Dynamics of Job Search Effort and Vacancies: Evidence from Classified Advertisements*

Huixin Bi[†], Nicolas Petrosky-Nadeau[‡], Nora Traum[§], and Greg Woodward[¶]

March 20, 2025

Abstract

We introduce new monthly U.S. city-level and national measures of job search effort and vacancies from 1900 to 1938, providing new insights of the response of search effort to labor market conditions. We provide breakdowns of these measures by gender and show trends are similar across gender, while posts for/by women are more volatile than those for/by men. Combining scanned images of U.S. newspapers with image processing operations, we construct local and national measures of job posts placed by both firms and job seekers in the classified advertisements. To our knowledge, we are the first to systematically document and use the "situations-wanted" advertisements placed by job seekers. We find that job search effort – measured as the ratio of the situations-wanted index to the unemployed – is strongly procyclical. Labor market tightness' cyclicality closely mirrors search effort over the period, and we find parallels with post-1950s labor market dynamics.

JEL Classification: J64, N32, E24, E32, C82

Keywords: job search effort; help-wanted; Great Depression

*The views expressed herein are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Banks of Kansas City or San Francisco, or the Federal Reserve System. We thank Brandyn Bok, Joël Cuerrier, Owen Engbretson, Christine El-Rami, August Hays-Ekeland, and Nish Sinha for excellent research assistance. Traum gratefully acknowledges support by funding from the Canadian Social Sciences and Humanities Research Council.

[†]Federal Reserve Bank of Kansas City. Email: huixin.bi@kc.frb.org

[‡]Federal Reserve Bank of San Francisco. Email: nicolas.petrosky-nadeau@sf.frb.org

[§]HEC Montreal. Email: nora.traum@hec.ca

[¶]Federal Reserve Bank of Kansas City. Email: greg.Woodward@kc.frb.org

1 Introduction

Labor market outcomes over the business cycle – such as the duration of unemployment spells, firms' hiring practices and compensation to workers – are impacted by the intensity with which individuals search for work. If the onset of a recession leads individuals to redouble their efforts to find work, theory suggests a more rapid labor market recovery ensues (e.g., Mukoyama *et al.*, 2018). If instead individuals lower their effort in response to fewer job opportunities, the process of matching job seekers to job openings declines, possibly prolonging a recession (e.g., Gomme and Lkhagvasuren, 2015). Moreover, job search effort is central to the welfare-efficiency trade-off in the provision of unemployment insurance benefits (e.g., Chetty, 2008). Quantifying the response of job search effort to economic conditions is crucial for understanding dynamics in the labor market. However, a lack of direct measures for search effort at regular frequencies and over significant amounts of time has constrained past research on this topic.

This paper presents novel measures of job search effort from a previously untapped resource: classified advertisements placed by job seekers – the *situations-wanted ads* – in early twentieth century U.S. newspapers. Most researchers are familiar with help-wanted ads, which have been used to proxy changes in job openings and labor demand. However, prior to the 1940s, advertisements directly placed by job seekers were routinely found alongside postings for help wanted by employers. These situation-wanted ads allow for a novel and deeper understanding of the dynamics of the labor market stretching back to the early twentieth century. To our knowledge, we are the first to systematically document and use the situations-wanted advertisements placed by job seekers.

We construct city-level and national help- and situations-wanted indices from 1900 to 1938 at a monthly frequency. In addition, we provide breakdowns for each city-level measure by gender. Using our national measure, we find that job search effort – measured as the ratio of the situations-wanted index to the unemployed – is strongly procyclical: it declines at a recession's onset, remains low during a recession, and rises during an expansion. We also find that the cylicality of labor market tightness – measured as the ratio of help-wanted ads to the unemployed – closely mirrors that of search effort. Trends across gender are similar, while posts for/by women are more volatile than those for/by men. Ads for women were particularly more sensitive to the Spanish flu, World War I, and the Great Depression.

To construct our labor indices, we extract individual advertisements from multiple newspapers spanning broad national coverage for the time using an innovative technique applied to scanned images of the classifieds sections in the *New York Times*, *Chicago Tribune*, *Atlanta Constitution*, *Los Angeles Times*, and *Washington Post* between January 1900 and Decem-

ber 1938.[1] Specifically, our approach uses morphological operations to identify columns and isolate each ad within a column. We then create template images of section headers within the classified pages, e.g. "Wanted Situations" or "Help Wanted," and identify sections based on matching images of headers. Finally, we use the ad counts within the help- and situations-wanted sections as input for city-level and national situations- and help-wanted indices. To the best of our knowledge, our methodology for systematically categorizing classified advertisements, which is based extensively on morphological operations, is novel in application to economics.

We validate our methodology for identifying individual labor ads against extensive manual checks and a contemporaneous newspaper-based index of help-wanted. A comparison of manual counts conducted at regular intervals for each newspaper in our sample with the counts from our methodology shows a high degree of precision in capturing both types of labor ads. That is, across newspapers, the average median error between the two methods is 4% for help wanted and 3% for situations wanted. We then externally validate our national help-wanted indicator against a measure available for a sub-period of our time frame: the help-wanted index constructed by the Metropolitan Life Insurance Company (MetLife), which is available from 1919 onward. Our national help-wanted index, constructed by weighting individual metropolises by their city-level population, has a correlation of 0.98 with MetLife's national help-wanted index for the overlapping period.[2]

Job advertisements from both employers and job seekers vary greatly during our sample period. Posts by firms decline sharply at the onset of a recession, and there is a strong negative correlation between unemployment and job openings - the Beveridge curve - similar to what is observed in data for the second half of the 20th century. Posts by job seekers display less of a cyclical pattern, tending to hold steady at the onset of a recession when the number of job seekers is likely rising, although the cyclicality is more pronounced in the period after World War I relative to the pre-war period. Nevertheless, job search effort – measured as the ratio of our situations-wanted measure to the unemployed – is strongly procyclical. It declines at the onset of a recession, remains low during a recession, and rises during an expansion. Furthermore, labor market tightness – measured as the ratio of our help-wanted ads to the unemployed – is highly positively correlated with job search effort.

This paper is related to several strands of the literature. From a measurement perspective,

[1]The scanned images are accessed through rights from ProQuest. Previous work suggests substantial challenges in identifying individual ProQuest ads, which often appear as a single raw text file for *all* classified ads on a single day in a newspaper. Atalay *et al.* (2020) relied on regular patterns of job ads to discern the boundaries, such as a help-wanted ad that begins with a word with uppercase letters. Unfortunately, for our sample, there is no such systematic pattern.

[2]Additionally, the appendix discusses records of help-wanted counts collected by local employment agencies between 1924 and 1932 and provides a comparison with our city-level help-wanted ad counts.

we provide new city- and national-level time series spanning several decades at a monthly frequency to help characterize the dynamics of firm job posting and worker job search behavior. Regarding the latter, most studies of job search effort have relied on time diaries in the annual American Time Use Survey supplement to the Current Population Survey administered by the Bureau of Labor Statistics. However, using this data, researchers have attained conflicting conclusions: job search effort of various individuals can be procyclical (Gomme and Lkhagvasuren, 2015), countercyclical (Mukoyama *et al.*, 2018, Ahn and Shao, 2017), or acyclical (DeLoach and Kurt, 2013). Separately, a recent literature studies job search effort using job applications from online job search platforms (e.g., Faberman and Kudlyak, 2019, Marinescu and Skandalis, 2021). Although this literature has provided new insights on how individuals search, the duration of the samples is short, and often too infrequent, to adequately address questions relative to the business cycle. Faberman *et al.* (2022) uses responses to specific job search questions in the Federal Reserve Bank of NY Survey of Consumer Expectations (SCE) to provide new evidence on variation in search behavior across employment status. The new evidence is used to calibrate a general equilibrium model with on-the-job search that results in pro-cyclical aggregate search effort.

Our paper is also related to several works that use data to examine labor markets in the early twentieth century. Lee (2016) uses data from public employment offices between 1924 and 1932 to estimate a matching function that abstracts from search effort. To our knowledge, very few researchers have explicitly and systematically utilized the situation-wanted advertisements. DeVaro and Gurtler (2018) examine how help- and situations-wanted ads vary over centuries by manually counting the ads on a specific date vicennially. Laurent (2020) manually counts three days of ads per year using the *Kansas City Star* to study how female employment varied from 1920-1936. Simon (2001) examines the asked wage rate for clerical positions in situations-wanted ads of the *New York Times* by focusing on the first Sunday paper of each month between 1929 and 1934.[3] By automatizing the procedure and process of extracting ads from historical newspapers, we improve on this literature by offering systematic measures using both the help- and situations-wanted advertisements over decades in the first half of the twentieth century.

The rest of this paper is organized as follows. Section 2 provides an overview of job postings in classified ads, as well as our methodology and strategy for developing indices based on these ads. Section 3 presents the time series properties of our help- and situations-wanted indices. Section 4 further disaggregates our city-level measures by gender. Section 5 concludes.

[3]In order to provide a perspective on the pre-Depression era, Simon (2001) also collects data for the period between 1925 and 1928 using the first Sunday paper of March, June, September, and December.

2 Background and Data Construction

Classified advertisements from both employers and job seekers were common in the early twentieth century. While limited information is available on the hiring process in this period, surveys suggest that firms utilized a mix of newspaper ads, employment agencies and informal methods to fill vacancies (see Simon, 2001). An advantage of newspapers was wide circulation and readership, providing an ideal outlet for advertising positions to a broad regional market.[4]

In this paper, we utilize the digitized records of widely-circulated newspapers in major metropolises to build a monthly database of advertisements by employers and job seekers. The number of newspapers entering our analysis is constrained by one main consideration: after 1923, U.S. newspapers became subject to copyright laws that restricted public access, see Heald (2014). We consider five newspapers with wide circulation in their respective cities with direct purchase of access rights from ProQuest between January 1, 1900 and December 31, 1938: (i) *New York Times*; (ii) *Chicago Tribune*; (iii) *Los Angeles Times*; (iv) *Atlanta Constitution*; and (v) *Washington Post*. According to the 1930 Census, the population of these five metropolitan areas accounted for close to 25 percent of all metropolitan populations and 10 percent of the total US population. In the following subsections, we first present an overview of classified ads, followed by our methodology to extract labor ads and verify the accuracy of our counts.

2.1 Classified advertisements

Advertisements for help wanted by employers and positions sought by workers generally ran next to each other in the classified section of newspapers in the first half of the 20th century. These advertisements cover a wide range of occupations and industries. They also often contain information on offered or asked compensation, as well as education and/or experience of job seekers and requirements from employers. In addition, ads frequently included information on race, ethnicity, gender, marital status and other characteristics required by employers.[5]

As an example, Figure 1 offers a snapshot of some ads seeking to employ women (first column) and some posted by women looking for jobs (second column) from the *New York Times* on January 19, 1930. Both sides of the market advertised for similar positions, such as

[4]Education had markedly increased by the early twentieth century, and the illiteracy rate was only 4 percent by 1930. In addition, overall enrollment rates for 5- to 19- year olds rose from 51 percent in 1900 to 75 percent in 1940 (Snyder, 1993).

[5]The quality of the Proquest data limits the possibility of systematically collecting this information at this time (see section 2.2 for further details). For this reason, in the following sections, we focus on systematically extracting information solely based on the counts of advertisements following individual header categories.

Figure 1: Help and Situations Wanted Ads: Examples from the *New York Times*, January 19, 1930.



Help Wanted—Female.

BOOKKEEPER, experienced; Yorkville section. R 387 Times.

CHEMIST, college training in standard methods of water analysis. G. R. Spalding, Hackensack Water Co., Oradell, N. J.

CONTROLLER, for first-class hotel; must be absolutely thorough and hard worker; give references and full particulars in first letter. X 2008 Times Annex.

DENTAL ASSISTANT-SECRETARY, refined, cultured, willing young lady. Apply in person Monday, 9-11 A. M., Dr. Alan Mishkin, 240 West 98th.

DENTAL ASSISTANT; write details, knowledge of dental experience; $18 to start. Write Box A. B., 19 Greenwich Av.

JANITRESS.—man and wife; experienced steam, hot water; salary and rooms. Apply at once, Heckman, 136 West 96th.

Situations Wanted—Female.

AMANUENSIS—Services available very capable secretary, 25, American, experienced business and personal work; difficult employers. Phone Sunday, Hackensack 2670J, or write M 421 Times.

ARTIST, experienced in women's fashions, accessories, pen and ink, wash, desires steady position. Miss Barbero, 1,760 Walton Av.

ASSISTANT to doctor or similar capacity; educated, refined, splendid personality, character, initiative, young, attractive; moderate salary. Adirondack 7584.

ASSISTANT BOOKKEEPER, stenographer, typist; experienced; high school graduate. Intervale 8088.

ASSISTANT BOOKKEEPER-TYPIST, thoroughly experienced detail worker; $23. V 672 Times Downtown.

BAKER, assistant to cake baker; experienced colored girl. Edgecombe 1743.

a bookkeeper or medical assistant in this example, as well as advertised or requested specific educational attainment or demographic background. The ads included positions for higher-paid jobs with higher qualification requirements (i.e., the help-wanted ad for a college trained chemist) and lower-paid jobs with less expertise (i.e., the help-wanted ad for a janitor). Although there is notable segregation by gender in the type of occupations advertised during this period (Laurent, 2020), the advertisements cover a wide set of occupations at the time.[6]

The posting of advertisements required a monetary cost that varied over time and across cities, with rates regularly printed in the Sunday edition of certain newspapers (the *Chicago Tribune* and *Atlanta Constitution* are exceptions). We manually collected the posting costs for the *New York Times* and the *Los Angeles Times* and report the nominal cost of an ad placed in the Sunday paper as dotted lines in Figure 2 .[7] Overall the cost was a small fraction of typical earnings. Using the *Los Angeles Times* as an example, in 1925 job seekers paid 23 cents a line on Sundays.[8] For comparison, the average retail price of a pound of sirloin steak in 1925 was 38.5 cents, while the hourly wage rate for a bricklayer in Los Angeles was $1.37

[6]Appendix section D provides a more detailed overview of the occupations advertised in newspaper jobs. The most common occupations for women include domestic services, office work, teachers, and nurses. For comparison, Goldin (1990), pg. 68, notes that the largest categories of female employment over the period were domestic servants, manufacturing operatives, and farm laborers, followed by professional and clerical workers. The occupations of men in the ads span a wide variety of jobs, including sales and advertising, managers, doctors, law enforcement, manual labor, and domestic help.

[7]Prior to December 1920, costs for the *Los Angeles Times* were the same for both help- and situations-wanted ads. The newspaper initially charged ads by the number of words: 1 cent per word, between 1900 and 1903; 1 cent per word with a minimum charge of 25 cents between 1903 to 1918. We assumed 25 cents per ad for the 1903-1918 period and abstract from the period prior to 1903 to keep things comparable. Costs are typically higher on Sundays compared to other days of the week. See Appendix C for additional details.

[8]The exception is the following categories of job seeker ads, which were subject to the same rates as employers, presumably because these categories represent more "self-employed" fields: painting, decorating, carpentering, dressmakers, seamstresses, and nurses.

(Bureau of Labor Statistics, 1925, 1926). Employer advertisements were normally subject to a higher rate. In 1925, it cost employers 36 cents to post a one-line ad on a Sunday. By 1929, the cost increased to 50 cents. In the same year, the *New York Times* charged job seekers 65 cents to post a situations-wanted ad and employers 75 cents to post a help-wanted ad on a Sunday.

Changes to advertising costs were infrequent, especially during the first two decades of the 20th century, and both the frequency and magnitude of changes appear unrelated to business cycle conditions. Interestingly, the real costs in both cities are relatively stable, but decline in the boom period of the 1920s, and initially increase in the Great Depression. To see this, the solid lines in Figure 2 plot the nominal cost deflated by the consumer price index.[9]

2.2 Extracting Labor Advertisements

Identifying labor advertisements requires a novel approach based on extracting labor advertisements from image files of newspapers. We outline the procedure in the following sections.

