ISSN 1936-5330

Research Working Papers

Firm Search in the Labor Market: Evidence from Help-Wanted Advertisements

Huixin Bi, Nicolas Petrosky-Nadeau, Nora Traum, and Greg Woodward
July 2025; updated May 2026
RWP 25-07
https://doi.org/10.18651/RWP2025-07
This paper supersedes the old version:
"Worker and Firm Search in the Labor Market: Evidence from Classified Advertisements"

Firm Search in the Labor Market: Evidence from Help-Wanted Advertisements*

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

May 15, 2026

Abstract

We construct new monthly city-level and national measures of firm search for workers from 1900 to 1938, drawing on approximately 5 million scanned help-wanted advertisements from five U.S. newspapers, with breakdowns by gender. We document four main findings: (1) firm search effort is procyclical, declining sharply at the onset of recessions; (2) posting costs affect advertising behavior, but the effect is modest, with an elasticity of −0.09; (3) the U.S. Beveridge curve has been stable for the past 125 years, with matching elasticities of 0.57 pre-WWII and 0.55 post-WWII; and (4) help-wanted advertisements for women are more responsive than those for men to both posting costs and the business cycle.

JEL Classification: J64, N32, E32, C82

Keywords: help-wanted; job vacancies; historical data

*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. Many excellent graduate, undergraduate and professional research assistants contributed to this project; see Appendix A for complete acknowledgments. We also appreciate comments from our discussants Ryan Michaels and Ronald Wolthoff, as well as those from Price Fishback, Claudia Goldin, Kris Mitchener, and participants at various seminars and conferences. 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

The early twentieth century was a period of significant economic upheaval in the United States. Several pivotal events reshaped the U.S. economy: two World Wars, the Great Depression, banking crises, and a pandemic. The period also brought sweeping social and political changes, including the rise of labor unions, the introduction of unemployment benefits, significant shifts in tariff policy, and immigration reform. However, despite its historical importance, our understanding of how the labor market behaved during this period is constrained by limited data. Existing studies rely mainly on annual or lower-frequency data series or cover only a subset of the economy, providing little systematic information on firms' demand for labor and making it difficult to evaluate the stability of labor-market relationships central to modern search and business-cycle models.

We fill this gap by introducing new monthly, geographically diverse, and gender-disaggregated measures of firm search activity that opens a new window onto how labor markets functioned in the first decades of the twentieth century. Specifically, we construct measures of labor demand from January 1900 to December 1938 using a single source: the classified ads of contemporary newspapers. Help-wanted ads are a familiar proxy for job openings and labor demand in the second half of the twentieth century. Beyond extending the available national series back to 1900, we provide novel breakdowns by location and gender.

The new series yields four central findings. First, firm search effort was highly procyclical, contracting sharply at the onset of recessions. Second, advertising costs shaped firms' posting decisions, although the estimated elasticity is modest at -0.09. Third, despite profound structural changes in the U.S. economy, the Beveridge curve remained remarkably stable over the last 125 years, with matching elasticities of 0.57 before World War II and 0.55 afterward. Finally, help-wanted advertisements for women were more responsive both to posting costs and to cyclical fluctuations. Taken together, these findings suggest that several empirical relationships central to modern search models are remarkably stable across different macroeconomic and institutional environments.

To construct the labor demand series, we extract individual advertisements from scanned images of five newspapers with a broad national coverage for the time: the *New York Times*, *Chicago Tribune*, *Atlanta Constitution*, *Los Angeles Times*, and *Washington Post*.[1] We process and catalog approximately 5 million labor advertisements using a pipeline of image

[1]The scanned images are accessed through rights from ProQuest.

processing techniques tailored to the dense classified pages of this period.[2] We first identify columns and individual advertisements within newspaper pages, then classify sections such as "Help Wanted" using template matching on classified-page headers. Because job advertisements remained legally gender-segregated until Title VII of the 1964 Civil Rights Act and persisted in practice into the early 1970s (Pedriana and Abraham, 2006), we can additionally classify help-wanted ads separately for men and women. We then use these counts to construct city-level and national help-wanted indices.

We validate the methodology with an extensive series of checks. First, our automated counts match closely with about one hundred thousand manual counts conducted at regular intervals across the five newspapers. The average median error across newspapers is 4 percent. Second, we 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), available from 1919 onward. Weighting cities by populations, our national help-wanted index has a correlation of 0.98 with the MetLife national help-wanted index over the overlapping period. Finally, we digitize help-wanted registrations from local public employment agencies between 1924 and 1932, demonstrating that our city-level series track these alternative measures closely.

