s	1970s	1980s	1990s	Avg.
2A1a Reading Comprehension	0.056	0.046	0.054	0.044	0.062	0.080	0.054
2A1b Active Listening	0.001	0.001	0.003	0.006	0.017	0.033	0.007
2A1c Writing	0.200	0.139	0.115	0.082	0.090	0.107	0.116
2A1d Speaking	0.041	0.029	0.031	0.037	0.062	0.092	0.043
2A1e Mathematics	0.073	0.086	0.123	0.120	0.144	0.184	0.118
2A1f Science	0.154	0.161	0.219	0.210	0.248	0.302	0.210
2A2a Critical Thinking	0.005	0.005	0.009	0.012	0.020	0.034	0.012
2A2b Active Learning	0.003	0.002	0.003	0.006	0.007	0.009	0.004
2A2c Learning Strategies	0.023	0.025	0.040	0.044	0.052	0.075	0.041
2A2d Monitoring	0.051	0.059	0.066	0.074	0.082	0.095	0.070
2B1a Social Perceptiveness	0.003	0.002	0.003	0.003	0.004	0.005	0.003
2B1b Coordination	0.026	0.035	0.064	0.089	0.117	0.139	0.074
2B1c Persuasion	0.001	0.001	0.001	0.002	0.002	0.004	0.002
2B1d Negotiation	0.008	0.005	0.006	0.008	0.022	0.044	0.012
2B1e Instructing	0.028	0.016	0.015	0.017	0.023	0.026	0.019
2B1f Service Orientation	0.096	0.075	0.086	0.086	0.108	0.125	0.091
2B2i Complex Problem Solving	0.011	0.014	0.019	0.022	0.028	0.043	0.021
2B3a Operations Analysis	0.065	0.106	0.155	0.181	0.181	0.182	0.148
2B3b Technology Design	0.104	0.147	0.181	0.198	0.212	0.239	0.179
2B3c Equipment Selection	0.073	0.064	0.054	0.043	0.034	0.023	0.050
2B3d Installation	0.089	0.071	0.076	0.086	0.085	0.075	0.080
2B3e Programming	0.014	0.011	0.045	0.084	0.091	0.132	0.058
2B3g Operation Monitoring	0.095	0.091	0.101	0.120	0.122	0.120	0.108
2B3h Operation and Control	0.129	0.119	0.112	0.121	0.110	0.097	0.116
2B3j Equipment Maintenance	0.179	0.149	0.151	0.149	0.138	0.115	0.148
2B3k Troubleshooting	0.075	0.065	0.073	0.075	0.083	0.106	0.076
2B3l Repairing	0.052	0.026	0.024	0.030	0.030	0.026	0.030
2B3m Quality Control Analysis	0.064	0.077	0.077	0.076	0.070	0.071	0.074
2B4e Judgment and Decision Making	0.022	0.019	0.022	0.024	0.032	0.047	0.025
2B4g Systems Analysis	0.093	0.146	0.200	0.211	0.230	0.265	0.190
2B4h Systems Evaluation	0.092	0.160	0.195	0.192	0.200	0.229	0.180
2B5a Time Management	0.008	0.007	0.006	0.010	0.011	0.013	0.009
2B5b Mgmt. of Financial Resources	0.153	0.115	0.135	0.156	0.191	0.239	0.154
2B5c Mgmt. of Material Resources	0.061	0.088	0.135	0.158	0.169	0.188	0.133
2B5d Mgmt. of Personnel Resources	0.064	0.070	0.102	0.134	0.191	0.244	0.124

Table 4: Summary Statistics: Knowledge Requirements (Keyword Mentions Per Ad)