2.2.1 Splitting Ads Previous studies often apply textual analysis to text files that are extracted from image scans using Optical Character Recognition (OCR). For instance, Atalay *et al.* (2020) relied on regular patterns of job ads in the text files, such as a help-wanted ad beginning with a word with upper case letters, to discern the boundaries between labor ads for the period between 1950 and 2000. While the OCR technology has advanced significantly over the years, this method still struggles with poorly scanned documents and dense text, which are particularly relevant for our data sample.[10] During our period of interest, classified ads were structured in columns of densely placed ads. If applied directly on an entire newspaper page with many columns or even a single column with many ads, state-of-art OCR technology still fails to identify all the break points between ads, leading to significant error rates.[11]

To overcome these challenges, we rely on digital image processing to identify ad boundaries using images of the classified pages. A common formatting style for the classified ads

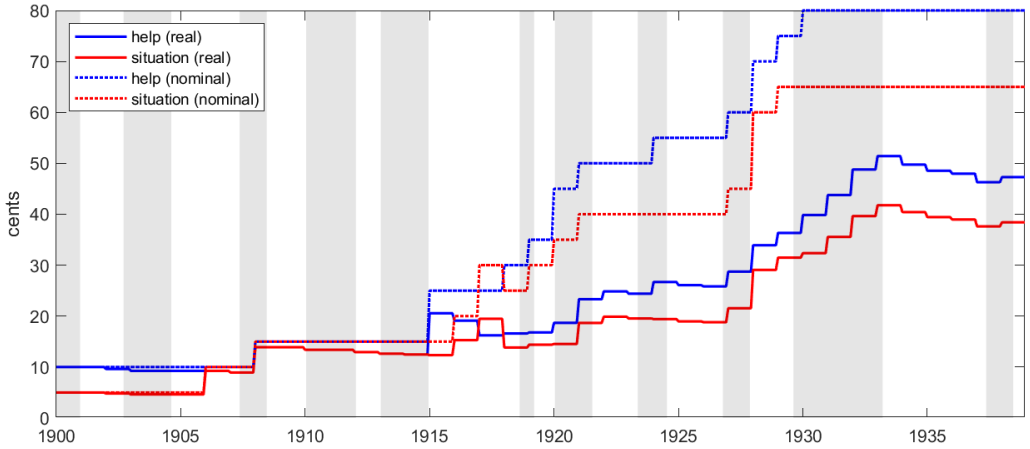
[9] Since 1913, CPI has been computed by the Bureau of Labor Statistics in a consistent manner. Prior to 1913, we use the cost of living index by Albert Rees, maintained by the Minneapolis Federal Reserve. Appendix C provides greater detail on the costs of advertisements over time in our sample of newspapers.

[10] Although ProQuest has produced text files from scanned images of newspaper pages, the text quality for the classified ads in our sample period is too poor to be used for textual analysis. Moreover, ProQuest created a single raw text file for *all* classified ads on a given page, making it challenging to identify individual ads even if the file quality is high.

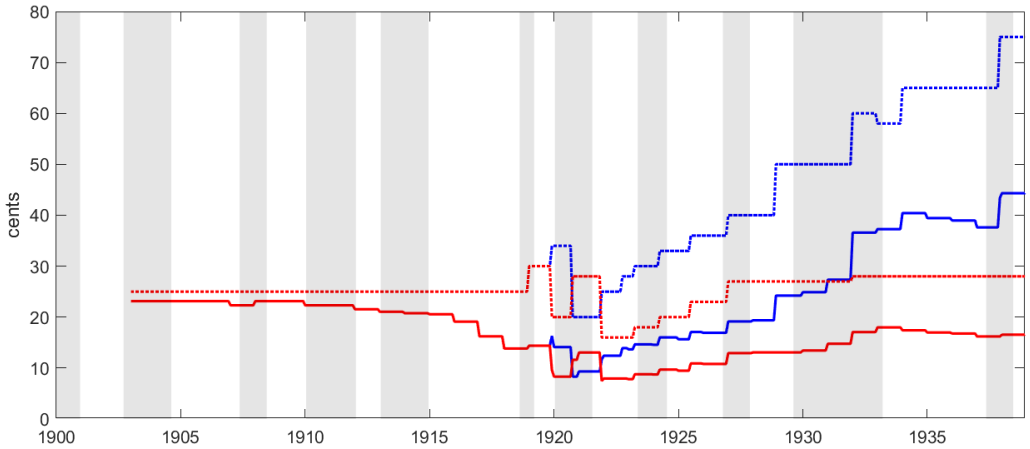
[11] We have also explored Artificial Intelligence technology, but the results are unsatisfying. For instance, for a newspaper column with more than 50 ads, the most recent version of ChatGPT-4o only identifies less than 10 ads.

Figure 2: Costs for posting labor ads: *New York Times* and *Los Angeles Times*.

(a) *New York Times*



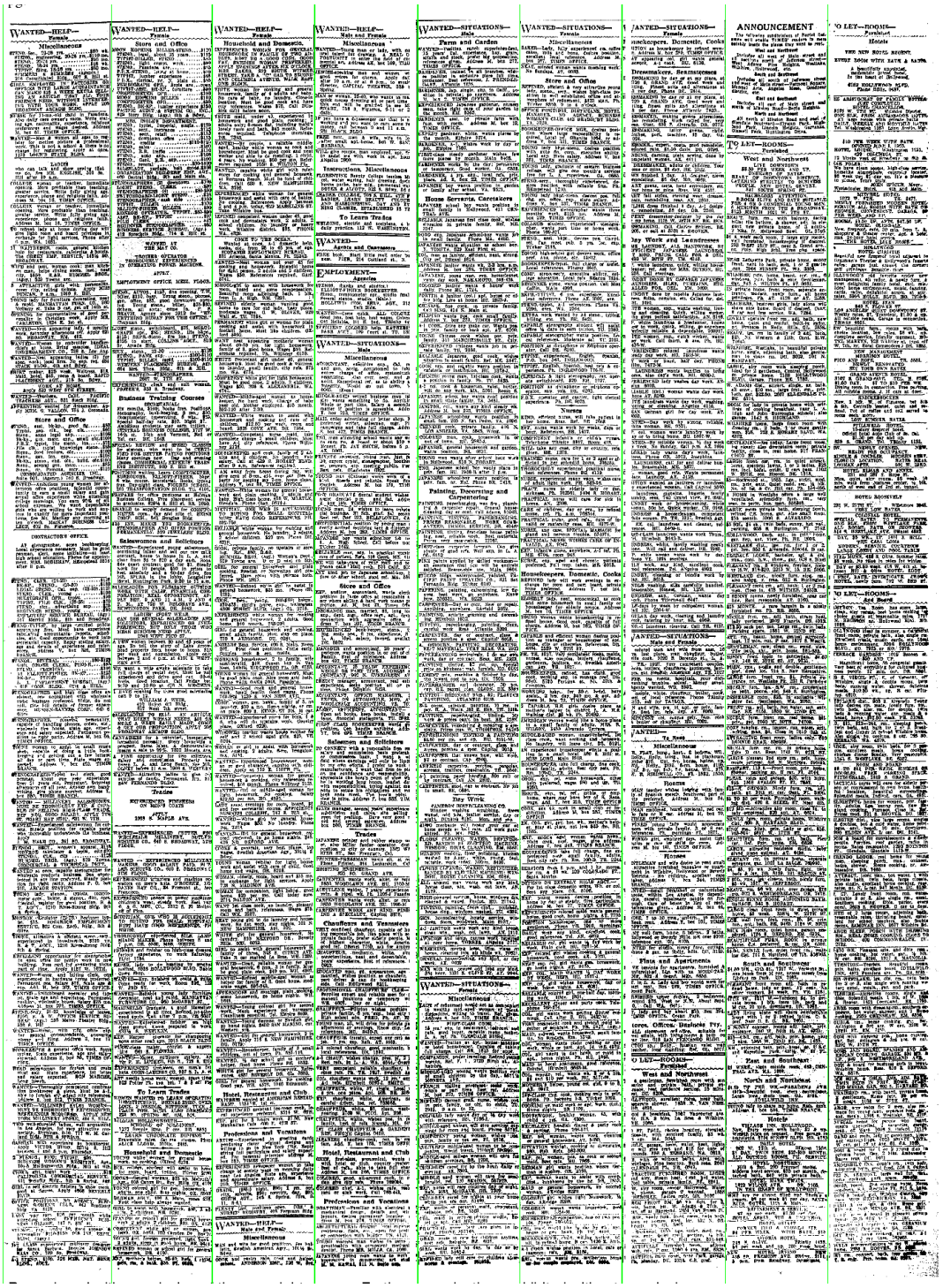
(b) *Los Angeles Times*



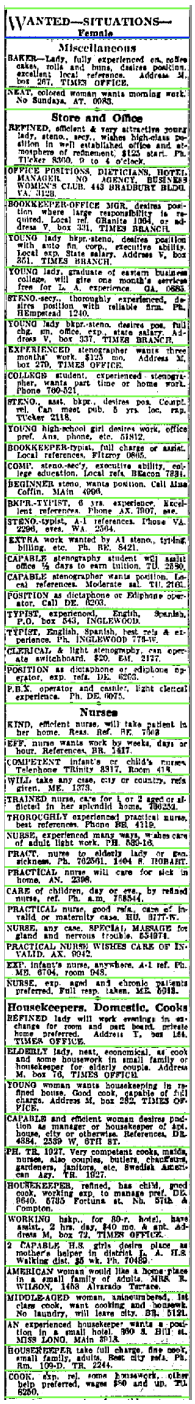
Notes: Prior to December 1920, costs for the *Los Angeles Times* were the same for both help- and situations-wanted ads.

Figure 3: Extracting individual classified ads: example from the *Los Angeles Times*.

(a) Column Splits



(b) Ad Splits



section featured black vertical bars that separated the page's columns as well as black horizontal lines that separated individual ads. We take advantage of this structure to identify and divide columns and ads, as illustrated in panels (a) and (b) of Figure 3, following three steps. First, we straighten any images that aren't properly oriented during the scanning process. We use the Separator Net Post Processor from NewsEye (2020) to identify any skew or slant to images and then use the warpAffine method from OpenCV, an open source computer vision library, to rotate images accordingly. Second, we use NewsEye's Separator Net Post Processor again to split newspaper pages into columns. Specifically, we identify the concentration of black pixels in a narrow vertical band by grouping them into clusters based on their horizontal coordinates. Finally, we split columns into ads, which is the most challenging step in the process because of the density of text. We start by inverting the image colors as it simplifies the process and then remove text from each column. We calculate the mean of each row of pixels, which should be close to a white pixel value for a line with inverted image colors. Depending on the distribution of all pixel values in a column, a prominence threshold is computed and used to identify ad dividers. Appendix A.1 provides details for each of these steps.

2.2.2 Identifying Help- and Situations-wanted Ads

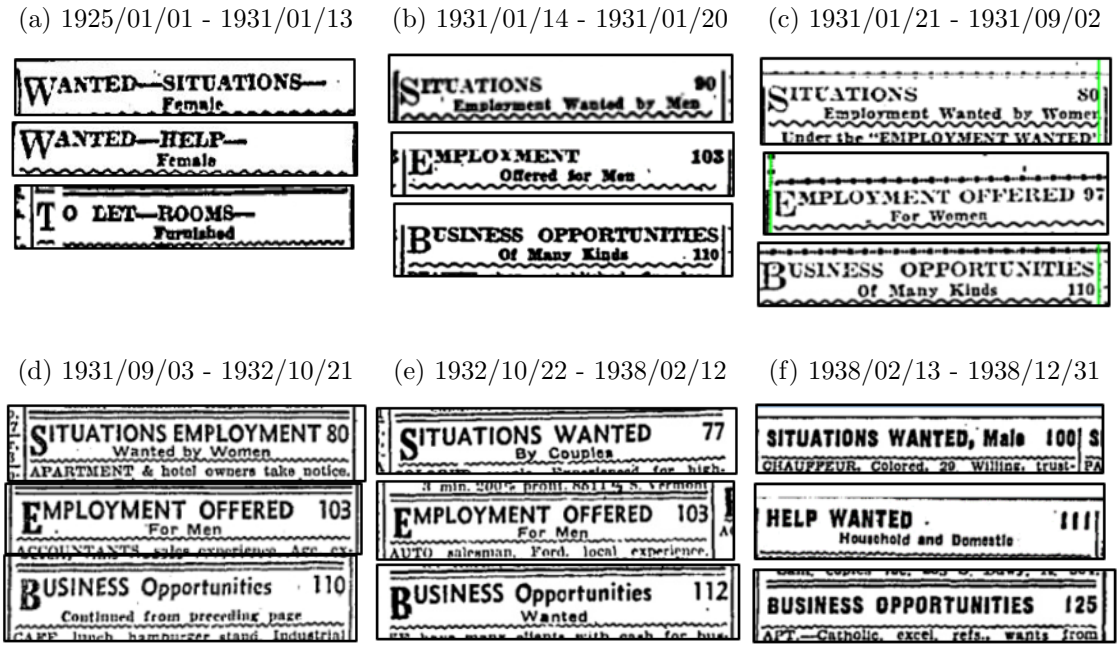
While the image processing procedure allows us to distinguish all ads in the classifieds section, it does not identify the type of ad, e.g., a labor ad posted by an employer versus a person advertising a room for rent. To classify ads, we identify specific types of section headers following an image matching procedure. This approach, detailed next, requires creating image templates that match the section headers used by a particular newspaper at a specific point in time.

The template-matching approach involves three steps. First, we manually identify ad section headers for labor supply (situations-wanted), labor demand (help-wanted), and sections that immediately follow labor ads. Next, we create image templates to match the typesetting of the headers as they appeared in the underlying section header images. Finally, we identify the beginning and the end of the relevant classified sections based on template matching scores for section headers, and sort individuals ads into their respective categories.

Section headers varied, both in terms of wordings and font styles, over time and across newspapers. That is, they often have offset lettering and large font compared to the words in an individual ad. By way of example, Figure 4 illustrates the frequent changes to section headers in the *Los Angeles Times* between January 1, 1925 and December 31, 1938. Until January 13, 1931, labor supply ads were denoted with "Wanted Situations," while labor demand ads started with the header "Wanted Help." The two sections are adjacent and, over this period, always followed by "To Let Rooms," which we use to identify the end of the labor advertisements. After January 13, 1931, the section header for labor supply became

"Situations Employment," while for labor demand it became "Employment Offered,"and the section that followed job postings became "Business Opportunities." The number of times that headers changed varies across newspapers. For instance, while the *Chicago Tribune* changed section headers only twice during the span of 38 years, the *New York Times* made 20 changes to their section headers. An extensive catalog of section headers is provided in Appendix A.2. In total, we create 346 different templates for headers across newspapers over our entire time frame.

Figure 4: Changing Section Headers: examples from the *Los Angeles Times* between January 1, 1925 and December 31, 1938.

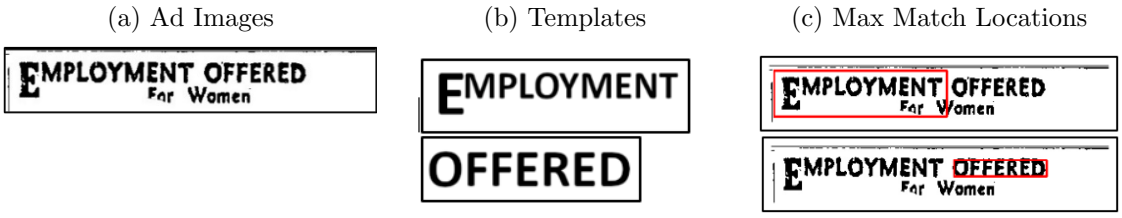


Section headers are identified by matching an image to a template created to align with the typesetting as it appears in the underlying ad images of a newspaper. The matching procedure is as follows: (i) create templates for each identified section header; (ii) use a 2-dimensional convolution to score the "likelihood" of a match between individual ad images and the templates; (iii) scores above a threshold determine whether an image is a section header and its type. In practice, scoring is done by sliding a template across each ad image and measuring the correspondence of the black-and-white pixel locations for the images within the ad. This generates a match score between 0 and 1, where 1 signifies that the template fully matches a pattern of pixels in the underlying ad image. The process is repeated for various templates sizes and the maximum score is recorded.

Finally, headers are classified as either labor supply, demand, or neither based on preset thresholds for image matching scores. We count all ads immediately following an identified

header as belonging to that category, up until another header is identified, or the page ends.

Figure 5: Steps to the Template Matching Procedure: example for identifying "Employment offered" section in the *Los Angeles Times*.

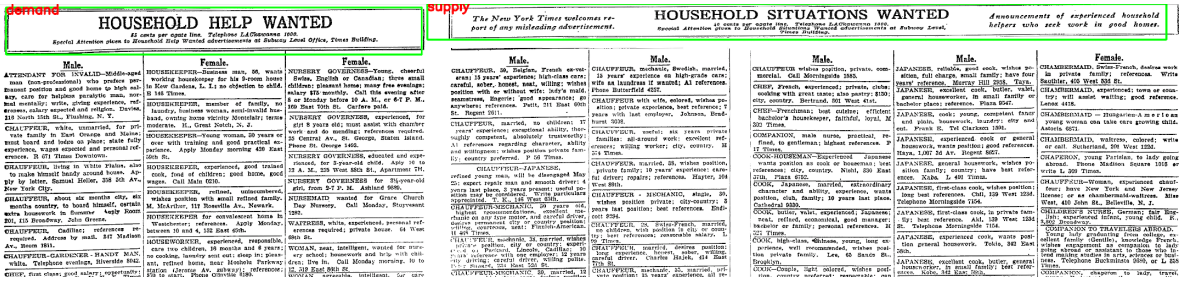


Notes: Templates in (b) are slid across ad images (a) search for the highest correspondence of pixels (c).

Figure 5 illustrates the three steps of template matching using "Employment Offered" in the *Los Angeles Times* in the 1930s. We first identify a header, as shown in the top left panel, and then create two templates, "Employment" and "Offered", that align with the typesetting, size, and font of the header (middle panel). Next, the 2-dimensional convolution finds matches of templates in individual ad images. The red bounding boxes in the last panel give the window with the highest correlation of the black and white pixels between the templates and the image. Finally, scores for "Employment" and "Offered", 0.66 and 0.68 respectively, are recorded and compared to a preset threshold, which is 0.55 in this example. After clearing the threshold, this header is classified as a labor demand header. The accuracy of template matching depends on the setting of thresholds. We use an iterative process to identify the appropriate thresholds for each template, the details of which are in Appendix A.2.