The paper makes three contributions. First, we construct new monthly measures of firm search at both the city and national levels, disaggregated by gender from 1900 to 1938. The national series extends systematic vacancy measurement nearly two decades before the MetLife index began in 1919, and the city-level and gender breakdowns are entirely new. Second, we establish that several empirical regularities central to labor market search theory also hold in pre-WWII data: firm search effort is strongly procyclical, the Beveridge curve is stable over 125 years, and the matching elasticity is similar across eras (0.57 pre-WWII versus 0.55 post-WWII). Third, we document the role of posting costs and gender in shaping vacancy dynamics. The elasticity of help-wanted advertising with respect to posting costs is -0.09 — statistically significant but economically modest — and help-wanted ads for women display roughly twice the cyclical sensitivity of those for men, with impulse responses to productivity shocks approximately double in magnitude.

At the city level, all five metropolitan areas display common cyclical patterns in help-wanted advertising, with sharp declines during the Great Depression. Meaningful regional variation nonetheless emerges: Atlanta rebounded earlier than the other cities, consistent

[2] The high density of classified ads appears challenging for the pre-trained layout-detection models of Shen *et al.* (2021). Instead, we rely heavily on computer vision to identify individual ads.

with milder employment losses in the South (Wallis, 1989). Gender differences are pronounced throughout the sample. Help-wanted ads for women were more volatile than those for men, rising sharply after World War I. Although female workers were a smaller share of the employed population, women were overrepresented among workers reached through newspaper ads. This finding is consistent with limited access to alternative search channels, as women were largely excluded from union-based referrals and from direct hiring at the work site (Kessler-Harris, 1982).

To examine the role of posting costs, we hand-collect advertising rates for three newspapers in our sample and estimate a panel regression of ad counts on costs. A period during which the *Washington Post* made classified ads free provides a natural experiment: the removal of posting costs led to an immediate and substantial increase in job postings, concentrated among positions for women. Combined with the greater cyclical sensitivity of female help-wanted ads, this suggests that the cost margin was particularly important for employers seeking female workers.

Our work contributes to several strands of the literature. First, it speaks to studies that use data to examine economic activity, and labor markets in particular, in the early twentieth century. A number of papers construct measures of employment and hours worked at lower frequencies (e.g., Ramey and Francis, 2009, Wallis, 1989), while others examine the effects of technological change on labor market outcomes (Alexopoulos and Cohen, 2016, Goldin and Katz, 1998). Goldin (1990) documents important shifts in employment patterns and wage structures for women over this period. More closely related to our paper, Lee (2016) uses data from public employment offices between 1924 and 1932 to estimate a matching function, while DeVaro and Gurtler (2018) study long-run variation in help-wanted advertising using manual newspaper counts. We instead automate the extraction of advertisements from historical newspapers to construct systematic high-frequency measures of help-wanted advertising for the first four decades of the twentieth century.

Our findings also speak to the modern business-cycle literature on labor market search. Merz (1995) and Shimer (2005) document the cyclical properties of vacancies and labor market tightness in post-war U.S. data, establishing the stylized facts that search models aim to match. Fujita and Ramey (2009) study the dynamic response of job vacancies to productivity shocks using structural VARs. We extend these analyzes to the pre-WWII period and show that the cyclical moments and impulse response patterns are qualitatively similar, despite substantially higher volatility before WWII. Our estimation of a matching function complements the survey by Petrongolo and Pissarides (2001): we find a matching elasticity in the pre-war period that is comparable to the post-war estimates.

The rest of this paper is organized as follows. Section 2 provides an overview of labor postings in classified ads, as well as our methodology and strategy for developing indices based on these ads. Section 3 focuses on the city-level indices, while section 4 presents the aggregate help-wanted index at the national level. Section 5 concludes.

2 Background and Data Construction

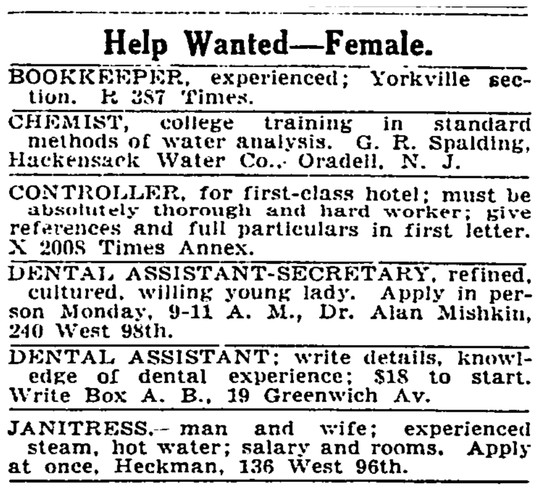
In the early twentieth century, firms used a variety of methods to connect with job seekers in the labor market: informal networks, direct application, newspaper ads, and employment agencies (see Fishback, 1998; Simon, 2001). Direct evidence on the hiring process during this period is limited, with most studies relying on small surveys of firms in specific industries (e.g., Schweinitz, 1932; Fishback, 1992; Rosenbloom, 1994).[3] Classified advertisements were used consistently by firms, and analyzing them offers valuable insights into the functioning of the labor market. A particular advantage of newspaper advertisements was their wide circulation and readership, which made them an effective outlet for advertising positions to a broad regional market.[4]