O*NET Knowledge Requirements	1940s	1950s	1960s	1970s	1980s	1990s	Avg.
2C1a Administration and Management	0.110	0.083	0.108	0.137	0.190	0.257	0.134
2C1b Clerical	0.400	0.317	0.249	0.184	0.202	0.101	0.246
2C1c Economics and Accounting	0.057	0.058	0.091	0.118	0.153	0.188	0.104
2C1d Sales and Marketing	0.347	0.287	0.264	0.259	0.330	0.422	0.298
2C1e Customer and Personal Service	0.121	0.130	0.141	0.160	0.215	0.289	0.164
2C1f Personnel and Human Resources	0.063	0.082	0.108	0.111	0.138	0.135	0.106
2C10 Transportation	0.069	0.035	0.026	0.019	0.021	0.018	0.030
2C1f Production and Processing	0.227	0.184	0.181	0.176	0.162	0.141	0.179
2C2b Food Production	0.133	0.097	0.073	0.059	0.038	0.024	0.071
2C3a Computers and Electronics	0.139	0.227	0.260	0.261	0.274	0.312	0.247
2C3b Engineering and Technology	0.107	0.158	0.203	0.211	0.230	0.272	0.195
2C3c Design	0.162	0.197	0.194	0.192	0.198	0.232	0.194
2C3d Building and Construction	0.148	0.090	0.075	0.058	0.058	0.048	0.077
2C3e Mechanical	0.256	0.273	0.233	0.189	0.154	0.096	0.211
2C4a Mathematics	0.073	0.086	0.123	0.120	0.144	0.184	0.118
2C4b Physics	0.190	0.221	0.273	0.251	0.276	0.311	0.253
2C4c Chemistry	0.070	0.074	0.088	0.074	0.079	0.086	0.079
2C4d Biology	0.128	0.108	0.132	0.114	0.134	0.155	0.125
2C4e Psychology	0.126	0.103	0.125	0.115	0.143	0.170	0.125
2C4f Sociology and Anthropology	0.015	0.016	0.029	0.032	0.041	0.049	0.029
2C4g Geography	0.027	0.017	0.011	0.011	0.012	0.014	0.014
2C5a Medicine and Dentistry	0.048	0.037	0.040	0.056	0.087	0.104	0.056
2C5b Therapy and Consulting	0.010	0.009	0.013	0.026	0.045	0.046	0.023
2C6 Education and Counseling	0.275	0.220	0.256	0.231	0.271	0.278	0.250
2C7a English Language	0.023	0.015	0.015	0.012	0.019	0.025	0.017
2C7b Foreign Language	0.114	0.067	0.053	0.042	0.058	0.083	0.063
2C7c Fine Arts	0.006	0.005	0.007	0.005	0.006	0.005	0.006
2C7d History and Archaeology	0.001	0.001	0.001	0.002	0.002	0.005	0.002
2C7e Philosophy and Theology	0.003	0.003	0.002	0.003	0.003	0.005	0.003
2C8a Public Safety and Security	0.017	0.013	0.012	0.020	0.024	0.034	0.018
2C8b Law and Government	0.065	0.048	0.042	0.061	0.086	0.089	0.061
2C9a Telecommunications	0.015	0.027	0.037	0.039	0.062	0.079	0.040
2C9b Communications and Media	0.083	0.073	0.091	0.100	0.146	0.197	0.106

Table 5: Summary Statistics: Work Activities (Keyword Mentions Per Ad)