Although the three-step procedure works well with most newspaper pages in our data sample, multicolumn headers appearing in the *New York Times* require some modifications to the approach (see Figure 6). That is, for this newspaper we use an object detection model, a computer vision technique that recognizes and localizes objects of interest through machine learning, to detect and locate the multicolumn headers. First, we use a training sample by manually classifying four types of objects on a large number of newspaper pages: 1) help-wanted multicolumn headers, 2) situations-wanted multicolumn headers, 3) other multicolumn headers, and 4) background (i.e., anything else). Next, an object detection model is trained with this sample. Then, we apply the trained model on each newspaper page to identify locations and classes of multicolumn headers, recording the location of the bounding box and matching probability. Each header is classified according to the matching probability against preset thresholds, for which details are given in Appendix A.3.

Figure 6: Template Matching with Multi-column Headers: example from the *New York Times*.



2.3 Accuracy

The accuracy of labor ad identification is tested with extensive manual counts. We manually counted help- and situations-wanted ads on the second Sunday in May between 1900 and 1938 for the *Los Angeles Times*, *Chicago Tribune*, the *New York Times*, *Atlanta Constitution*, and *Washington Post*, totaling about 1328 newspaper image files and approximately 150 thousand individual ads. We consistently count on the same Sunday, since there were more postings in the classifieds section on Sundays compared to other days of the week. Figure 7 provides a snapshot of the accuracy of our method by comparing the number of ads counted with our algorithm relative to the manual counts.[12] The left panels show the (log) counts of help-wanted ads, while the right panels are for situations-wanted ads. In general, our method works well to correctly identify and count labor advertisements: the observations are well in line with the 45-degree line, suggesting that manual counts and template matching results are very similar.

We further evaluate the performance of our template matching method by providing comparison summary statistics. Specifically, we define a measure of error on a particular day as:

$$E_t = \frac{|C_t^{template} - C_t^{manual}|}{C_t^{manual}}, \tag{1}$$

where C_t^{manual} is the manual count and $C_t^{template}$ is the template matching count.

In general, errors are small across newspapers and across time, despite the frequent changes to section header wording and typesetting. Table 1 summarizes the distribution of errors by newspaper, reporting the median error along with the first and third tertiles.

[12]For the *Los Angeles Times*, the month of May 1933 uses the first Sunday due to poor image quality on the second Sunday. Additional details are documented in Appendix A.4. In addition, we alternatively identified ads based on machine learning with a Convolutional Neural Network (CNN). This method requires more time and human effort, but results are comparable to our template matching approach. See Appendix A.5 for more details.

Table 1: Accuracy of Template Matching Approach: Comparison with Manual Counts

	Number of ads	1st Tertile	Median	3rd Tertile
New York Times				
Help-wanted E	23,656	0.01	0.03	0.05
Situations-wanted E	23,001	0.01	0.04	0.08
Chicago Tribune				
Help-wanted E	47,723	0.02	0.03	0.06
Situations-wanted E	18,551	0	0.01	0.06
Los Angeles Times				
Help-wanted E	20,455	0.03	0.06	0.18
Situations-wanted E	13,859	0.01	0.02	0.07
Atlanta Constitution				
Help-wanted E	4,634	0.03	0.04	0.11
Situations-wanted E	582	0	0.05	0.11

Notes. Errors between machine and manual counts computed according to equation (1). Manual counts collected on the second Sunday in May. See the Appendix for details.

The identification of help- and situations-wanted ads is best for the *Chicago Tribune*, as the median share of misclassified ads is 3% for help-wanted ads and 1% for situations-wanted ads. The template matching method performs the least well in *Los Angeles Times*, which we suspect is due to the relatively poor print quality of several images for this newspaper.

3 Help- and Situations-wanted Time Series

We use the daily counts of categorized ads on Sundays, a day with a significant spike in postings in the classifieds section, to construct monthly measures of help- and situations-wanted ads for each metropolitan area in our sample of newspapers. We then construct national help- and situations-wanted indices using estimates of metro area population as weights in a manner similar to MetLife's methodology for its contemporaneous help-wanted index. The latter serves as an external validation for our aggregated help-wanted counts.

3.1 City-level counts

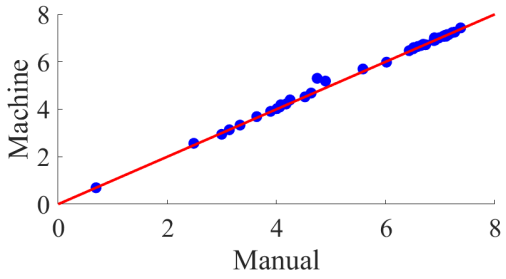
We create monthly series of city-level help- and situations-wanted ads by averaging the Sunday counts within a month. Figure 8 plots the resulting series after seasonal adjustment and taking a 3-month moving average.[13] For New York, we report ad counts beginning in January 1919 due to the paper's low circulation in the earlier years.

Counts of help-wanted ads display similar cyclical patterns across cities between 1900

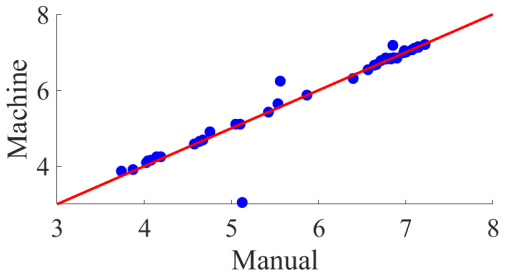
[13]See Appendix B for a comparison of the raw and seasonally adjusted monthly series.

Figure 7: Accuracy of ad counts

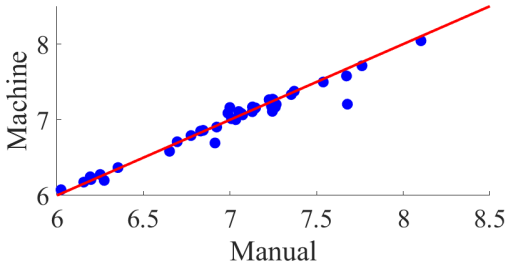
(a) Help Wanted: *New York Times*



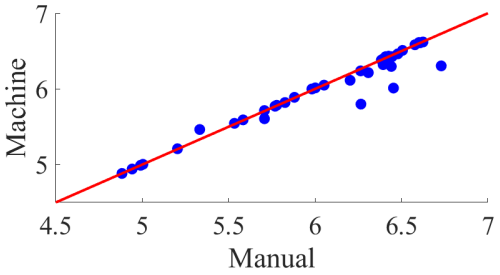
(b) Situations Wanted: *New York Times*



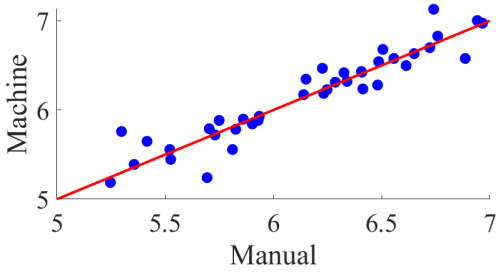
(c) Help Wanted: *Chicago Tribune*



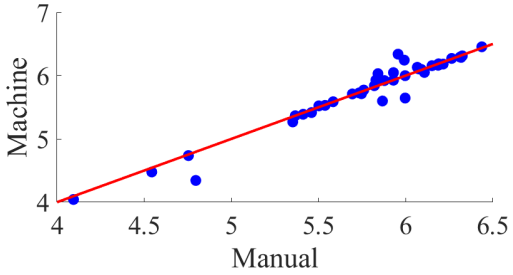
(d) Situations Wanted: *Chicago Tribune*



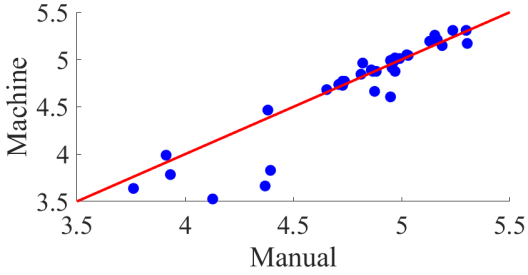
(e) Help Wanted: *Los Angeles Times*



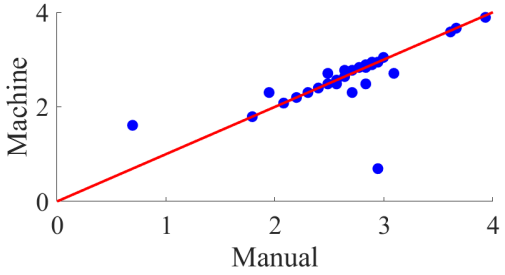
(f) Situations Wanted: *Los Angeles Times*



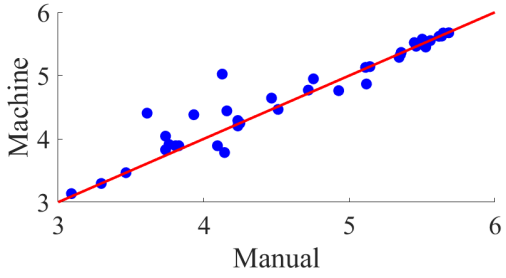
(g) Help Wanted: *Atlanta Constitution*



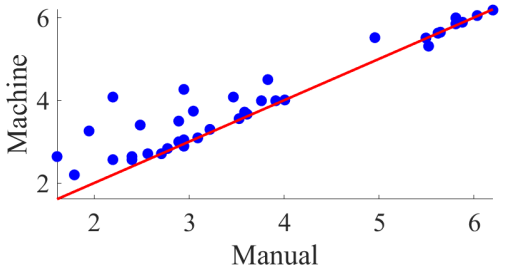
(h) Situations Wanted: *Atlanta Constitution*



(i) Help Wanted: *Washington Post*



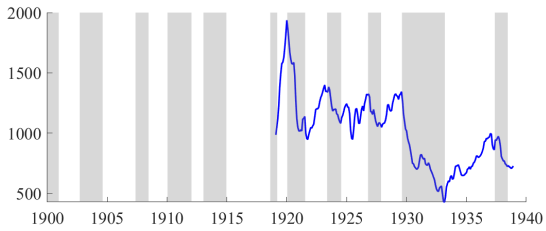
(j) Situations Wanted: *Washington Post*



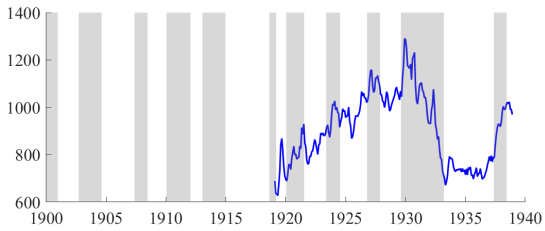
Note: The plots are in log-log scale. Manual counts collected on the second Sunday in May. See the Appendix for details.

Figure 8: City-Level Help- and Situations-Wanted ads, Jan. 1900 to Dec. 1938

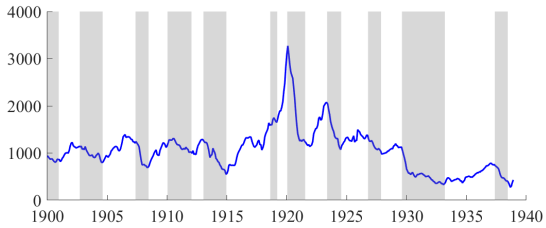
(a) Help Wanted: *New York Times*



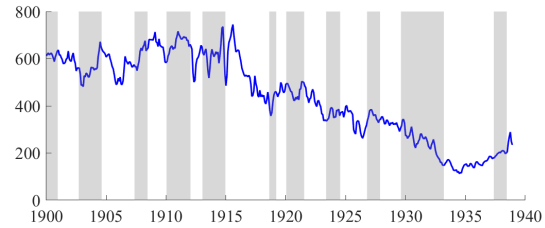
(b) Situations Wanted: *New York Times*



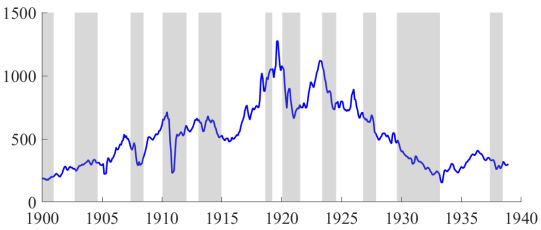
(c) Help Wanted: *Chicago Tribune*



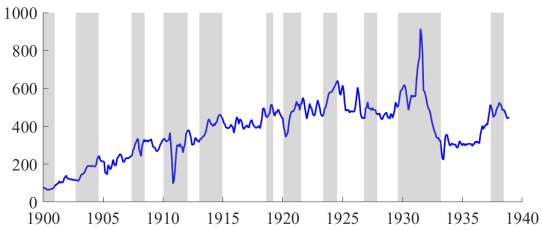
(d) Situations Wanted: *Chicago Tribune*



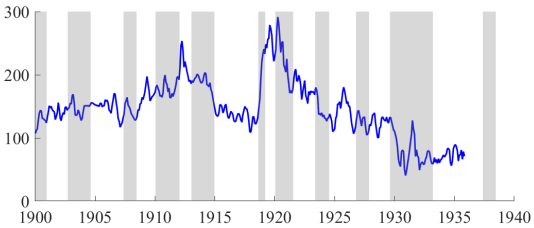
(e) Help Wanted: *Los Angeles Times*



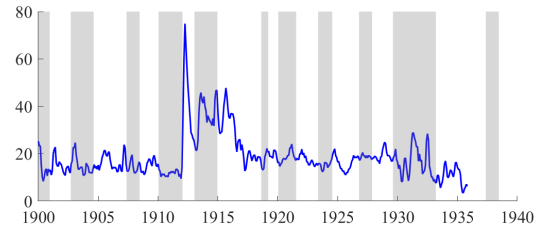
(f) Situations Wanted: *Los Angeles Times*



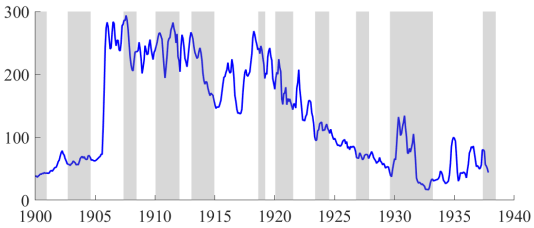
(g) Help Wanted: *Atlanta Constitution*



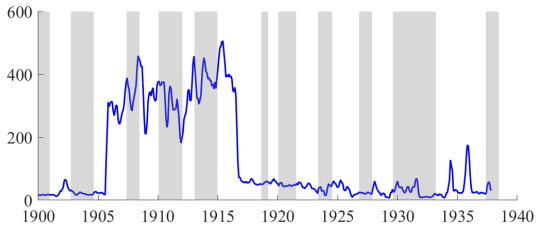
(h) Situations Wanted: *Atlanta Constitution*



(i) Help Wanted: *Washington Post*



(j) Situations Wanted: *Washington Post*



Note: Seasonally adjusted monthly averages. NBER recession dates in grey. *New York Times* ad count reported from 1919 onward due to low circulation in prior years. See the appendix for additional discussion.

and 1938. The number of ads in Chicago and Atlanta is relatively stable until the end of the first World War. A large spike in 1907 occurs in the *Washington Post* because the newspaper made labor advertisements free to encourage broader circulation. Apart from this, the series is also relatively stable until the end of WW1. There is a moderate upward trend in Los Angeles during the first 7 years of the sample, likely due to the rapid growth in the city's population.

Beginning around mid-1918, each city in our sample experiences a very rapid rise in the number of help-wanted ads, peaking around 1920 before declining rapidly during the recession of 1920-21. All five cities experienced a sharp decline to their lowest levels during the years that followed the onset of the Great Depression.

Larger variation of situations-wanted ads appears across cities. The level of situations-wanted ads in Chicago remains fairly stable until around 1916 when a slow-moving downward trend takes hold, while the *Los Angeles Times* displays an upward trend from 1900 similar to the city's trend for help-wanted ads. Situations-wanted ads increase notably at the beginning of the Great Depression in Los Angeles and New York before experiencing a significant decline. The *Washington Post* experiences a large increase over the period the newspaper made labor advertisements free. In all cities, the mid-1930s economic recovery is accompanied by an increase in situations-wanted ads.

Relating to Other City-Level Measures Our data provide city-level measures of economic activity at the monthly frequency. To our knowledge, very few high-frequency measures at the city-level exist over our sample period. In this section, we show that our series provide similar trends as other city-level measures in the literature and discuss how our measures may provide a more accurate assessment of overall economic activity.

Correia *et al.* (2022) construct a city-level measure of economic disruptions for 1917-1922 using information from the "Trade at a Glance" report published as part of the weekly *Bradstreet's — A Journal of Trade, Finance, and Public Economy.* "Trade at a Glance" provides a summary of the current outlook for major cities in the United States for five categories: wholesale trade, retail trade, manufacturing and industry, collections,and crops (excluding winter months). The report tended to use one word to describe the current state of each category (see the Appendix for further details), and Correia *et al.* (2022) use the description to classify the categories manufacturing, wholesale, and retail sectors as poor, fair, or good. We extend their approach by additionally digitizing the records from 1923-1932 to create a measure of the economic activity in each city.