We use digitized records of widely-circulated newspapers in major metropolises to build a monthly database of advertisements by employers. The number of newspapers in our analysis is constrained primarily by access: after 1923, U.S. newspapers became subject to copyright laws that restricted public access (Heald, 2014). We focus on five newspapers with wide circulation in their respective cities, accessed under direct rights from ProQuest for the period from January 1, 1900 through December 31, 1938: (i) the *New York Times*; (ii) the *Chicago Tribune*; (iii) the *Los Angeles Times*; (iv) the *Atlanta Constitution*; and (v) the *Washington Post*. According to the 1930 Census, the populations of these five metropolitan areas together accounted for more than 33 percent of all U.S. city populations and roughly 10 percent of the total U.S. population. The following subsections describe classified ads during this period, our methodology for extracting labor ads, and our checks on the accuracy of the resulting counts.

[3]These studies suggest that informal networks and direct applications were the preferred job search methods in mining and manufacturing, but considerably less is known about search practices in other sectors. Classified advertisements provide a valuable complementary perspective on search across a broad range of sectors. Simon (2001) notes that classified ads were a significant part of the search process for clerical workers.

[4]Education had increased markedly by the early twentieth century: the illiteracy rate was only 4 percent by 1930, and overall enrollment among 5- to 19-year-olds rose from 51 percent in 1900 to 75 percent in 1940 (Snyder, 1993). Employment agencies also frequently posted in the classifieds, so newspaper ads partially encompass other search methods.

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



Help Wanted—Female.

BOOKKEEPER, experienced; Yorkville section. R 287 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.

2.1 Classified advertisements

In the first half of the twentieth century, employer advertisements in the classified section of newspapers were organized by gender into distinct sections for men and women. Ads sometimes listed information on offered compensation as well as education and/or experience requirements. They could include other characteristics requested by employers, such as race, ethnicity, or marital status.

As an example, Figure 1 shows a sample of ads seeking female workers in the *New York Times* on January 19, 1930. The advertisements include higher-paid jobs with substantial qualification requirements, such as a college-trained chemist, alongside lower-paid jobs with little required expertise, such as a janitor.[5]

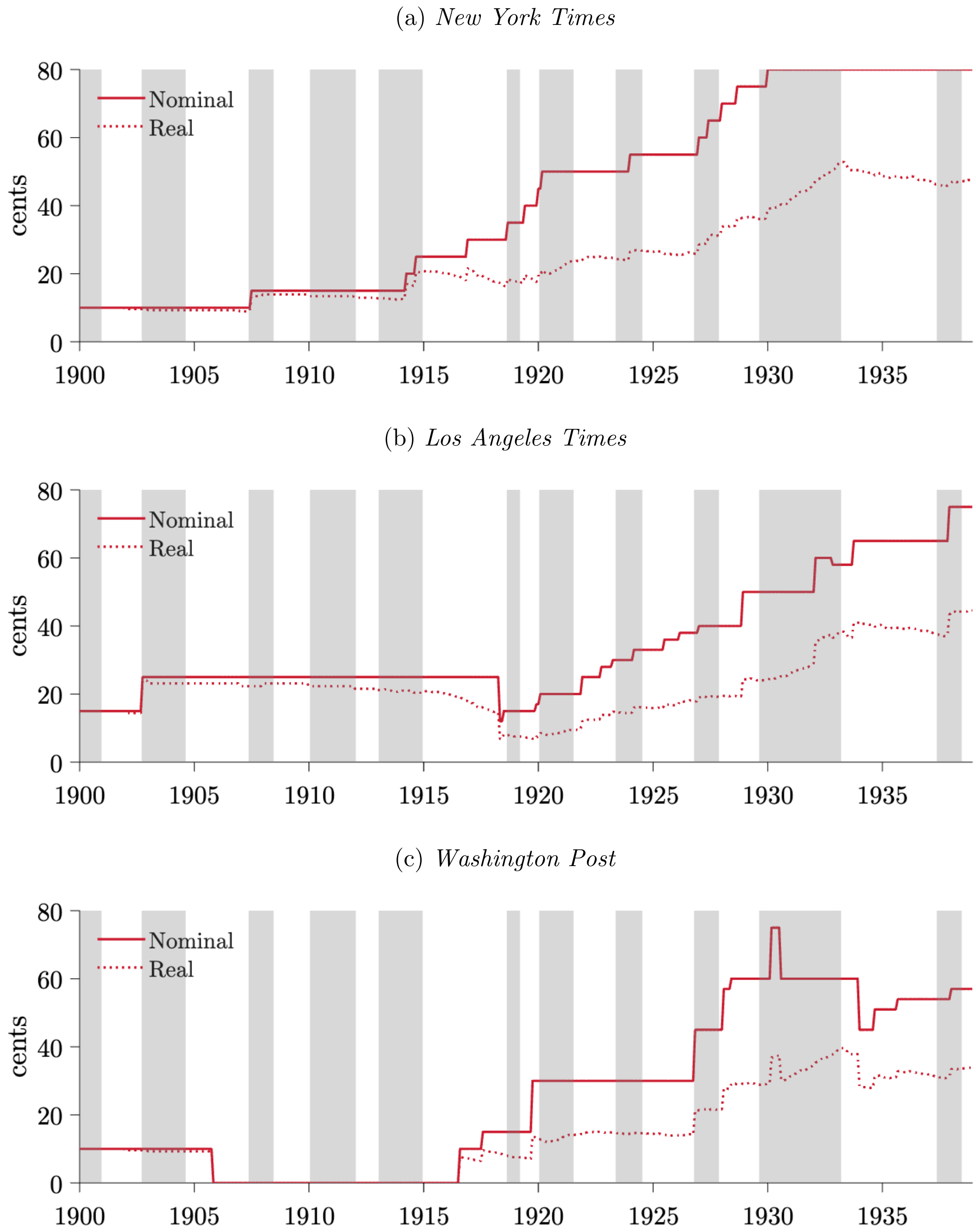
The posting of advertisements required a monetary cost that varied over time and across cities, with rates regularly printed in the Sunday edition of some newspapers. We hand-collect the posting costs for the *New York Times*, the *Los Angeles Times*, and *Washington Post*.[6] Figure 2 reports the nominal cost of help-wanted ads placed in the Sunday paper in solid lines and the real cost adjusted by the Consumer Price Index in dashed lines.