O*NET Work Activity	1940s	1950s	1960s	1970s	1980s	1990s	Avg.
4A1a1 Getting Information	0.068	0.043	0.027	0.019	0.017	0.014	0.030
4A1a2 Monitor Processes, Materials	0.125	0.115	0.112	0.098	0.095	0.097	0.107
4A1b1 Identifying Objects, Actions, Events	0.012	0.010	0.017	0.026	0.037	0.055	0.023
4A1b2 Inspecting Equipment, Structures	0.187	0.170	0.132	0.110	0.084	0.048	0.126
4A1b3 Estimating Products Characteristics	0.380	0.324	0.324	0.352	0.358	0.353	0.344
4A2a1 Judging the Qualities of Things, People	0.002	0.002	0.003	0.004	0.007	0.010	0.004
4A2a2 Processing Information	0.009	0.015	0.027	0.036	0.047	0.049	0.030
4A2a3 Evaluating Information on Compliance	0.014	0.022	0.032	0.039	0.047	0.066	0.035
4A2a4 Analyzing Data or Information	0.027	0.032	0.043	0.050	0.055	0.079	0.045
4A2b1 Making Decisions, Solving Problems	0.016	0.017	0.023	0.026	0.032	0.055	0.025
4A2b2 Thinking Creatively	0.030	0.029	0.036	0.033	0.034	0.042	0.033
4A2b3 Updating, Using Relevant Knowledge	0.091	0.084	0.107	0.110	0.124	0.154	0.107
4A2b4 Developing Objectives and Strategies	0.047	0.075	0.100	0.123	0.148	0.213	0.111
4A2b5 Scheduling Work and Activities	0.061	0.057	0.061	0.071	0.081	0.087	0.068
4A2b6 Organizing, Planning, Prioritizing Work	0.020	0.021	0.035	0.050	0.072	0.102	0.045
4A3a1 Performing General Physical Activities	0.062	0.061	0.083	0.102	0.117	0.105	0.088
4A3a2 Handling and Moving Objects	0.070	0.070	0.059	0.054	0.041	0.026	0.056
4A3a3 Controlling Machines and Processes	0.207	0.202	0.204	0.196	0.167	0.134	0.191
4A3a4 Operating Vehicles, Mechanized Devices	0.195	0.197	0.179	0.151	0.137	0.108	0.166
4A3b1 Interacting With Computers	0.056	0.060	0.086	0.115	0.166	0.260	0.110
4A3b2 Drafting, Specifying Devices	0.222	0.255	0.249	0.231	0.205	0.180	0.231
4A3b4 Repairing, Maintaining Mech. Equip.	0.171	0.153	0.131	0.110	0.086	0.054	0.122
4A3b5 Repairing, and Maintaining Elec. Equip.	0.170	0.161	0.167	0.161	0.144	0.116	0.157
4A3b6 Documenting/Recording Information	0.020	0.019	0.019	0.022	0.034	0.049	0.024
4A4a1 Interpreting for Others	0.007	0.012	0.025	0.037	0.047	0.048	0.029
4A4a2 Communicating Inside Organization	0.112	0.112	0.138	0.147	0.183	0.184	0.143
4A4a3 Communicating Outside Organization	0.053	0.055	0.077	0.106	0.133	0.164	0.092
4A4a4 Establishing, Maintaining Relationships	0.009	0.009	0.016	0.026	0.060	0.117	0.030
4A4a5 Assisting and Caring for Others	0.006	0.005	0.007	0.011	0.017	0.021	0.010
4A4a6 Selling or Influencing Others	0.159	0.146	0.135	0.137	0.183	0.213	0.154
4A4a7 Resolving Conflicts, Negotiating	0.009	0.010	0.013	0.014	0.017	0.031	0.014
4A4a8 Working Directly with the Public	0.480	0.379	0.371	0.429	0.468	0.500	0.422
4A4b1 Coordinating Others ' Work/Activities	0.100	0.109	0.163	0.204	0.257	0.291	0.180
4A4b2 Developing and Building Teams	0.034	0.047	0.054	0.068	0.085	0.132	0.064
4A4b3 Training and Teaching Others	0.037	0.026	0.033	0.040	0.056	0.065	0.040
4A4b4 Guiding, Directing, Motivating	0.016	0.013	0.017	0.024	0.036	0.059	0.024
4A4b5 Coaching and Developing Others	0.014	0.030	0.036	0.048	0.058	0.082	0.042
4A4b6 Provide Consultation and Advice	0.014	0.016	0.018	0.016	0.018	0.024	0.017
4Ac1 Performing Administrative Activities	0.118	0.128	0.132	0.136	0.160	0.124	0.135
4Ac2 Staffing Organizational Units	0.078	0.095	0.125	0.131	0.156	0.135	0.122
4Ac3 Monitoring, Controlling Resources	0.043	0.051	0.059	0.063	0.068	0.081	0.060

Table 6: Summary Statistics: Technologies (Keyword Mentions Per Thousand Ads)