Figure 9 presents the index of the Bradstreet measure for each city (where higher values denote more favorable economic conditions), along with our measures of help wanted.[14] In

[14]Bradstreet does not report information for New York over a large part of the sample and never reports

general, the Bradstreet index (right column) is positively correlated with the help-wanted series (left column); the correlation coefficients range from 0.3 to 0.7. The Bradstreet index is also much more volatile, but this is mainly by construction: The Bradstreet measure only has three possible states (good, fair, and poor) and cannot account for the quantitative degree of change over time. Moreover, Bradstreet's "Trade at a Glance" does not report economic activity for the same cities in every publication, resulting in breaks in our measures for all cities.

In addition to the Bradstreet's publication, from January 1924 to January 1932, the Department of Labor, in cooperation with state and local employment services, produced monthly reports on the number of registered job seekers in their local offices, along with the number of requests for help wanted. These reports provide an alternative measure of help wanted at the city-level. According to Lee (2007), these offices were mainly used by unskilled workers and represented a small (approximately 4%) portion of the labor force. We digitized these records and find that they have similar trends and business cycles as our measures, but the employment agency data tend to be more volatile (see Appendix E.2). In addition, our measures cover a range of skilled and unskilled workers (see the discussion of occupations in Appendix D).

3.2 Building National Indices

Our national help- and situations-wanted indices (HWI and SWI, respectively) are constructed as:

$$Index_t = 100 \times \frac{\sum_{c=1}^{C} AD_{c,t}\omega_{c,t}}{\sum_{c=1}^{C} AD_{c,t=1927:1}\omega_{c,t=1927:1}}, \tag{2}$$

where C is our set of metropolitan areas, $AD_{c,t}$ is the seasonally adjusted monthly ad counts from area c, and $\omega_{c,t}$ is a weighting factor for area c based on its share of the national population in period t, obtained by interpolating between decennial censuses. The national indices are normalized to 100 in January 1927 after scaling by an estimate of the size of the labor force during our sample period. We discuss these indices in detail in the following sections.

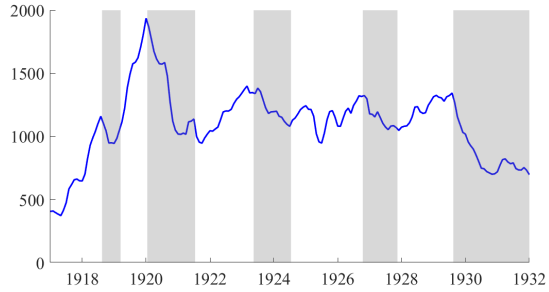
External Validation Our approach to building national help- and situations-wanted indices mirrors the methodology for MetLife's help-wanted index, initiated in 1927, with past issues of print newspapers collected to gather data back to 1919.[15] Doing so allows us to

information for Washington, D.C. Details of the construction of the Bradstreet index are in the Appendix.

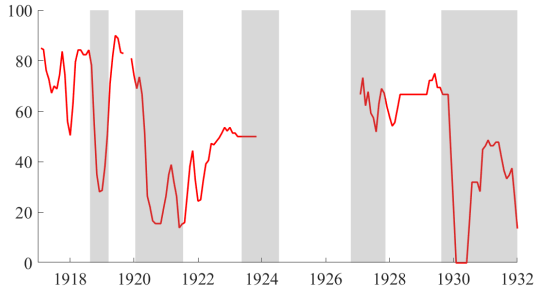
[15]The program, created by William A. Berridge, was the precursor to the Conference Board's Help-Wanted series. One-third of the 100 newspapers in 45 cities in the vacancy index were incorporated in 1927 (Berridge, 1929).

Figure 9: Help-Wanted ads vs. City-Level Bradstreet Indicators, Jan. 1900 to Dec. 1938

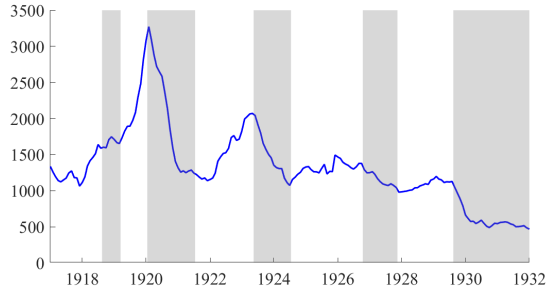
(a) Help Wanted: *New York Times*



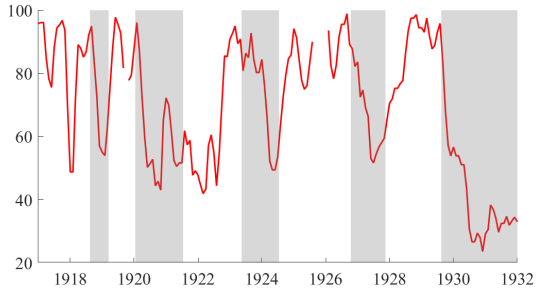
(b) Bradstreet Index: New York City



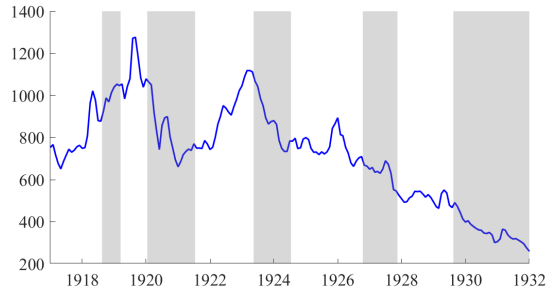
(c) Help Wanted: *Chicago Tribune*



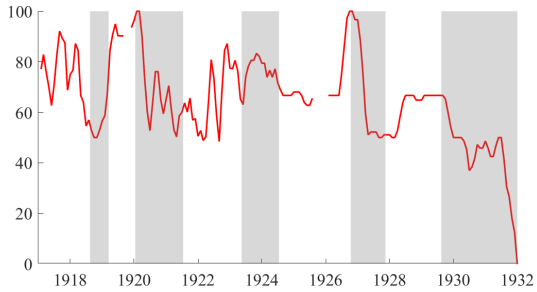
(d) Bradstreet Index: Chicago



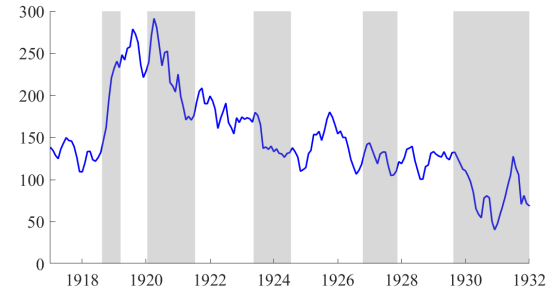
(e) Help Wanted: *Los Angeles Times*



(f) Bradstreet Index: Los Angeles



(g) Help Wanted: *Atlanta Constitution*



(h) Bradstreet Index: Atlanta

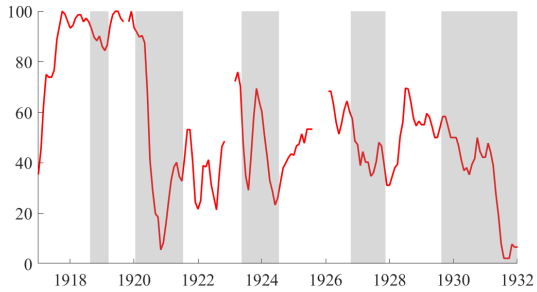
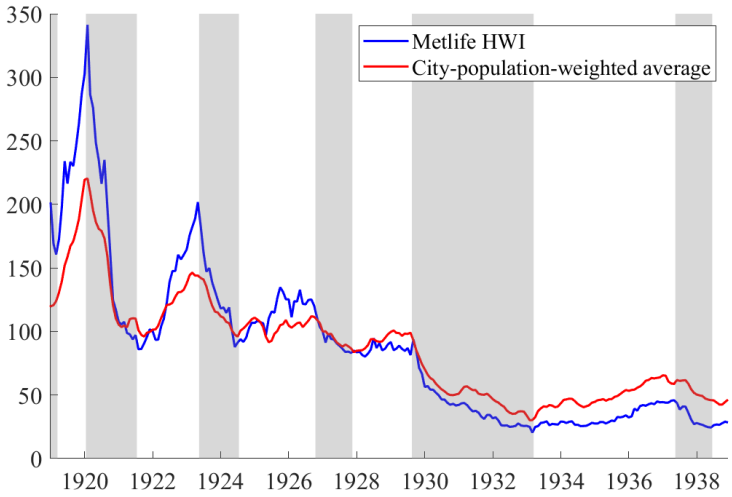


Figure 10: Help Wanted Index: A Comparison with MetLife



Note: The city-weighted HW Index is defined in equation (2). Both series are normalized to Jan. 1927=100 for comparability. NBER recession dates in grey.

validate our national help-wanted index by comparing it to the MetLife index during the period of overlap. Relative to the MetLife program, data limitations restrict us to a subset of newspapers that nevertheless provide significant national coverage at the time.

Our help-wanted index tracks the MetLife series markedly well during the overlapping period with a 0.98 correlation between the two series (see figure 10). Interestingly, our measure is less volatile in the pre-1927 period when there are concerns about the accuracy of the MetLife data (see Zagorsky, 1998). That is, our measure mirrors the Metlife business cycle troughs but not the elevated peaks during expansions, which were previously cited as a reason for concern with the reliability of the MetLife index prior to 1927.

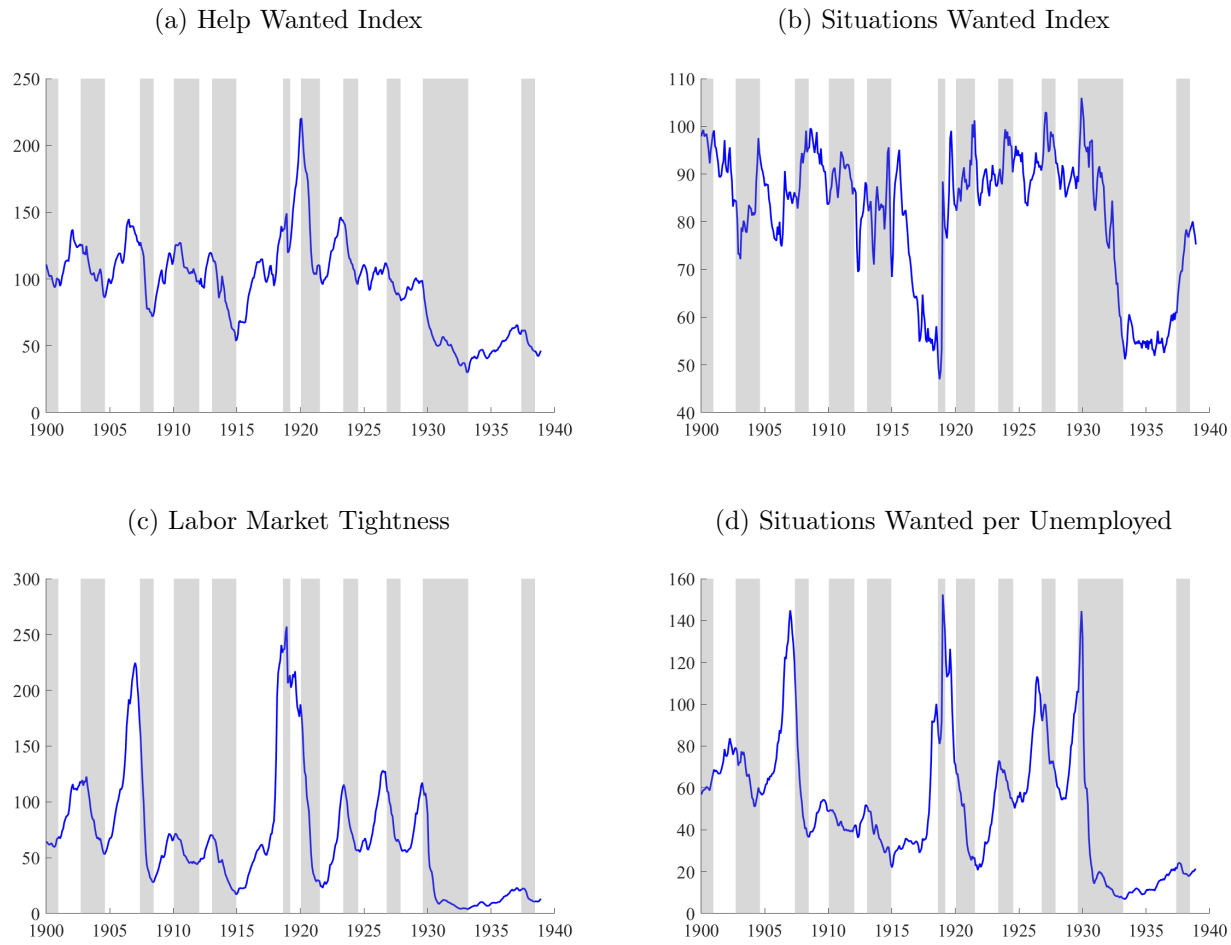
3.3 Cyclicality of Help- and Situations-wanted

In this section, we focus on the cyclicality of our national help- and situations-wanted indices, scaled by estimates of the labor force to account for growth in the workforce. We complement our newly constructed data with estimates of the rate of unemployment, labor productivity and wages in the first half of the 20th century.[16] The details for the latter data are provided in the appendix.

The help-wanted index, shown in panel (a) of Figure 11, tends to decline rapidly at the onset of a recession and to recover with the following expansion. This pattern is similar to the

[16]Data limitations restrict the ability to use city-level measures as comprehensive city-level economic data over a substantial time span is not available. See the discussion in section 3.1.

Figure 11: Aggregate Help- and Situations-Wanted, Tightness and Search Effort



Note: All series are normalized to Jan. 1927=100. NBER recession dates in grey.

behavior of job posting data in the second half of the century. Interestingly, the expansion beginning in March 1919, following the most severe wave of the Spanish flu in late 1918 and the end of World War I, is accompanied by an unprecedented increase in help-wanted advertising. The series peaks on the eve of the recession of 1920, drawing parallels with the spike in job openings during the recovery from the COVID-19 pandemic of 2020.

The situations-wanted index, panel (b) of Figure 11, displays a less distinct cyclical pattern relative to the help-wanted index, as movements over a recession/expansion tend to be less systematic. For instance, during the recession of 1920, which witnesses a sharp decline in the help-wanted index, the situations-wanted index remains largely unchanged or even increases. Even during the most severe economic downturn in our sample period, the first years of the Great Depression, the situations-wanted index remains relatively stable, despite a surge in the unemployment rate from under 3 percent to 18 percent by the fall

of 1931. The significant decline in the situations-wanted index only materializes after that point, when the rate of unemployment further increases to 25 percent. By the end of the 1929 recession, the situations-wanted index declines by nearly 50 percent. The other significant change to the situations-wanted index occurs between 1917 and 1918, which we suspect is due to the decline in Chicago and may not be nationally representative. Future incorporation of additional cities could address this issue.

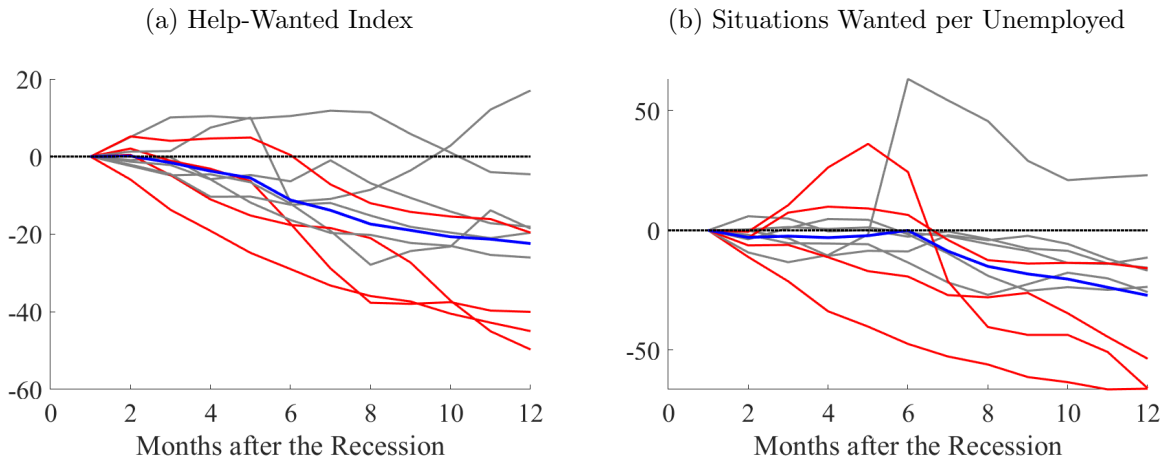
The ratio of job postings to the unemployed, shown in panel (c) of Figure 11, measures the tightness of the labor market from the perspective of employers. Two observations stand out. First, this series is highly pro-cyclical, with a clear pattern of increasing during expansions and falling during recessions. Second, the amplitude is affected by two large movements around 1906 and 1921. The first is largely due to an unusually low estimate of the unemployment rate. The second coincides with an unprecedented increase in help-wanted advertising following the Spanish flu and the low rate of unemployment at end of World War I. These two factors lead to a fourfold increase in the ratio of job postings to unemployment. The ratio drops sharply during the 1920 recession, returning to a similar level as in 1917. The recessions of 1903, 1923 and 1926 are also associated with large swings in the measure of labor market tightness, yet the largest decline to its lowest level in the sample occurs during the Great Depression when it remains at a low level even as the economy recovers in the late 1930s.