Although nominal costs rarely changed before 1915, posting costs were adjusted frequently between 1915 and 1938, and the frequency and magnitude of changes varied across

[5] Appendix B discusses the distribution of jobs listed relative to the 1930 U.S. Census. In general, help-wanted advertisements were dominated by trade occupations and excluded manufacturing and mining industries.

[6] Appendix C provides additional details. Information on posting costs were not printed in the *Chicago Tribune* and *Atlanta Constitution*.

Figure 2: Costs for posting help-wanted ads: *New York Times*, *Los Angeles Times*, and *Washington Post*.



Notes: Costs are for posting help-wanted ads one time in the Sunday paper. Solid lines are nominal costs; dashed lines are real costs adjusted by the Consumer Price Index. Shaded bar correspond to NBER dated recessions.

cities. The *New York Times* and the *Los Angeles Times* made more frequent changes at an incremental pace, while the *Washington Post* made fewer changes but with larger magnitude. The *Los Angeles Times*, for example, changed its help-wanted rate 14 times between 1918 and 1933, by an average of less than 3 cents, while the *Washington Post* changed its rate only 5 times, by an average of 10 cents. The *Washington Post* also made labor ads free between November 1905 and July 1916. The changes appear unrelated to business-cycle conditions.[7] Real costs were relatively stable until 1926, as nominal rate changes broadly tracked the Consumer Price Index. After 1926, the increase in real costs was particularly pronounced during the Great Depression, as the economy entered deflation.

2.2 Extracting Labor Advertisements

Identifying labor advertisements at scale requires a novel approach, as our raw data are image files of newspaper pages. The procedure is outlined in the following subsections.

2.2.1 Splitting Ads Previous studies often apply textual analysis to text files produced from image scans by Optical Character Recognition (OCR). For instance, Atalay *et al.* (2020) used regular textual patterns, such as a help-wanted ad beginning with a word in upper-case letters, to discern boundaries between labor ads for the period 1950–2000. During our period of interest, classified ads were arranged in columns of densely-packed listings.[8] Although OCR technology has advanced significantly and now performs well on cleaner text, state-of-the-art OCR still fails to identify most break points between ads when applied directly to an entire newspaper page with many columns, or even to a single column with many ads. Properly separating labor ads is crucial for constructing accurate labor demand indices: failure to do so can produce large error rates.