	1940s	1950s	1960s	1970s	1980s	1990s	Avg.
APL	0.3	0.2	0.3	0.5	0.8	0.3	0.4
BAL (Basic assembly language)	1.2	0.8	1.6	5.6	2.4	0.8	2.4
CAD	0.0	0.0	0.0	0.2	1.0	0.4	0.3
CICS	0.0	0.0	0.0	2.8	6.3	4.8	2.1
Cobol	0.0	0.0	3.1	12.3	10.6	7.8	5.8
C++	0.0	0.0	0.0	0.0	0.0	1.9	0.1
DB2	0.0	0.0	0.0	0.0	1.1	6.2	0.6
DOS	1.6	0.9	2.4	8.0	10.5	13.3	5.4
EDP	0.1	0.5	9.2	12.5	7.7	1.2	6.5
FORTRAN	0.0	0.0	1.4	3.7	3.7	0.9	1.9
Foxpro	0.0	0.0	0.0	0.0	0.0	2.1	0.1
HTML	0.0	0.0	0.0	0.0	0.0	4.3	0.3
IBM 360	0.1	0.4	2.6	2.7	0.2	0.0	1.3
IBM 370	0.1	0.2	0.1	2.9	1.5	0.1	1.0
IMS	0.9	0.7	1.0	2.9	4.0	3.2	2.1
JAVA	0.3	0.1	0.2	0.5	0.4	5.8	0.6
Job Control Language	0.1	0.1	0.3	2.6	2.4	1.8	1.2
LAN (Local Area Network)	1.2	1.0	0.9	1.1	1.5	9.8	1.7
Lotus 123	0.0	0.0	0.0	0.0	2.5	6.2	0.9
Lotus Notes	0.0	0.0	0.0	0.0	0.0	3.6	0.2
Microsoft Excel	0.0	0.0	0.0	0.0	0.1	4.5	0.3
Microsoft Power Point	0.0	0.0	0.0	0.0	0.0	5.8	0.4
Microsoft Word	0.0	0.0	0.0	0.0	0.6	17.3	1.2
MVS	0.0	0.0	0.0	0.6	3.9	3.2	1.1
Novell	0.0	0.0	0.0	0.0	0.2	7.2	0.5
Oracle	0.1	0.0	0.1	0.1	0.5	9.8	0.7
PASCAL	0.1	0.0	0.1	0.1	1.5	0.6	0.4
Quark	0.0	0.0	0.0	0.0	0.1	8.2	0.5
RPG	0.0	0.0	0.1	0.5	0.8	0.3	0.3
SQL	0.3	0.2	0.1	0.1	0.4	7.2	0.6
Sybase	0.0	0.0	0.0	0.0	0.0	5.5	0.3
TCP (Transmission Control Protocol)	0.0	0.0	0.0	0.0	0.0	3.2	0.2
TSO (TCP Segment Offloading)	0.1	0.0	0.1	0.6	1.3	0.9	0.4
UNIVAC	0.0	0.2	0.6	0.8	0.4	0.0	0.5
UNIX	0.0	0.0	0.0	0.1	2.3	18.8	1.6
VAX	0.1	0.0	0.0	0.1	3.1	2.9	0.8
VisualBasic	0.0	0.0	0.0	0.0	0.0	4.4	0.3
VMS (Operating System)	0.1	0.0	0.0	0.1	1.6	2.7	0.5
VSAM	0.0	0.0	0.0	0.2	1.4	1.3	0.4
WordPerfect	0.0	0.0	0.0	0.0	2.1	11.0	1.1

Table 7: Summary Statistics: Schooling and Experience (Keyword Mentions Per Ad)

	1940s	1950s	1960s	1970s	1980s	1990s	Avg.
Experience: 1 year	0.007	0.010	0.017	0.021	0.025	0.020	0.017
Experience: 2 years	0.019	0.027	0.038	0.041	0.049	0.052	0.038
Experience: 3 years	0.015	0.019	0.023	0.030	0.038	0.042	0.027
Experience: 4 years	0.009	0.007	0.006	0.006	0.008	0.011	0.007
Experience: 5 years	0.017	0.019	0.017	0.022	0.027	0.032	0.021
Associates Degree	0.000	0.000	0.000	0.001	0.002	0.001	0.001
Bachelor of Arts	0.015	0.014	0.015	0.015	0.024	0.051	0.019
Bachelor of Sciences	0.005	0.010	0.014	0.017	0.019	0.020	0.014
Masters	0.037	0.029	0.036	0.045	0.071	0.036	0.043
MBA	0.001	0.001	0.006	0.012	0.012	0.015	0.007
PHD	0.002	0.004	0.004	0.004	0.004	0.003	0.004
CPA	0.028	0.023	0.018	0.019	0.021	0.020	0.021