Variation in the situations-wanted index can be attributed to two potential sources: (1) changes in the intensity of job seekers' search through the posting of ads and (2) fluctuations in the number of individuals actively seeking employment (given a certain likelihood of placing a situations-wanted ad). We use estimates of unemployment to approximate the magnitude of changes in the number of job seekers, acknowledging that it may not fully capture variation in people actively searching for jobs. Consequently, we define a measure of aggregate job search intensity as the ratio of situations-wanted ads to the number of unemployed individuals and present the resulting series, normalized to 100 in January 1927, in panel (d) of Figure 11.

Job search effort (per unemployed) declines at the start of recessions and increases during periods of economic expansions. It is pro-cyclical and mirrors the business cycle of labor market tightness. Similarly, the amplitude of job search effort is more pronounced after World War I up until the Great Depression compared to the period prior to 1917, though the expansions of the 1920s are accompanied by relatively large increases in situations-wanted advertising. Search effort reaches peaks in 1907, 1919, and at the onset of the Great Depression.

Figure 12 further illustrates these points by plotting the help-wanted index relative to its position at the start of a recession. The blue line indicates the mean response, while the

Figure 12: Dynamics of Help and Situations Wanted Following the Start of a Recession



Note: Blue line indicates the mean response, while red lines denote major recessions, including 1907, 1920, 1929 and 1937. Grey lines indicate all other recessions.

red lines denote "major recessions" as categorized by Bordo and Haubrich (2017) (gray lines indicate the remaining recessions). The four major recessions include: (i) the recession of 1907, following a stock market crash and banking panic; (ii) the recession of 1920, following a sharp tightening by the Federal Reserve; (iii) the Great Depression of 1929, following a major stock market crash and banking panic; and (iv) the recession of 1937, which was the third worst of the twentieth century and often attributed to monetary policy. The figure shows that recessions are similar at their onset, but severe recessions are followed by pronounced declines and slower recoveries in help-and situations-wanted advertising per unemployed.

To further characterize the business-cycle dynamics of the early 20th century labor market, we remove longer-run trends from each times series converted to a quarterly frequency and focus on standard second moments of the resulting cyclical components. To do so, we extract the cyclical component by removing fluctuations at horizons shorter than one year and longer than 10 years through a bandpass filter. The business-cycle moments of interest – standard deviations and contemporaneous cross-correlations – are computed as proportional deviations from trend. Table 2 provides the resulting moments for the full sample period (panel A) and a subperiod ending with the start of the Great Depression in Oct. 1929 (panel B). In addition, panel C reports the same set of moments in the post-World War II sample from 1955 to 2019, ending before the recent pandemic. The details for the latter data are presented in the Appendix.

Unemployment and job postings by firms are very volatile over the business cycle in our full sample, with unemployment about twice as volatile as help wanted. The negative slope of the Beveridge curve can be measured as a negative contemporaneous correlation between

Table 2: Empirical Business Cycle Moments, Quarterly: 1900 to 1938, pre-Great Depression, and post-1955 Samples

Panel A: 1900:I to 1938:IV

		U	V	$S \times U$	θ	S	X	W
Std. dev.		0.44	0.19	0.10	0.63	0.44	0.06	0.02
Corr. matrix	U	1.00	-0.57	0.20	-0.67	-0.74	-0.37	0.11
	V		1.00	-0.08	0.71	0.57	0.39	-0.21
	$S \times U$			1.00	-0.25	-0.06	0.12	0.17
	θ				1.00	0.91	0.41	-0.21
	S					1.00	0.40	-0.21
	X						1.00	0.04

Panel B: 1900:I to 1929:III

		U	V	$S \times U$	θ	S	X	W
Std. dev.		0.37	0.18	0.09	0.54	0.37	0.05	0.02
Corr. matrix	U	1.00	-0.64	0.47	-0.86	-0.81	-0.38	0.12
	V		1.00	-0.11	0.68	0.53	0.40	-0.19
	$S \times U$			1.00	-0.34	-0.14	-0.06	0.21
	θ				1.00	0.93	0.46	-0.18
	S					1.00	0.46	-0.15
	X						1.00	0.21

Panel C: 1955:I to 2019:IV

		U	V	$S \times U$	θ	S	X	W
Std. dev.		0.16	0.17		0.31		0.01	0.01
Corr. matrix	U	1.00	-0.93		-0.93		-0.16	-0.38
	V		1.00		0.97		0.31	0.49
	$S \times U$							
	θ				1.00		0.24	0.49
	S							
	X						1.00	0.56

Notes: *U unemployment rate from updating Petrosky-Nadeau and Zhang (2021); V: 1900 to 1938 - our aggregate Help-Wanted index, 1955 to 2019 - updated data from Petrosky-Nadeau and Zhang (2021); $S \times U$: our aggregate Situations-Wanted Index; V/U: labor tightness; S search effort, which equals the Situations-Wanted index divided by unemployment; X: Labor productivity, 1900 to 1938 - see appendix, 1955 to 2019 - nonfarm output per worker; W: hourly earnings. Standard deviations and contemporaneous cross-correlations computed from proportional deviations from the band-pass filtered trend. See the Appendix for additional details.*

unemployment and vacancies in Table 2. The volatility of situations wanted, while elevated, is about half of that of help wanted and mildly pro-cyclcical. Search effort per unemployment, in contrast, is strongly pro-cyclical, and moves with firm's help-wanted advertising and the tightness of the labor market. The correlations of the cyclical component of wages is quite mild and even in an unexpected direction: wages are negatively correlated with help wanted and positively correlated with labor productivity.

The Great Depression, and its large departures from normal business-cycle ups and downs, could distort the characterization of the cyclical dynamics of labor. Panel B of Table 2 addresses this issue by presenting the same set of moments on a sample spanning 1900 to 1929. The patterns are markedly similar to the pre-Depression era, though the volatility of unemployment, labor tightness and search effort are lower.

For comparison, panel C of Table 2 presents the moments of variables in the post-World War II period (1955-2019). Consistent with the literature, the volatility of variables is considerably lower in the modern era. Vacancies and labor tightness are much more positively correlated, but in general the variables display the same correlation patterns as the early twentieth century samples.

Finally, panel (a) of Figure 13 shows the U.S. Beveridge curve between January 1900 and December 1938 by plotting the help-wanted index relative to the unemployment rate. Consistent with more recent sample periods, the scatter points display a downward-sloping and convex relationship between the variables. The high unemployment rates and low vacancy rates of the Great Depression (red dots) describe a particularly flat portion of the Beveridge curve, as noted in Petrosky-Nadeau and Zhang (2021). The figure also distinguishes the period from May 1919 to December 1922 (teal dots) when the Beveridge curve appears to have shifted outward for a few years before returning to its original position.

Panel (b) of Figure 13 plots the log of search effort against the log of labor market tightness. Throughout the sample, there is a strong and positive relationship between the two variables. In particular, this relation did not change when the Beveridge curve shifted between May 1919 and December 1922, nor did it change during the Great Depression.

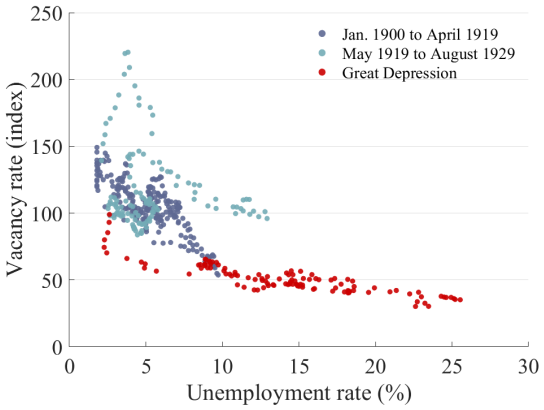
4 Help- and Situations-wanted Series by Gender

We further explore how help- and situations-wanted ads evolved by gender over our sample. Figure 1 shows that labor ads were segregated by gender, allowing us to extract this information using section headers.

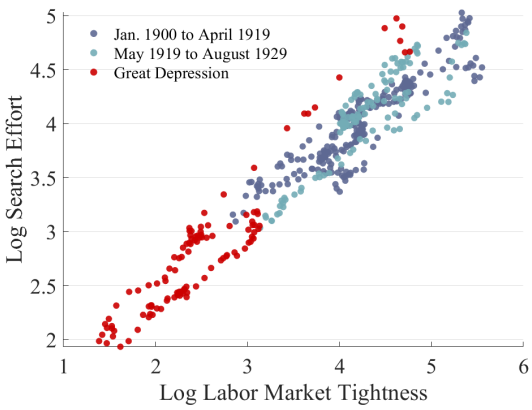
We apply state-of-the-art OCR tools to extract gender information from labor ads. As highlighted in Section 2.2, the format of newspaper ad pages restricts us to use morphological operations on image files, instead of the conventional approach of OCR on text files, in order

Figure 13: Beveridge and Search Effort-Tightness Curves

(a) Beveridge Curve



(b) Search Effort and Labor Market Tightness



to identify and split ads. However, once classified pages are split into individual ads, OCR tools can be used to convert individual ad images to text files. We have tested several OCR tools, including Amazon's Textract, Google's Cloud Vision, and the open source Tesseract. Similarly to Correia and Luck (2023), we find that Google Cloud Vision provides the best results for our application.

One straightforward way to extract the text is to convert each image to a text file. However, the cost of this approach is prohibitively high, as the state-of-the-art OCR tools charge by the number of image files, rather than by the size of files, and we have hundreds of ads every day for each newspaper. Instead, we first store the bounding box of each ad when splitting classified pages into ads. Then, we apply OCR to each newspaper page that contains hundreds of classified ads. The OCR output includes bounding boxes for all the letters and words in an image. Any word that overlaps with the bounding box of an ad is classified as text of that ad. To characterize gender, we focus on the text files identified as headers (see Section 2.2.2) and reclassify the header as related to men or women, depending on whether words related to gender appear in the header (e.g, male, males, men). Table 3 summarizes the distribution of errors with our method, relative to manual counts, by gender and newspaper (errors are calculated using equation 1).

Figure 14 highlights the count comparisons between male and female labor ads for both help- and situations-wanted categories across cities. The counts of male labor ads are generally higher than those of female labor ads. For example, the average number of ads for help wanted is 640 for men and 380 for women throughout the data sample for the *Chicago Tribune*. The difference is even more notable for the *Atlanta Constitution*, as help-wanted ads for women are quite low.

In terms of cyclical fluctuations, female vacancies often move together with male vacancies

Table 3: Accuracy of Template Matching Approach: Comparisons with Manual Counts across Genders

	Number of ads	1st Tertile	Median	3rd Tertile
New York Times				
Help-wanted Male	12,993	0.03	0.05	0.09
Help-wanted Female	9,269	0.01	0.03	0.08
Situations-wanted Male	11,715	0.01	0.05	0.15
Situations-wanted Female	11,571	0.01	0.03	0.11
Chicago Tribune				
Help-wanted Male	30,050	0.02	0.04	0.09
Help-wanted Female	17,621	0.01	0.02	0.08
Situations-wanted Male	9,835	0	0.01	0.06
Situations-wanted Female	8,716	0	0.02	0.13
Los Angeles Times				
Help-wanted Male	10,099	0.03	0.13	0.30
Help-wanted Female	7,553	0.04	0.11	0.34
Situations-wanted Male	5,970	0.02	0.06	0.10
Situations-wanted Female	7,253	0.04	0.08	0.25
Atlanta Constitution				
Help-wanted Male	1,972	0.03	0.06	0.13
Help-wanted Female	768	0	0.02	0.08
Situations-wanted Male	340	0	0	0.1
Situations-wanted Female	241	0	0	0

Notes. Errors between machine and manual counts computed according to equation (1). Manual counts collected on the second Sunday in May. For *Atlanta Constitution*, we report the statistics between 1900 and 1929.

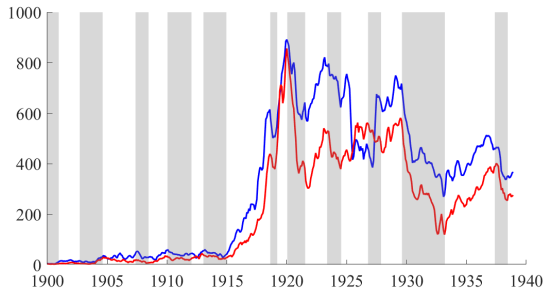
but have higher volatility. For instance, during the 1920 recession, the number of help-wanted ads posted in the *New York Times* falls from the peak of above 800 to 600 for men, but the decline is more significant for women, decreasing from 800 to less than 400. However, help-wanted ads for women increased more rapidly than those for men during the Spanish flu and World War I according to the *Los Angeles Times*. Similarly, the number of situations-wanted ads increases more notably for women during the Great Depression according to the *New York Times* and the *Los Angeles Times*. Table 4 highlights the standard deviations by gender and between cities, showing that there is generally higher volatility for female advertisements.

5 Conclusion

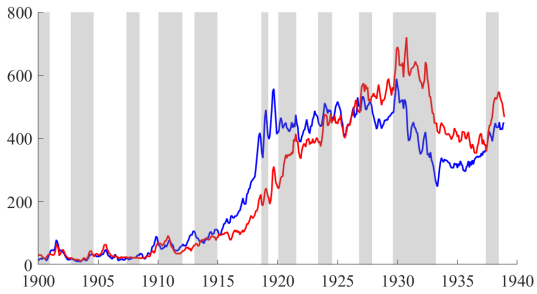
This paper provides new city-level and national measures of search effort based on the classified advertisements by job seekers — the *situations-wanted ads* — in early twentieth century

Figure 14: City-Level Help- and Situations-Wanted ads by Gender

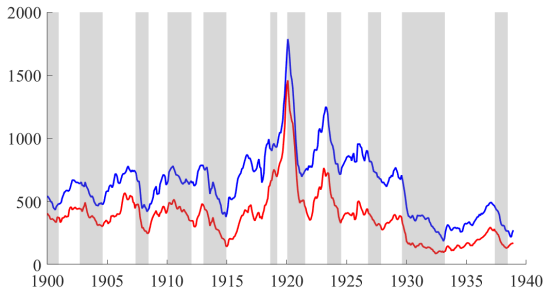
(a) Help Wanted: *New York Times*



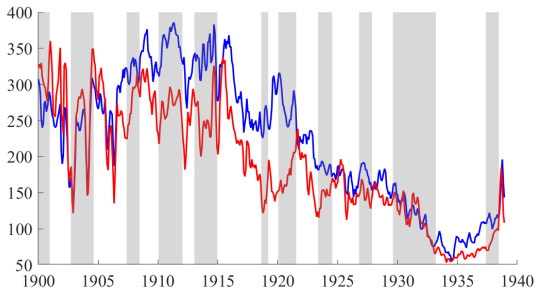
(b) Situations Wanted: *New York Times*



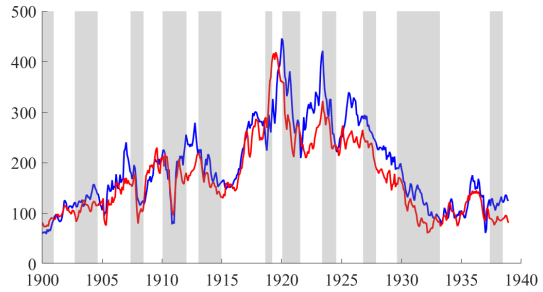
(c) Help Wanted: *Chicago Tribune*



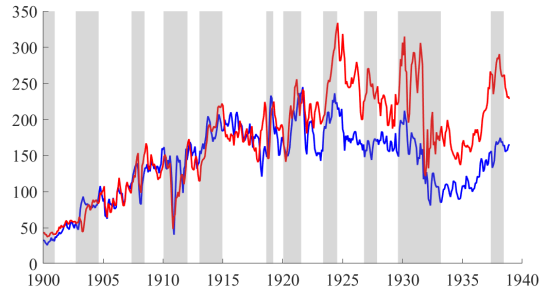
(d) Situations Wanted: *Chicago Tribune*



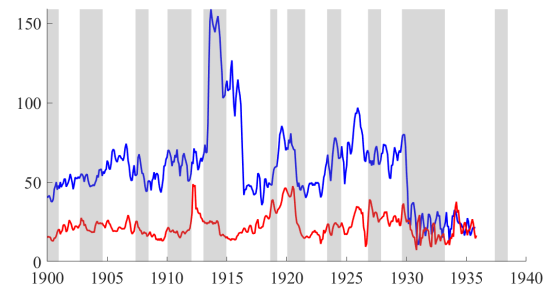
(e) Help Wanted: *Los Angeles Times*



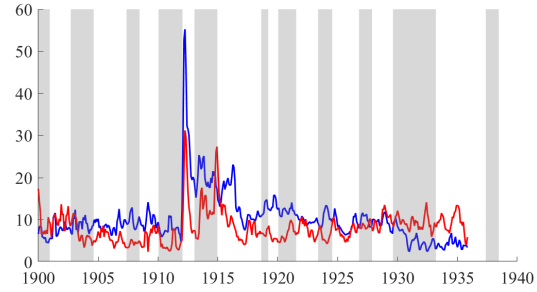
(f) Situations Wanted: *Los Angeles Times*



(g) Help Wanted: *Atlanta Constitution*



(h) Situations Wanted: *Atlanta Constitution*



Note: Seasonally adjusted monthly averages. Red lines are for female labor ads, while blue lines are for male labor ads. NBER recession dates in grey.