To overcome these challenges, we rely on digital image processing to identify ad boundaries directly from images of the classified pages. A common formatting style for classified sections used black vertical bars to separate the page's columns and black horizontal lines to separate individual ads. We exploit this structure to identify and divide columns and ads, as illustrated in panels (a) and (b) of Figure 3, in three steps. First, we straighten and rotate any images that are misaligned from the scanning process. Second, we split newspaper

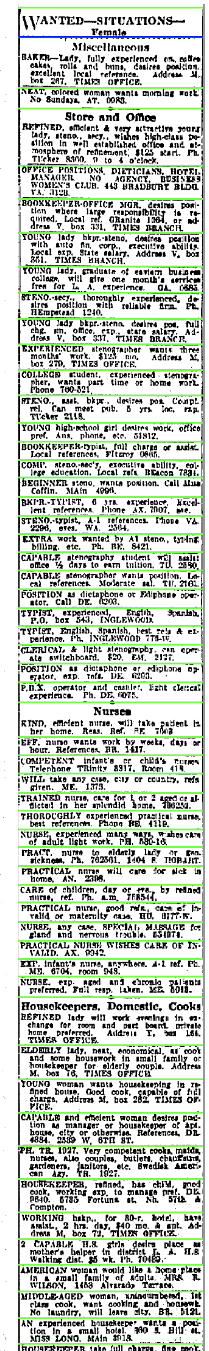
[7] Kirchhoff (2011) reports that total newspaper advertising revenue fell by 45 percent between 1929 and 1933. Our data indicate that for classified advertisements specifically, this decline was driven primarily by a reduction in the number of postings rather than by changes in advertising rates (see Section 3). In Section 3.3, we exploit changes in posting costs across cities and over time to estimate their impact on firm search effort.

[8] ProQuest provides text files derived from scanned images of newspaper pages, but the text quality for the classified ads in our sample period is too poor for textual analysis. ProQuest also outputs a single raw text file for *all* classified ads on a given page, making it nearly impossible to identify individual ads even when the file quality is adequate.

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

(a) Column Splits

(b) Ad Splits



WANTED—SITUATIONS—
Female

Miscellaneous

BAKER—Lady, fully experienced on coffee cakes, rolls and buns, desires position. excellent local reference. Address M, box 267, TIMES OFFICE.

NEAT, colored woman wants morning work. No Sundays, AT. 0083.

Store and Office

REFINED, efficient & very attractive young lady, steno., secy., wishes high-class position in well established office and atmosphere of refinement. $125 start. Ph. TUcker 8300, 9 to 4 o'clock.

OFFICE POSITIONS, DIETICIANS, HOTEL MANAGER. NO AGENCY. BUSINESS WOMEN'S CLUB. 443 BRADBURY BLDG. VA. 3128.

BOOKKEEPER-OFFICE MGR. desires position where large responsibility is required. Local ref. GRanite 1004, or address V, box 331, TIMES BRANCH.

YOUNG lady bkpr.-steno. desires position with auto fin. corp., executive ability. Local exp. State salary. Address V, box 351. TIMES BRANCH.

YOUNG lady, graduate of eastern business college, will give one month's services free for L. A. experience. GA. 0685.

STENO.-secy., thoroughly experienced, desires position with reliable firm. Ph. HEmpstead 1240.

YOUNG lady bkpr.-steno. desires pos. full chg. sm. office, exp., state salary. Address V, box 337, TIMES BRANCH.

EXPERIENCED stenographer wants three months' work. $125 mo. Address M, box 270, TIMES OFFICE.

COLLEGE student, experienced stenographer, wants part time or home work. Phone 760-521.

STENO., asst. bkpr., desires pos. Compt. rel. Can meet pub. 5 yrs. loc. exp. TUcker 2118.

YOUNG high-school girl desires work, office pref. Ans. phone, etc. 51812.

BOOKKEEPER-typist, full charge or assist. Local references, FItzroy 0605.

COMP. steno.-sec'y, executive ability, college education. Local refs. BEacon 7831.

BEGINNER steno. wants position. Call Miss Coffin. MAin 4996.

BKPR.-TYPIST, 6 yrs. experience. Excellent references. Phone AX. 7007, eve.

STENO.-typist, A-1 references. Phone VA. 2296, eves. WA. 2564.

EXTRA work wanted by A1 steno., typing, billing, etc. Ph. RE. 8421.

CAPABLE stenography student will assist office ½ days to earn tuition. TU. 2890.

CAPABLE stenographer wants position. Local references. Moderate sal. TU. 2161.

POSITION as dictaphone or Ediphone operator. Call DE. 6203.

TYPIST, experienced, English, Spanish, P.O. box 543, INGLEWOOD.

TYPIST, English, Spanish, best refs & experience. Ph. INGLEWOOD 778-W.

CLERICAL & light stenography, can operate switchboard. $20. EM. 2137.

POSITION as dictaphone or ediphone operator, exp. refs. DE. 6203.

P.B.X. operator and cashier, light clerical experience. Ph. DE. 6075.