Table 8: Summary Statistics: Spitz-Oener (2006) Task Groups (Keyword Mentions Per Ad)

	1940s	1950s	1960s	1970s	1980s	1990s	Avg.
Literature							
Non-routine Analytic	0.082	0.119	0.141	0.133	0.145	0.174	0.132
Non-routine Interactive	0.157	0.105	0.091	0.085	0.115	0.145	0.107
Non-routine Manual	0.112	0.069	0.070	0.066	0.086	0.095	0.078
Routine Manual	0.032	0.018	0.010	0.007	0.007	0.004	0.012
Routine Cognitive	0.021	0.011	0.007	0.005	0.003	0.001	0.008
Literature+CBOW							
Non-routine Analytic	0.147	0.190	0.242	0.248	0.281	0.369	0.239
Non-routine Interactive	0.350	0.298	0.281	0.284	0.374	0.478	0.322
Non-routine Manual	0.156	0.097	0.097	0.088	0.112	0.127	0.106
Routine Manual	0.136	0.090	0.056	0.036	0.040	0.029	0.062
Routine Cognitive	0.068	0.033	0.024	0.014	0.008	0.003	0.024

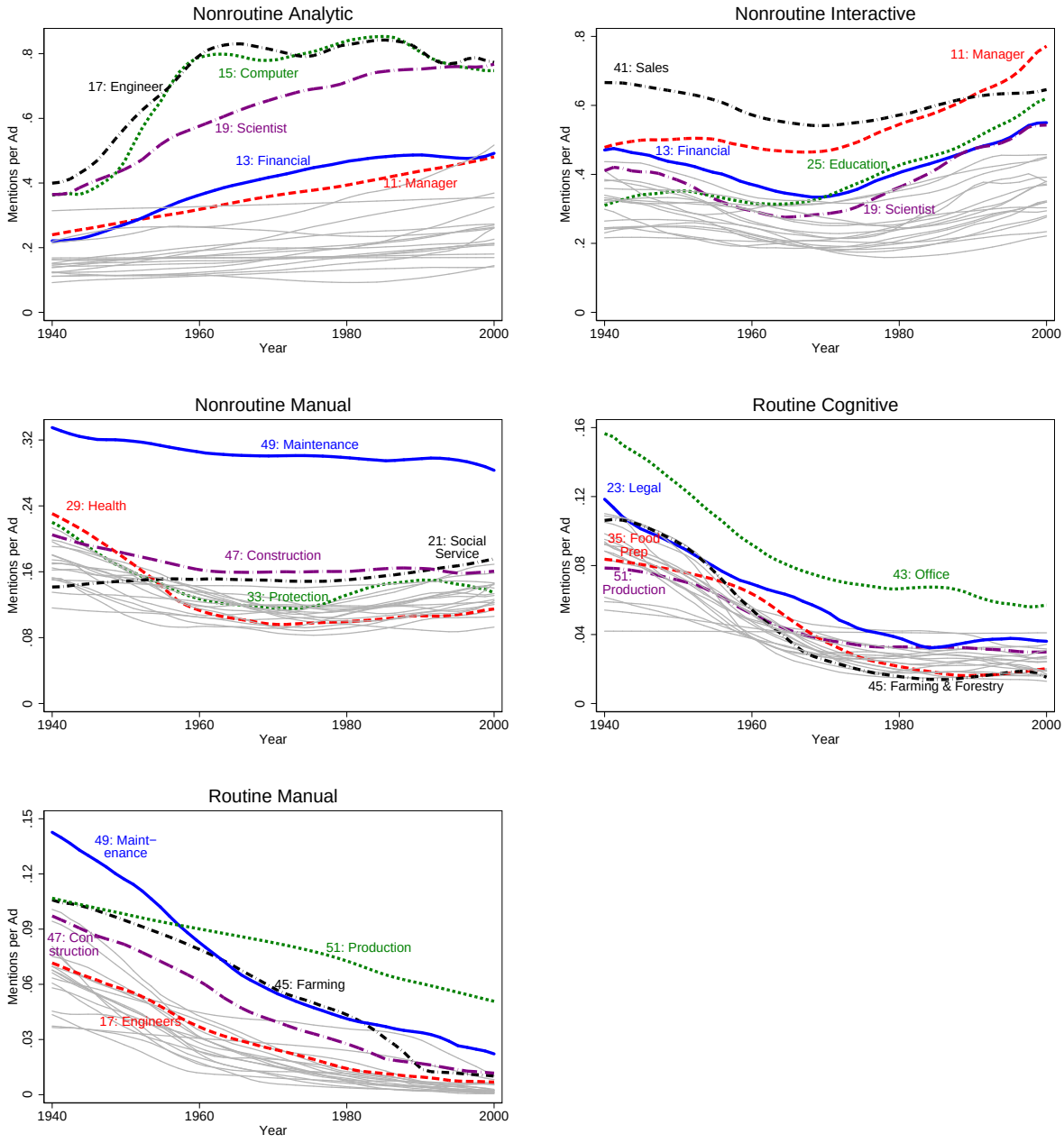
Notes: The set of rows marked "literature" use Spitz-Oener (2006)'s mappings between words and task groups. In the rows marked "Literature+CBOW," we include in our mappings words that are deemed characteristic of the task groups according to our continuous bag of words.