Table 4: Standard deviations by gender and across cities

	NYC	**Chicago**	**LA**	**Atlanta**
Help Wanted				
Men	0.18	0.22	0.18	0.35
Women	0.26	0.35	0.22	0.31
Situation Wanted				
Men	0.10	0.13	0.15	0.40
Women	0.11	0.18	0.18	0.39

U.S. newspapers along side help-wanted advertising placed by employers. To do so, we first employ a novel technique using morphological operations to identify labor advertisements and construct monthly help- and situations-wanted indices from 1900 to 1938. We validate our methodology using extensive manual checks, as well as comparing our series relative to national indicators of help-wanted ads available for sub-periods within our sample.

The job advertisements from both firms and job seekers vary extensively over our sample period. Posts by employers decline sharply at the beginning of a recession, and there is a strong negative correlation between unemployment and job openings, i.e. the Beveridge curve. The total of job-seeking posts show less of a cyclical pattern, tending to remain stable at the onset of a recession when the number of job seekers is likely to rise. However, job search effort –measured as the ratio of our situations-wanted ads to the unemployed – is strongly pro-cyclical: it declines at a recession's onset, remains low during a recession, and rises during an expansion. In addition, labor market tightness – measured as the ratio of our help-wanted ads to the unemployed – is highly positively correlated with job search effort. Trends across gender are similar, while posts for/by women are more volatile than those for/by men.

Our work offers new high-frequency measures of economic activity in the early twentieth century. Moreover, to the best of our knowledge, our methodology for systematically categorizing classified advertisements, relying extensively on morphological operations, is novel in application to economics. Future work can further this approach by complementing with textual analysis to further characterize the evolution in the early twentieth century of occupations, demographics, and potentially asked-for wages. Moreover, our method can be applied to more newspapers and other dimensions of classified ads, for example, to systematically categorize the real estate market.

References

Ahn, H. J. and Shao, L. (2017). *Precautionary On-the-Job Search over the Business Cycle*. Finance and Economics Discussion Series 2017-025, Board of Governors of the Federal Reserve System (U.S.).

Atalay, E., Phongthiengtham, P., Sotelo, S. and Tannenbaum, D. (2020). The evolution of work in the united states. *American Economic Journal: Applied Economics*, **12** (2), 1–34.

Berridge, W. A. (1929). Labor turnover. *Monthly Labor Review*, (29), 62–65.

Bezdan, T. and Bacanin, N. (2019). Convolutional neural network layers and architectures. pp. 445–451.

Bordo, M. D. and Haubrich, J. G. (2017). Deep Recessions, Fast Recoveries, And Financial Crises: Evidence From The American Record. *Economic Inquiry*, **55** (1), 527–541.

Bureau of Labor Statistics (1925). *Union Scale of Wages and Hours of Labor*. Bulletin of the United States Bureau of Labor Statistics 404.

Bureau of Labor Statistics (1926). Prices and cost of living. *Monthly Labor Review*, **22** (4), 32–59.

Chetty, R. (2008). Moral hazard versus liquidity and optimal unemployment insurance. *Journal of Political Economy*, **116** (2), 173–234.

Correia, S. and Luck, S. (2023). Digitizing historical balance sheet data: A practitioner's guide. *Explorations in Economic History*, **87**, 101475.

—, — and Verner, E. (2022). Pandemics Depress the Economy, Public Health Interventions Do Not: Evidence from the 1918 Flu. *The Journal of Economic History*, **82** (4), 917–957.

DeLoach, S. and Kurt, M. (2013). Discouraging Workers: Estimating the Impacts of Macroeconomic Shocks on the Search Intensity of the Unemployed. *Journal of Labor Research*, **34** (4), 433–454.

DeVaro, J. and Gurtler, O. (2018). Advertising and Labor Market Matching: A Tour through the Times. *Journal of Labor Economics*, **36** (1), 253–307.

Faberman, R. J. and Kudlyak, M. (2019). The Intensity of Job Search and Search Duration. *American Economic Journal: Macroeconomics*, **11** (3), 327–357.

—, Mueller, A. I., Şahin, A. and Topa, G. (2022). Job search behavior among the employed and non-employed. *Econometrica*, **90** (4), 1743–1779.

Goldin, C. D. (1990). *Understanding the Gender Gap: An Economic History of American Women*. New York: Oxford University Press.

Gomme, P. and Lkhagvasuren, D. (2015). Worker search effort as an amplification mechanism. *Journal of Monetary Economics*, **75**, 106–122.

Heald, P. J. (2014). The public domain. *Handbook on the Economics of Copyright*, **87**, 93–104.

Laurent, J. D. (2020). *Great Expectations: Women's Help Wanted Ads in Kansas City, 1920-1936*. Master's thesis, University of Kansas.

Lee, W. (2007). *Private Deception and the Rise of Public Employment Offices in the United States, 1890 - 1930*. NBER Working Papers 13695, National Bureau of Economic Research, Inc.

— (2016). Slack and Slacker: Job Seekers, Job Vacancies, and Matching Functions in the U.S. Labor Market during the Roaring Twenties and the Great Contraction, 1924-1932. *The Journal of Economic History*, **76** (3), 840–873.

Marinescu, I. and Skandalis, D. (2021). Unemployment insurance and job search behavior. *The Quarterly Journal of Economics*, **136** (2), 887–931.

Mukoyama, T., Patterson, C. and Şahin, A. (2018). Job search behavior over the business cycle. *American Economic Journal: Macroeconomics*, **10** (1), 190–215.

NewsEye (2020). Newseye: A digital investigator for historical newspapers. https://www.newseye.eu/, accessed: 2023-08-16.

Petrosky-Nadeau, N. and Zhang, L. (2021). Unemployment crises. *Journal of Monetary Economics*, **117**, 335–353.

Simon, C. J. (2001). The Supply Price Of Labor During The Great Depression. *The Journal of Economic History*, **61** (4), 877–903.

Snyder, T. D. (ed.) (1993). *120 Years of American Education: A Statistical Portrait*. Washington, D.C.: U.S. Dept. of Education, Office of Educational Research and Improvement, National Center for Education Statistics.

Zagorsky, J. L. (1998). Job Vacancies In The United States: 1923 To 1994. *The Review of Economics and Statistics*, **80** (2), 338–345.

Appendix

A Extracting Individual Advertisements: Additional Details and Robustness Checks

In this section, we provide additional details and consider robustness checks for the template matching method in Section 2. First, we explain how newspaper pages of classified ads are split into ads through image processing. Second, we provide more details on the template matching method, including the iterative process that we use to determine the preset thresholds. We also discuss the process of identifying multicolumn headers in the *New York Times*. Next, we provide additional manual counts by randomly selecting two pages of each newspaper for each month and comparing them to the results of our template matching procedure. Finally, we compare our process with an alternative method based on machine learning using a Convolutional Neural Network.

A.1 Image Processing

Classified ads use a strict column format with ads flowing from top to bottom and left to right, with vertical lines separating columns and varying types of horizontal lines separating individual ads. Horizontal dividers can be a single or double solid line or a wavy line. Columns typically start with a categorical header that features larger font. In some cases, headers can span multiple columns. We split labor ads through image segmentation in three steps: 1) image alignment; 2) splitting the page into columns (vertical); and 3) splitting the columns into ads (horizontal) as shown in Figure 3.

Image Alignment During the digitization process, text documents are often not oriented properly while scanned. Ensuring column/ad dividers are as vertical/horizontal as possible improves accuracy. While there are a number of transformations that could be used to straighten the image, a simple rotation is sufficient to provide good results in our case. We use the Separator Net Post Processor from NewsEye (2020) to identify any skew or slant to images and then use the warpAffine method from OpenCV, an open source computer vision library, to rotate images accordingly. Specifically, NewsEye's Separator Net Post Processor identifies poly-lines delineating the separation of columns and then applies a linear regression for each poly-line to identify the slope. The median angle is used to rotate the image using OpenCV.

Splitting Pages into Columns We use NewsEye's Separator Net Post Processor to split newspaper pages into columns. The aligned poly-lines are grouped into clusters based on

their horizontal coordinates and smoothed using a Savitzky-Golay filter. Discontinuous line segments are connected by straight lines, and vertical lines are added to the top and bottom of the poly-lines to span the entire image from top to bottom. After all the column dividers are identified, the page is split into separate images.

Occasionally, column dividers are faint or non-existent, resulting in multiple columns being grouped together. At the other extreme, noise in the scanned image can result in over splitting of the image into too many columns. To alleviate these issues, we calculate a typical column width by using the average difference of the modes of each poly-line's horizontal coordinates.[17] Columns that are close to a multiple of the typical column width are divided, while dividers that create too narrow columns are removed.

Splitting Columns into Ads Due to the density of text, NewsEye's Separator Net Post Processor does not adequately separate individual ads in columns. Instead, the black horizontal lines that separate ads are identified by using a combination of tools from different python libraries, including OpenCV, Scikit-Learn, Scipy, and Shapely. We follow three steps:

1. Invert image colors: as shown in Figure A1b, the image colors are inverted, from black to white or from white to black. In computer graphics, white and black are represented in monochrome images by values of 255 and 0, respectively. Shades of gray are represented by a value between 0 and 255. Inverting the image colors simplifies the separating process by giving the foreground values greater than zero.[18]

2. Remove text: for each column, the text is identified by grouping together neighboring white pixels. It is then removed based on the aspect ratio of their bounding box, leaving just the horizontal lines behind, as shown in Figure A1c.

3. Identify ad dividers: the mean of each row of pixels is normalized by dividing by 255, the value of a white pixel. The value should be close to one for a horizontal line and close to zero everywhere else, see the blue line in Figure A1d. A Hanning window is applied to the array of values to filter out any noise in the underlying data, see the orange line in Figure A1d.

   Ad dividers can be identified by peaks in the array: if the pixel value is higher than a certain prominence threshold, we classify it as a line. However, a fixed threshold can not be used, as the image quality varies across newspaper pages. Instead, a threshold is calculated for each column. First, all peak values within a column are binned into

[17] We use the Density-Based Spacial Clustering of Applications with Noise (DBSCAN) from Scikit-Learn to remove any outliers from the differences.

[18] Images can have varying illumination across the page. Following the image processing literature, we use an adaptive threshold to improve the outcome.

Figure A1: Identifying ad dividers

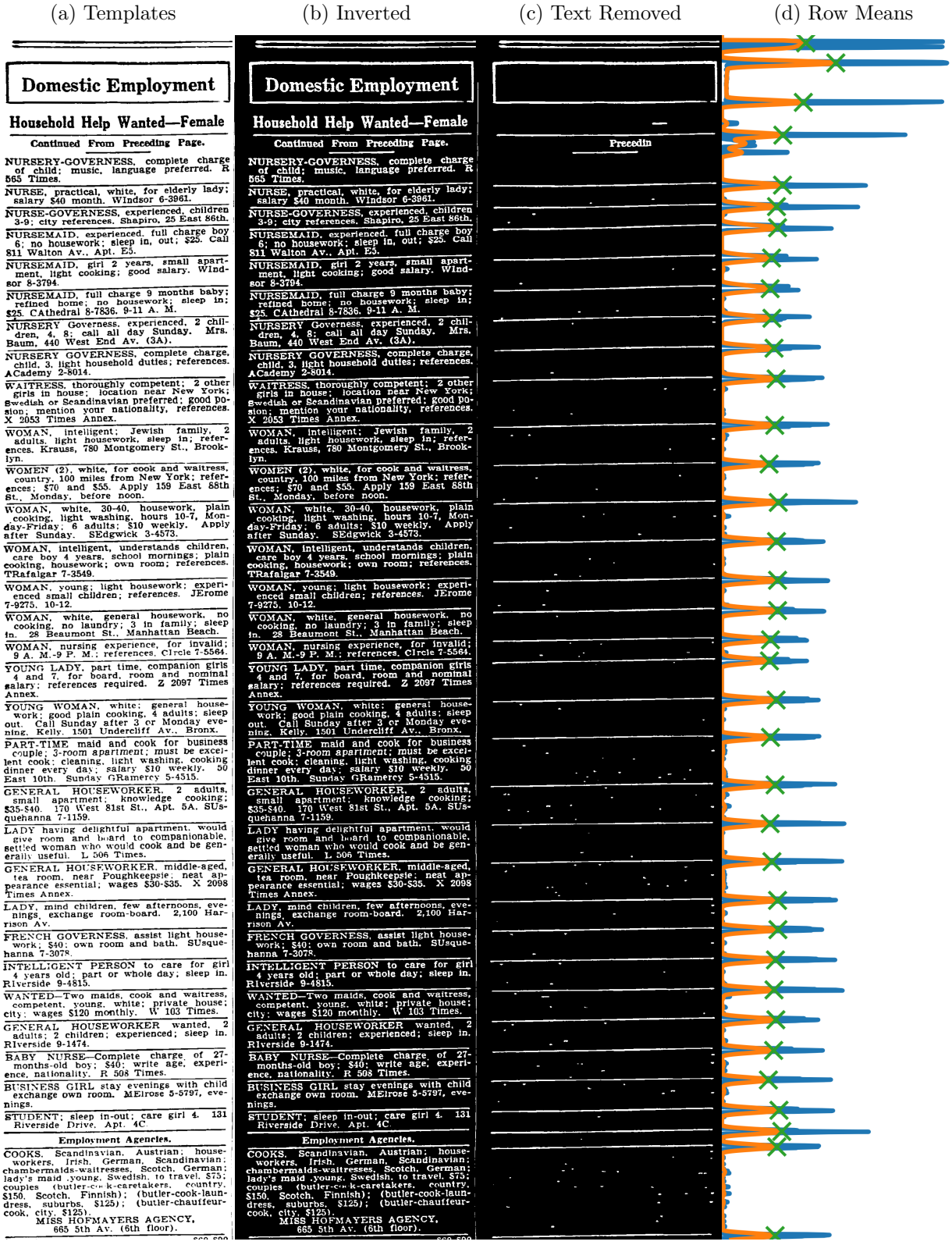
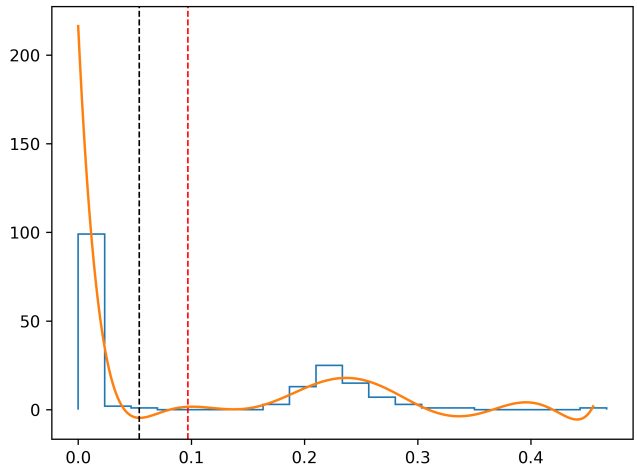


Figure A2: Prominence Threshold Calculation



a histogram, see the blue line in Figure A2. Then, a high-order polynomial is fit to the histogram, as shown by the orange line. The first local minimum is used as the prominence threshold, as shown by the black dashed line. If no local minimum is found, then the mean of all peak values is used as the threshold, see the red dashed line in Figure A2. Finally, peaks with a prominence above this threshold are used to divide the column into individual ads, as shown by green cross signs in Figure A1d.

A.2 Additional Details on Template Matching

Newspaper publishers changed the wording and typesetting for section headers over time. Table A1 documents the dates when either change occurred for our five newspapers between 1900 and 1938. The frequency varied considerably across newspapers. While the *Chicago Tribune* changed section headers only twice during the span of 38 years, other newspapers made more frequent changes. To account for these issues, we created 131 templates for the *Los Angles Times*, 37 for the *Chicago Tribune*, 178 for the *New York Times*, XX for the *Atlanta Constitution*, and XX for the *Washington Post*.

We use a 2-dimensional convolution to match individual ad images with templates. Each template is compared against every ad using OpenCV's matchTemplate method. The template is moved across the targeted image, which returns a matching score between 0 and 1. A blank border is added to the outside of each ad to ensure the best possible match for ads that may not have been segmented properly. Figure A3 highlights that the matching score can vary greatly depending on the template size. Therefore, we use a range of template sizes and rotations and choose the maximum matching score for each pair of ad and template.