Nurses

KIND, efficient nurse, will take patient in her home. Reas. Ref. RE. 7503

EFF. nurse wants work by weeks, days or hour. References, BR. 1417.

COMPETENT infant's or child's nurse. Telephone TRinity 8317, Room 418.

WILL take any case, city or country, refs given. ME. 1373.

TRAINED nurse, care for 1 or 2 aged or afflicted in her splendid home. 786252.

THOROUGHLY experienced practical nurse, best references. Phone BE. 4119.

NURSE, experienced many ways, wishes care of adult light work. PH. 580-16.

PRACT. nurse to elderly lady or gen. sickness. Ph. 702561. 1404 S. HOBART.

PRACTICAL nurse will care for sick in home. AN. 2398.

CARE of children, day or eve., by refined nurse, ref. Ph. a.m. 768544.

PRACTICAL nurse, good refs., care of invalid or maternity case. HU. 8177-W.

NURSE, any case. SPECIAL MASSAGE for gland and nervous trouble. 554974.

PRACTICAL NURSE WISHES CARE OF INVALID. AX. 9042.

EXP. infant's nurse, anywhere. A-1 ref. Ph. ME. 6704, room 948.

NURSE, exp. aged and chronic patients preferred. Full resp. taken. ME. 6048.

Housekeepers. Domestic. Cooks

REFINED lady will work evenings in exchange for room and part board, private home preferred. Address T, box 184. TIMES OFFICE.

ELDERLY lady, neat, economical, as cook and some housework in small family or housekeeper for elderly couple. Address M, box 76, TIMES OFFICE.

YOUNG woman wants housekeeping in refined house. Good cook, capable of full charge. Address M, box 282, TIMES OFFICE.

CAPABLE and efficient woman desires position as manager or housekeeper of apt. house, city or otherwise. References, DR. 4384. 2339 W. 6TH ST.

PH. TR. 1027. Very competent cooks, maids, nurses, also couples, butlers, chauffeurs, gardeners, janitors, etc. Swedish American Agy. TR. 1027.

HOUSEKEEPER, refined, has child, good cook, working exp. to manage pref. DE. 9640. 5785 Fortuna st. Nr. 57th & Compton.

WORKING hskp., for 80-r. hotel, have assist., 2 hrs. day, $40 mo. & apt. Address M, box 72, TIMES OFFICE.

2 CAPABLE H.S. girls desire place as mother's helper in district L. A. H.S. Walking dist. $5 wk. Ph. 70489.

AMERICAN woman would like a home place in a small family of adults. MRS. R. WILSON, 1463 Alvarado Terrace.

MIDDLE-AGED woman, unincumbered, 1st class cook, want cooking and housewk. No laundry, will leave city. BR. 5121.

AN experienced housekeeper wants a position in a small hotel. 360 S. Hill st. MISS LONG. MAin 8913.

HOUSEKEEPER take full charge, fine cook, small family, adults. Best city refs. Ph. Rm. 109-D. TR. 2244.

COOK, exp. rel. some housework, other help preferred, wages $50 and up. TU. 6250.

pages into columns by identifying concentrations of black pixels in narrow vertical bands and grouping them into clusters based on their horizontal coordinates. Third, we split columns into individual ads, the most challenging step because of the density of text. We invert the image colors and remove text from each column to simplify the search for dividers, then compute the mean of each row of pixels: a row corresponding to a (now inverted) divider line should lie close to a white pixel value. Based on the distribution of pixel values within each column, we compute a prominence threshold and use it to identify ad dividers. Appendix D.1 provides further details for each of these steps.

2.2.2 Identifying Help-Wanted Ads

While the image-processing procedure allows us to identify every ad in the classifieds section, it does not tell us the type of ad, that is, whether the ad is a labor ad posted by an employer or, say, a person advertising a room for rent. To classify ads, we identify section headers using an image-matching procedure. As detailed below, we create image templates that match the section headers used by a particular newspaper at a particular point in time.

The template-matching approach proceeds in three steps. First, we manually identify the headers used for labor demand (help-wanted) ads and for the sections that immediately follow them. Second, we create image templates that match the typesetting of these headers as they appear in the underlying section header images. Third, we use template-matching scores to identify the beginning and end of each relevant classified section and sort individual ads into their respective categories.

Section headers varied in wording and font style across newspapers and over time, even though they typically appear in offset lettering and a larger font than the words in an individual ad. As an example, Figure 4 illustrates some changes to section headers in the *Los Angeles Times* between 1925 and 1938. The frequency of header changes varies considerably across newspapers. For instance, the *Chicago Tribune* changed its section headers only twice over the data sample period, while the *New York Times* made 20 such changes. Appendix D.2 provides an extensive catalog. In total, we create 947 distinct header templates across newspapers over our entire time frame.