Table 9: Summary Statistics: Deeming and Kahn (2017) Skill Groups (Keyword Mentions Per Ad)

	1940s	1950s	1960s	1970s	1980s	1990s	Avg.
Literature							
Problem Solving	0.033	0.045	0.057	0.044	0.053	0.068	0.050
Social	0.009	0.009	0.012	0.015	0.038	0.076	0.021
Character	0.010	0.005	0.004	0.005	0.018	0.028	0.009
Writing	0.012	0.011	0.012	0.010	0.017	0.026	0.013
Customer Service	0.109	0.115	0.116	0.129	0.164	0.198	0.132
Project Management	0.000	0.000	0.000	0.001	0.002	0.004	0.001
People Management	0.016	0.025	0.041	0.050	0.069	0.080	0.045
Financial	0.066	0.063	0.076	0.090	0.105	0.123	0.084
Computer	0.000	0.005	0.020	0.039	0.063	0.122	0.034
Literature+CBOW							
Problem Solving	0.057	0.070	0.088	0.072	0.095	0.134	0.082
Social	0.023	0.019	0.026	0.036	0.079	0.133	0.043
Character	0.330	0.256	0.270	0.280	0.386	0.381	0.304
Writing	0.030	0.027	0.028	0.024	0.039	0.061	0.031
Customer Service	0.182	0.171	0.163	0.181	0.253	0.336	0.198
Project Management	0.106	0.169	0.216	0.243	0.266	0.321	0.217
People Management	0.086	0.111	0.160	0.184	0.225	0.215	0.163
Financial	0.152	0.162	0.158	0.162	0.177	0.203	0.165
Computer	0.018	0.047	0.090	0.123	0.174	0.291	0.109

Notes: The set of rows marked "literature" use Deming and Kahn (2017)'s mappings between words and skill groups. In the rows marked "Literature+CBOW," we include in our mappings words that are deemed characteristic of the task groups according to our continuous bag of words.

Figure 6: Trends in Keyword Frequencies: Spitz-Oener (2006) Task Groups



Notes: Each panel plots the trends in keyword frequencies, at the 2-digit SOC level. We highlight five of the SOCs with the highest keyword frequencies using thick, colored lines. The other SOCs we plot with lightly-shaded thinner lines.