Table A1: Changes of Ad Headers Over Time and Across Newspapers

Los Angles Times	*Chicago Tribune*	*New York Times*	*Atlanta Constitution*	*Washington Post*
January 1, 1900	January 1, 1900	January 1, 1900	January 1, 1900	January 1, 1900
March 22, 1905	March 26, 1930	June 16, 1903	January 31, 1903	June 1, 1903
December 4, 1909		February 20, 1904	October 17, 1904	April 15, 1906
October 1, 1910		May 22, 1907	March 26, 1907	April 11, 1908
February 22, 1911		May 4, 1908	January 28, 1909	April 6, 1915
February 22, 1918		November 19, 1908	January 18, 1910	April 3, 1926
April 8, 1920		March 20, 1909	June 15, 1913	March 25, 1930
February 11, 1922		October 7, 1909	November 24, 1913	September, 29, 1931
January 14, 1931		May 9, 1910	September 27, 1918	February 24, 1936
May 29, 1931		August 22, 1910	January 16, 1921	
September 3, 1931		April 7, 1911	April 11, 1921	
October 22, 1932		April 11, 1911	September 19, 1923	
February 13, 1938		June 9, 1912	March 20, 1924	
		August 10, 1912	January 1, 1930	
		May 8, 1915		
		May 14, 1915		
		October 4, 1915		
		February 10, 1916		
		October 12, 1916		
		March 29, 1926		

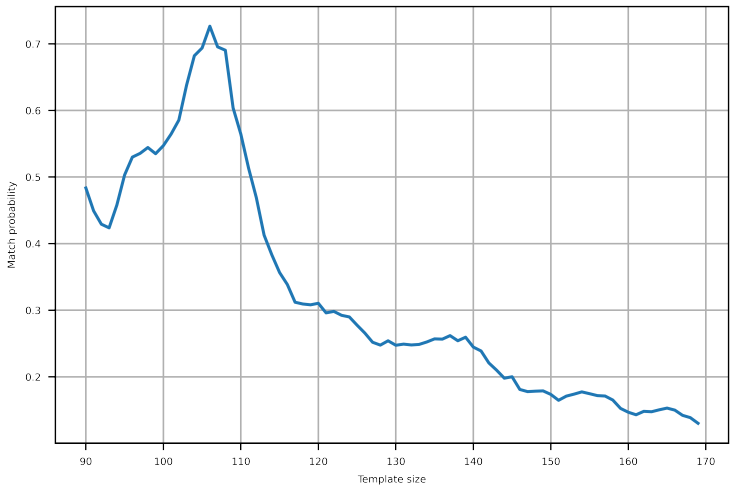
Finally, preset thresholds are applied to template match scores to identify and categorize headers. The thresholds are obtained through an iterative process.

1. An initial threshold of 0.7 is used for narrow templates and 0.6 for wider templates. The initial thresholds are determined based on experience and a small sample analysis.

2. For each period shown in Table A1, we run the template matching procedure on all image files within that time period and generate a maximum match score for each pair of image and template.

   The match scores usually have a bimodal distribution with two peaks. For example, the left panel of Figure A4 shows the match score distribution for the template of "situations" for the *New York Times* between March 29, 1926 and December 31, 1938. The horizontal axis shows the range of match scores between 0 and 1, and the vertical axis shows the number of ads in logarithm terms. The bimodal distribution has one peak around the match score of 0.4 and another peak around the match score above 0.8. The match score value at the trough—0.66 in this case—is chosen as the new threshold for the template of situations. Similarly, the right panel of figure A4 shows the match score distribution for the template of "wanted" for the *New York Times* in the same period. The value at the trough of this bimodal distribution is 0.73, which is used to update the threshold for the template of wanted.

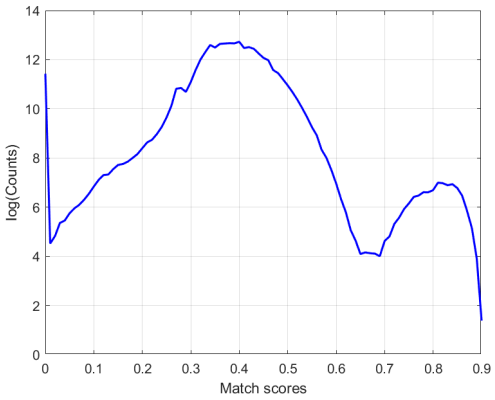
Figure A3: Template size and matching score



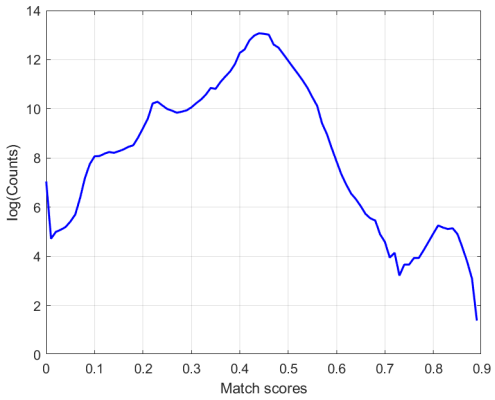
3. We repeat the previous step by using the updated thresholds. Finally, we also fine-tune the thresholds by comparing template matching results with manual counts.

Figure A4: Match score distributions for selected templates: *New York Times* between March 29, 1926 and December 31, 1938

(a) Situations



(b) Wanted



A.3 Multi-Column Header Identification

The *New York Times* employs a more complex layout with categorical headers spanning multiple columns, see Figure 6. These multi-column headers varied in size and kerning (i.e., the spacing between individual characters), and using our template matching method alone is less effective. Instead, a machine learning object detection approach is used for the identification and categorization of these headers. Using 6,060 instances of multi-column

headers tagged in 3,797 different images, a Convolutional Neural Network (CNN) model is trained to detect and classify regions of images into four different classes: background, help-wanted headers, situations-wanted headers, and other headers.

This model is applied to each image before it is split into columns and ads, and the classification and bounding box of the multi-column headers is stored and compared against the bounding box of each ad. If the bounding box of the multi-column header overlaps the bounding box of an ad by more than 50%, the multi-column header class is used to categorize the ad as a header of the same class. Since both single and multi-column headers are used together, this method is used in conjunction with the template matching method previously described.

A.4 Additional Manual Counts

In addition to manually counting ads on the second Sunday in May between 1900 and 1938, we conduct alternative manual checks by randomly picking two pages of each newspaper for every month for the *New York Times* between 1900 and 1938 as well as for the *Los Angeles Times* and the *Chicago Tribune* between 1925 and 1938. We manually count the numbers of the help- and the situations-wanted ads. Table A2 reports that the alternative manual checks deliver similar results in terms of the method's precision.

Table A2: Accuracy of Template Matching Approach Based on Alternative Manual Counts

	1st Tertile	Median	3rd Tertile
New York Times			
Help-wanted E	0.00	0.03	0.14
Situations-wanted E	0.00	0.02	0.12
Chicago Tribune			
Help-wanted E	0.01	0.04	0.11
Situations-wanted E	0.01	0.02	0.10
Los Angeles Times			
Help-wanted E	0.02	0.05	0.10
Situations-wanted E	0.01	0.02	0.05

Notes. Errors between machine and manual counts computed according to equation (1). Manual counts collected from two pages of each newspaper per month.

A.5 Machine Learning Method to Identify Headers

Besides the template matching method, we also explored a machine learning method by using a CNN model to identify classified ad headers. We applied the approach to the *Los Angeles Times*, on which the template matching method has relatively poor performance.

CNN is an algorithm that imitates the way humans learn and typically consists of three parts: convolutional layer, pooling layer, and fully connected layer (see Bezdan and Bacanin, 2019). We build a CNN on ResNet50 architecture, a deep residual network containing 50 layers.[19] The implementation of the procedure involves the following steps:

1. We collect more than 9,000 images between 1925 and 1938 and then manually classify those images to the following categories: help-wanted headers (female), help-wanted headers (male), situations-wanted headers (female), situations-wanted headers (male), non-labor headers, as well as non-header images. Table A3 provides a summary of the image files. We split the period into two sub-periods, 1925-30 and 1931-38, since the wordings for help- and situations-wanted headers switched in January 1931 as shown in Section 4. These classified files are split into a training sample and a validation sample on an 80-20% basis.

Table A3: Summary statistics of data input for the machine learning model.

Period	Category	Number of images
1925-1930	Total	4446
	Help wanted	1103
	Situations wanted	693
	Non-labor	1460
	Other	1182
1931-1938	Total	4667
	Help wanted	730
	Situations wanted	879
	Non-labor	1666
	Other	1202

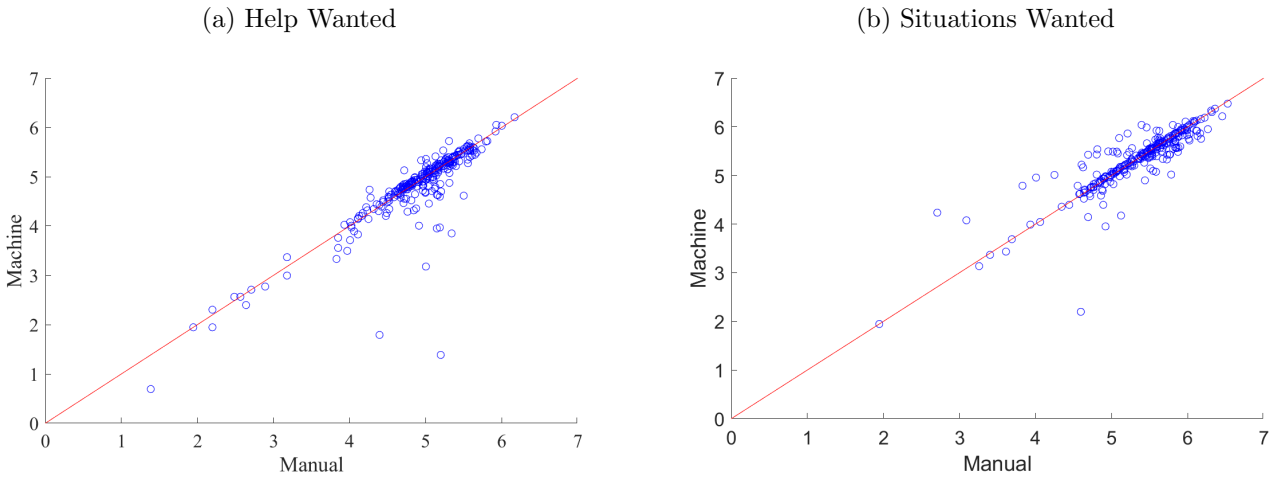
2. Using the training sample, the model learns the features of each image category based on a neural network. After the training, we check the model's performance using the validation sample, and the accuracy rate is 98% for the period of 1925-30 and 97% for the period of 1931-38.

3. Finally, we apply the trained CNN model on our data sample. We count the number of help- and situations-wanted ads based on the classification of each image file.

[19] Our codes are written based on an open-sourced deep learning framework, Pytorch, which was developed by Facebook.

Figure A5 shows the comparisons between the machine learning method against the manual counts. The scatter dots are clustered on the 45 degree line, suggesting that the machine learning method does a good job in identifying headers in general.

Figure A5: Comparisons between machine learning and manual counts in the *Los Angeles Times*.



B Seasonal adjustment

Figure B1 plots a comparison of our monthly counts for help- and situations-wanted ads when the series are and are not seasonally adjusted across newspapers. In general, seasonal adjustment smooths each series and, in turn, reduces the volatility of each series. The situations-wanted series from the *New York Times* has stronger seasonal patterns than other series.

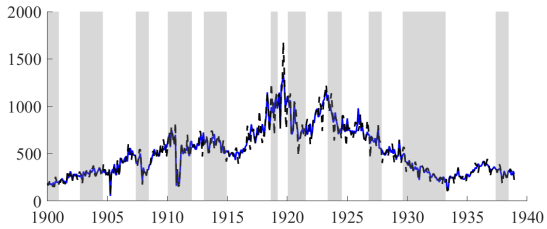
C Classified Ad Rates

The cost for posting labor ads varied over time and across newspapers. Moreover, the costs varied depending on the number of postings purchased. Many newspapers regularly printed the current rates in their Sunday publications. We manually collected the posting costs for the *New York Times* and the *Los Angeles Times*. Information for the *Chicago Tribune* and *Atlanta Constitution* were not printed in the newspapers.

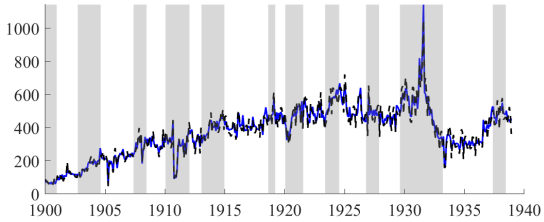
Figure 2 presents the nominal and real costs over time of posting a help-wanted or situations-wanted ad in the *New York Times* and *Los Angeles Times*. The cost decreased in the number of postings purchased. For instance, in the *New York Times*, the cost for posting a help-wanted ad one time in 1901 was 10 cents, 24 cents for three times, and 42 cents for seven times in a row.

Figure B1: Comparison of seasonally adjusted and non-seasonally adjusted measures

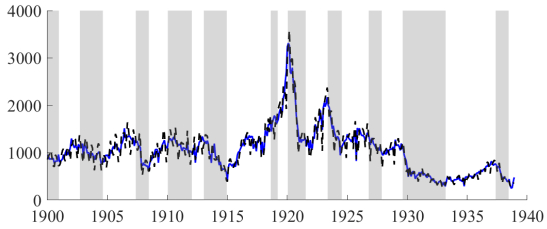
(a) Help Wanted: *Los Angeles Times*



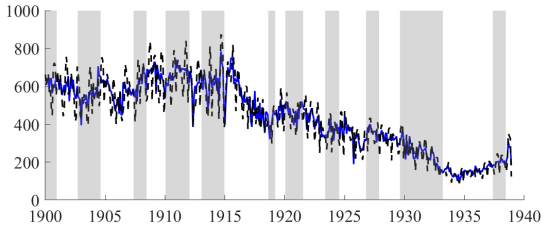
(b) Situation Wanted: *Los Angeles Times*



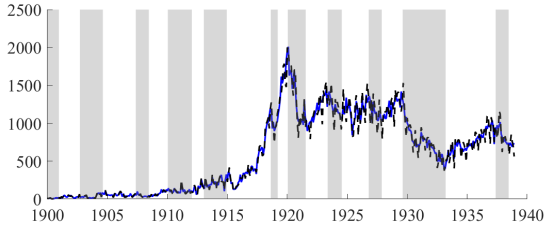
(c) Help Wanted: *Chicago Tribune*



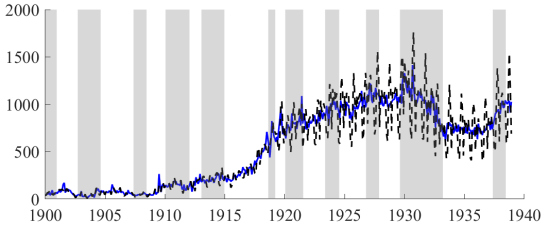
(d) Situation Wanted: *Chicago Tribune*



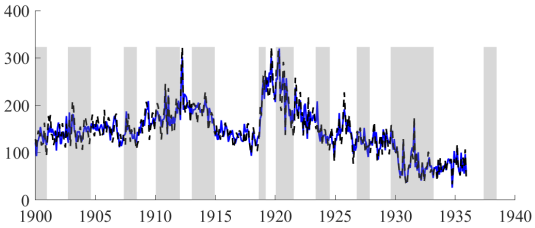
(e) Help Wanted: *New York Times*



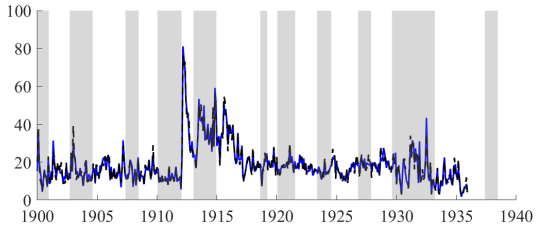
(f) Situation Wanted: *New York Times*



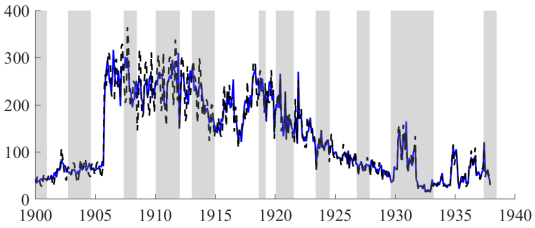
(g) Help Wanted: *Atlanta Constitution*



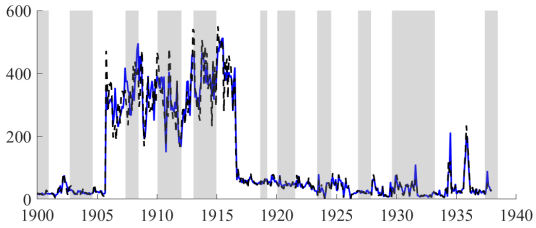
(h) Situation Wanted: *Atlanta Constitution*



(i) Help Wanted: *Washington Post*



(j) Situation Wanted: *Atlanta Constitution*



Note: The blue lines are seasonally adjusted, while the black dashed lines are not. NBER recession dates in grey.