Section headers are identified by matching ad images to templates designed to mirror the typesetting of headers in the underlying scans. The matching procedure proceeds as follows: (i) create a template for each identified section header; (ii) compute a match score between each candidate ad image and each template; (iii) compare the score against a threshold to determine whether the image is a section header and, if so, of what type. The match score

Figure 4: Changing Section Headers: examples from the *Los Angeles Times* between 1925 and 1938.

(a) 1925 (b) 1931 (c) 1938

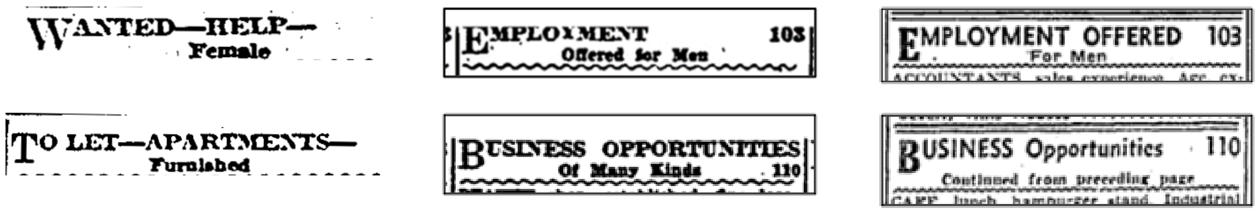


Figure 5: Steps to the Template Matching Procedure: example for identifying "Employment offered" section in the *Los Angeles Times*.

(a) Ad Images (b) Templates (c) Max Match Locations

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

lies between 0 and 1, with 1 indicating that the template exactly matches the pattern of pixels in the underlying image. Headers are then classified as either labor demand or other categories based on preset thresholds. We count every ad immediately following an identified header as belonging to that category, up to the next identified header or the end of the page. Appendix D.2 provides further details.

Figure 5 illustrates the three steps of template matching for the "Employment Offered" header in the *Los Angeles Times* in the 1930s. We first identify the header in question (top left panel) and create two templates, "Employment" and "Offered," that mirror its typesetting, size, and font (middle panel). A 2-dimensional convolution then locates the best match of each template within the candidate ad image. The red bounding boxes in the right panel mark the windows with the highest correlation between template and image pixels. The resulting match scores for "Employment" and "Offered" (0.66 and 0.68, respectively) clear the preset threshold of 0.55, and the header is classified as a labor demand header. Because the accuracy of template matching depends on the chosen thresholds, we tune them through an iterative process, described in Appendix D.2.

Although the three-step procedure works well for most newspaper pages, multicolumn headers in the *New York Times* require modifications to the approach (see Figure D4). We use an object-detection model, a computer-vision technique that recognizes and localizes objects of interest, to detect and locate these multicolumn headers. We first build a training sample by manually classifying objects on a large number of newspaper pages into three types: (1) help-wanted multicolumn headers, (2) other multicolumn headers, and (3) anything else. We then train the object-detection model on this sample, apply the trained model to each newspaper page, and record the location of every detected bounding box together with its predicted class probability. Each header is classified by comparing its predicted probability to preset thresholds; details are given in Appendix D.3.

To convey the scale of the image-processing effort, consider the *Chicago Tribune*. We begin with over 40 thousand image files of newspaper pages, about 10 thousand of which contain labor ads. These files are split into almost 4 million individual ads, of which 3 million are identified as labor ads. Appendix D.4 provides analogous figures for all five newspapers. The next subsection discusses the accuracy of our identification procedure.

2.3 Accuracy

We test the accuracy of our labor-ad identification with extensive manual counts. For each year between 1900 and 1938, we manually counted help-wanted ads on the second Sunday in May for all five newspapers, yielding more than 100 thousand individual help-wanted ads. We use the same Sunday across years for consistency. As a robustness check, we also randomly select two newspaper pages per month across newspapers and count those manually (see Appendix D.5). Figure 6 compares the number of ads counted by our algorithm with the manual counts.[9] The method works well overall: the observations cluster tightly around the 45-degree line, indicating close agreement between manual counts and template matching.

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.

[9]Appendix D.5 provides further details on the manual checks. As a further alternative, we identified ads using a Convolutional Neural Network (CNN). This method is more time- and labor-intensive but produces results comparable to template matching; see Appendix D.6.

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

	Number of ads	Error rate 1st Tertile	Median	3rd Tertile
New York Times	23,486	0.01	0.04	0.06
Chicago Tribune	47,723	0.02	0.03	0.06
Los Angeles Times	20,478	0.03	0.06	0.18
Atlanta Constitution	4,634	0.03	0.04	0.11
Washington Post	5,114	0.01	0.02	0.05

Notes. Errors between machine and manual counts computed according to equation (1). Manual counts collected on the second Sunday in May between 1900 and 1938.