The solid lines of Figure 2 represent the real cost for posting an advertisement once in a Sunday newspaper. To compute the real cost, we use the historical consumer price index values from the Minneapolis Federal Reserve. Since 1913, CPI has been computed by the Bureau of Labor Statistics in a consistent manner. Prior to 1913, we use the cost of living index by Albert Rees, maintained by the Minneapolis Federal Reserve.[20]

D Occupations

Table D1 provides a distribution of the occupations listed in the classified ads for newspapers on May 11, 1930. To categorize occupations, we used the classification of occupation groups from the 1930 Census of the United States. We focus on these broad occupation categories as ads often listed multiple occupations, making it difficult to classify one ad by a single job-type (e.g., an ad for a cook/chauffeur/butler or an ad for a painter/carpenter). In addition, we include a count for "no category." Several of the situations-wanted ads are very generic and do not specify an occupation (e.g., "Any work"). Likewise, sometimes a help-wanted ad can be too vague to classify an occupation (e.g., "Call to inquire about work"). Furthermore, it is sometimes impossible to read all or part of the text of an ad, making it impractical to classify the occupation. We count all of these types of occupations as "no category."

The table shows that both help- and situations-wanted ads covered a broad spectrum of occupations. Moreover, there is consistent overlap, with occupation categories appearing in both help- and situations-wanted ads. Help-wanted ads tend to be dominated by trade occupations, followed by advertising, banking, retail jobs, managers, and salesmen. Situations-wanted ads tend to be dominated by domestic and personal services (including, among others, all domestic jobs, cleaners, launderers, hotel and restaurant workers, and porters).[21] For comparison, the last column of Table D1 provides the distribution of gainfully employed individuals at the State-level in the 1930 Census.[22]

E Additional Historical Data: Details

This section documents the sources of other historical city-level and macroeconomic data.

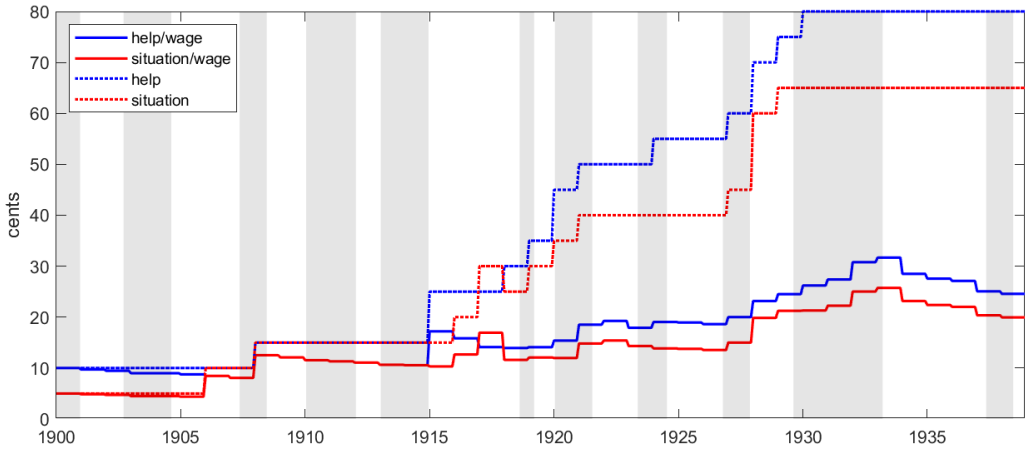
[20] https://www.minneapolisfed.org/about-us/monetary-policy/inflation-calculator/consumer-price-index-1800-

[21] We categorize all ads for a chauffeur as domestic and personal services, since these ads often are associated with domestic work (e.g., a situations-wanted ad for a cook/chauffeur/butler). The 1930 U.S. Census generally categorizes a chauffeur in the transportation and communication category, and our classification skews the distribution away from this category.

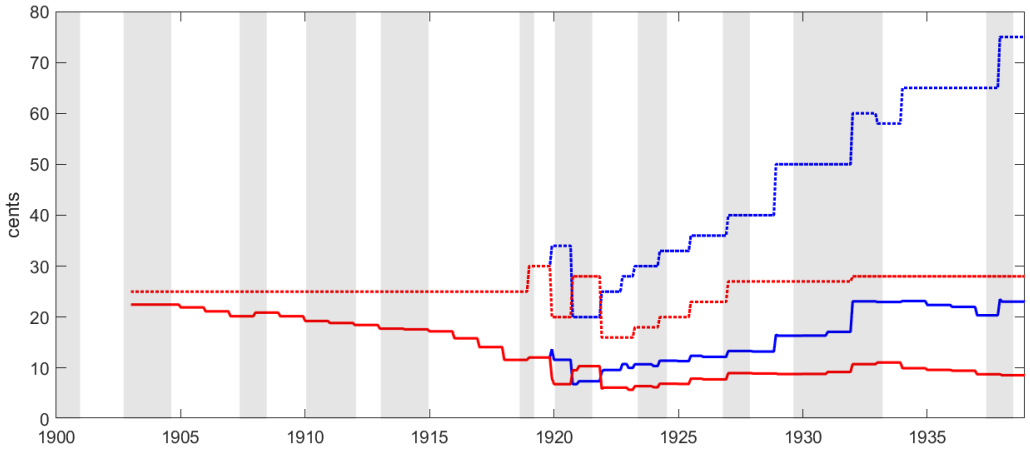
[22] State-level statistics are the most disaggregated level provided by the 1930 U.S. Census.

Figure C1: Costs for posting labor ads relative to wages: *New York Times* and *Los Angeles Times*.

(a) *New York Times*



(b) *Los Angeles Times*



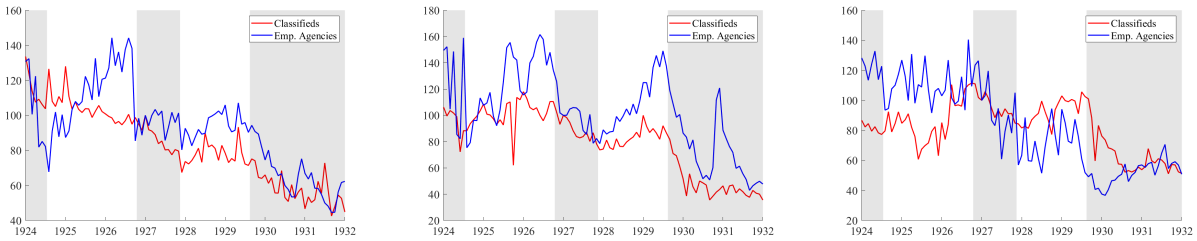
Notes: Prior to December 1920, costs for the *Los Angeles Times* were the same for both help- and situations-wanted ads.

Table D1: Frequency of Occupations on May 11, 1930

Los Angeles Times, total ads = 907			
Occupation category	percent of ads		
	Help wanted	Situation wanted	California US Census
Agriculture, fishing and forestry	<1	<1	14
Extraction of minerals	0	0	2
Manufacturing and mechanical industries	5	8	25
Transportation and communication	<1	2	8
Trade	34	4	17
Public Service	<1	0	2
Professional Service	8	3	9
Domestic and personal service	22	60	11
Clerical	15	10	10
No category	14	12	-
Chicago Tribune, total ads = 837			
Occupation category	percent of ads		
	Help wanted	Situation wanted	Illinois US Census
Agriculture, fishing and forestry	<1	<1	11
Extraction of minerals	0	0	2
Manufacturing and mechanical industries	6	6	32
Transportation and communication	<1	<1	9
Trade	59	17	15
Public Service	0	0	2
Professional Service	6	3	7
Domestic and personal service	9	46	10
Clerical	11	21	12
No category	10	4	-

Figure E1: City-Level Help Wanted: Newspapers vs. Public Employment Agency Data

(a) *Los Angeles* (b) *Chicago* (c) *New York*



Note: Newspaper and public employment agency series are normalized to 1927=100 for comparability. NBER recession dates in grey.

E.1 Bradstreet Index

To be included

E.2 State and local employment services

From January 1924 to January 1932, the U.S. Department of Labor, in cooperation with state and local employment services, produced monthly reports on the number of requests for help wanted from local businesses registered with their offices.[23] These reports, which we digitized, provide an alternative measure of city-level help wanted during this period, which we use as a check against our newspaper counts. The results, shown in Figure E1 after indexing both series to 1927, show similar trends and business cycles in both sources, though the employment agency data tend to be more volatile, suggesting the classified ads capture similar information about local labor demand.

E.3 Aggregate Time Series

- The unemployment rate and labor force are obtained from Petrosky-Nadeau and Zhang (2021).
- Labor productivity: The source data and procedure for constructing our measure of labor productivity from January 1884 to December 1940 is as follows:
  - Indexes of national productivity, by sector and type of input: 1889-1957, from Historical Statistics of the United States (Table Cg265-272, base year=1929. Source:

[23]The monthly reports also include the number of registered job seekers referred and the number placed during the reference period. According to Lee (2007), about four percent of the U.S. labor force searched for jobs through public employment offices at this time and were mainly unskilled labor.

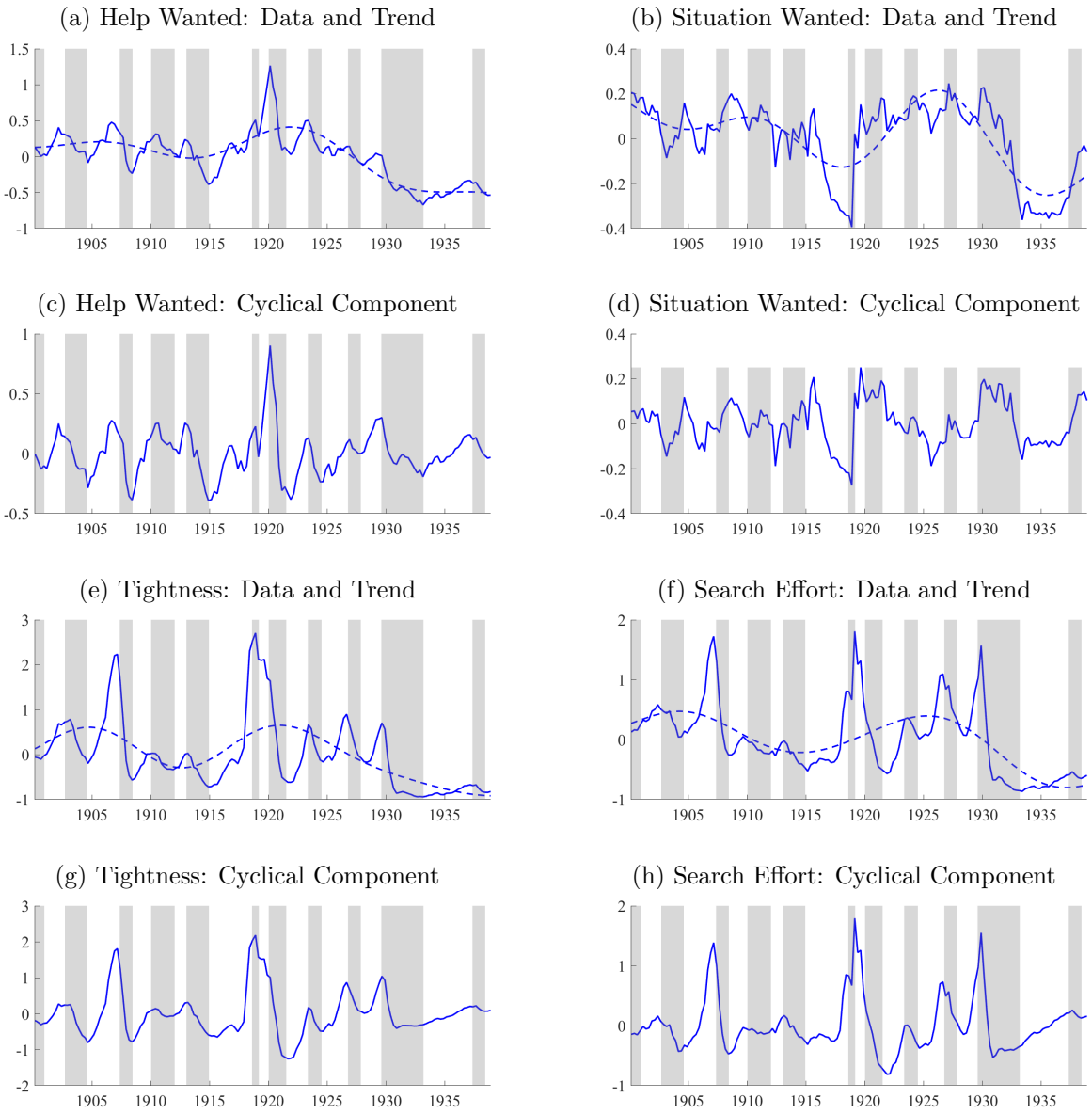
John W. Kendrick, Productivity Trends in the United States, Princeton University Press, 1961, and Postwar Productivity Trends in the United States, National Bureau of Economic Research, 1973).

  - The Miron-Romer monthly index of industrial production, January 1884-December 1940, not seasonally adjusted, from Historical Statistics of the United States (Table Cb28-31, base year=1909. Source: Jeffrey A. Miron and Christina D. Romer, "A New Monthly Index of Industrial Production, 1884-1940," Journal of Economic History 50 (June 1990): 321-35, with corrections to the original supplied by the authors). Note that a missing value in the Indexes of national productivity for March 1902 to the average value in February and April of 1902. In addition we seasonally adjust the series with the X-11-ARIMA program from US Census Bureau.
  - A monthly labor productivity series from January 1884 to December 1940 is constructed by temporly disaggregating the annual productivity series with a Denton proportional first difference procedure and the monthly industrial production series as indicators.

- Wages: The source data and procedure for constructing our monthly wage series from January 1913 to July 1946 is as follows:

  - U.S. Index of Composite Wages 01/1910-12/1920, NBER macrohistory series m08061b (January 1910-December 1920, index, base year=1913. Source: Federal Reserve Board Files, also published in Monthly Reviews of the New York Federal Reserve Bank).
  - U.S. Index of Composite Wages 01/1919-07/1946, NBER macrohistory series m08061c (January 1919- July 1946, index, base year=1926 Source: Federal Reserve Board, Monthly Reviews of Credit and Business Conditions of the New York Federal Reserve Bank).
  - Consumer Price Index for All Urban Consumers: All Items in U.S. City Average, available from Jan. 1912. U.S. Bureau of Labor Statistics, Consumer Price Index for All Urban Consumers: All Items in U.S. City Average [CPIAUCNS], retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/CPIAUCNS. We seasonally adjust the series with the X-11-ARIMA program from US Census Bureau.
  - We merge the two wage series by first rescaling series such that the average index value equals to 100 in 1920. We then deflate the nominal wage series with the seasonally adjusted CPI.

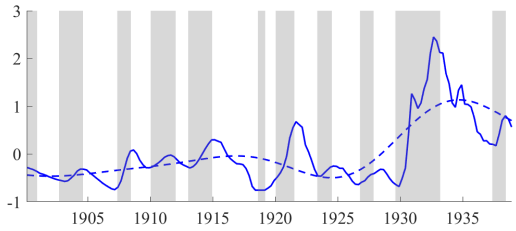
Figure E2: Actual and Band Pass Detrended Time Series



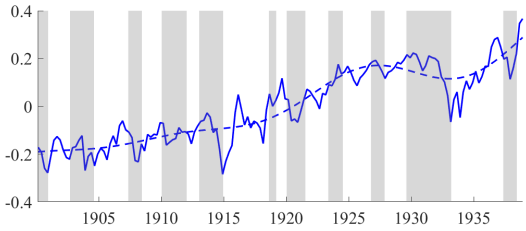
Note: NBER recession dates in grey.

Figure E3: Actual and Band Pass Detrended Time Series

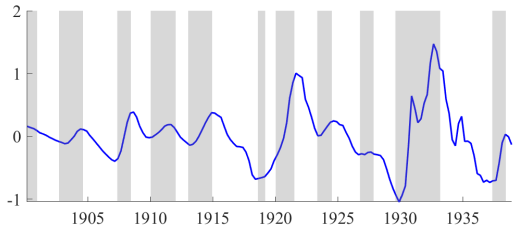
(a) Unemployment Rate: Data and Trend



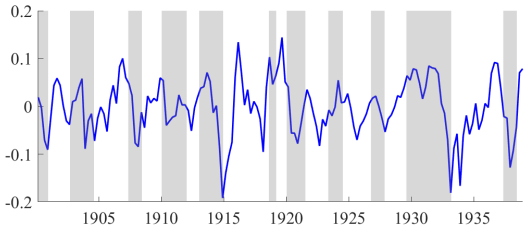
(b) Labor Productivity: Data and Trend



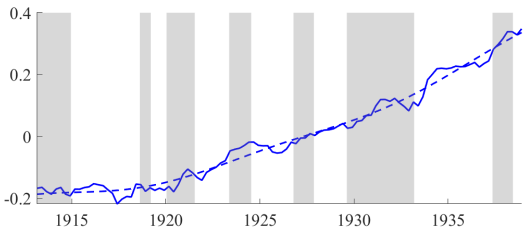
(c) Unemployment Rate: Cyclical Component



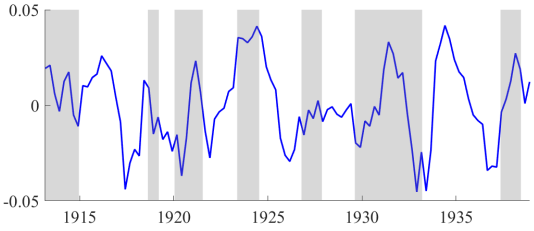
(d) Labor Productivity: Cyclical Component



(e) Real Wage: Data and Trend



(f) Real Wage: Cyclical Component



Note: NBER recession dates in grey.