Errors are generally small across newspapers and over time, despite the frequent changes in header wording and typesetting. Table 1 summarizes the distribution of errors by newspaper, reporting the median share along with the first and third tertiles. Identification is most accurate for the *Chicago Tribune*, with a median misclassification share of 3 percent. It is least accurate for the *Los Angeles Times*, which has relatively poor print quality, but accuracy remains high there as well, with a median error of 6 percent.

2.4 Extracting Gender Based Help-Wanted ads

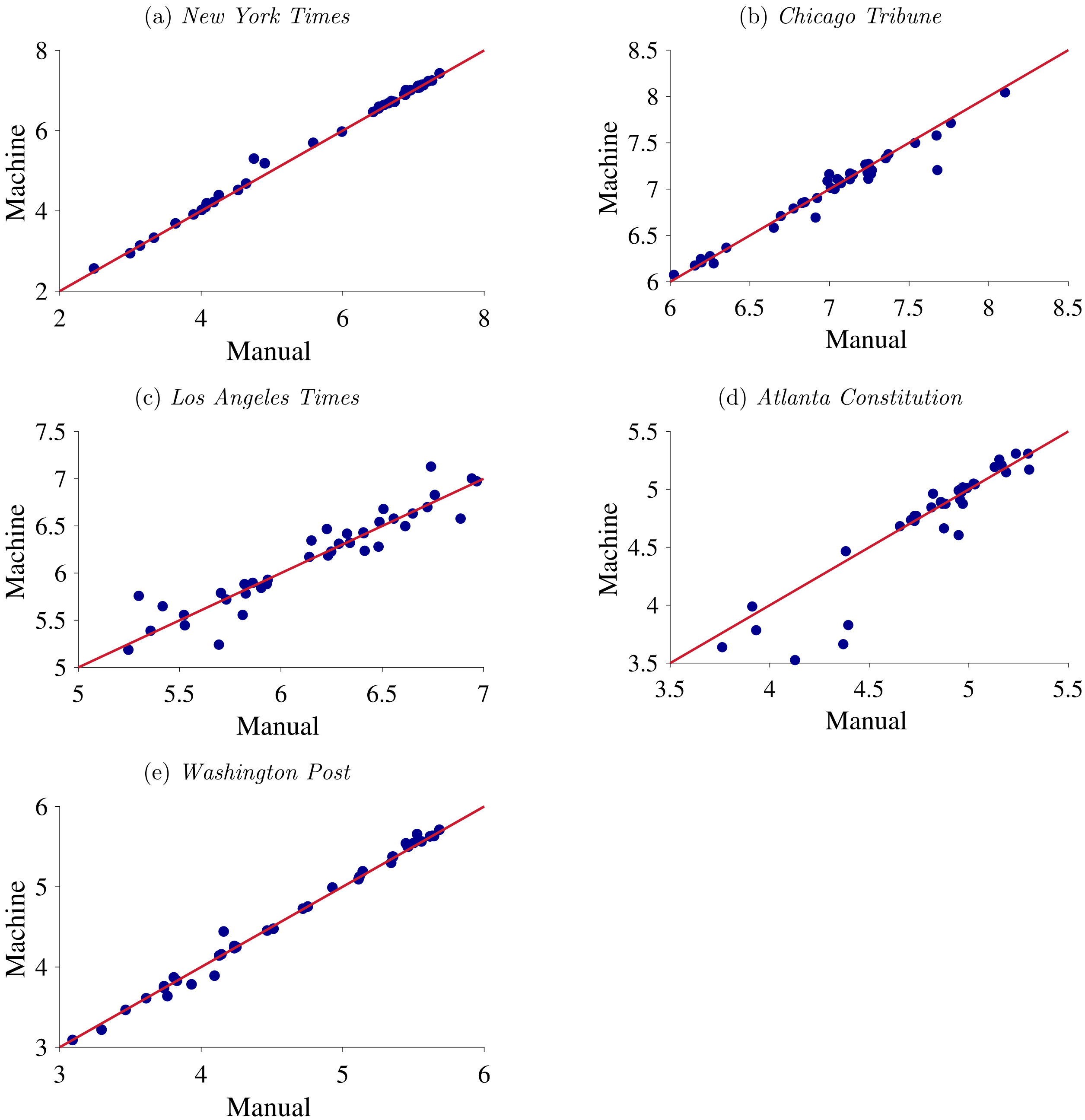
Labor demand ads were grouped by gender in the early twentieth century, as shown in Figures 1 and 3. We exploit this distinctive feature to separate our measures by gender. The pattern of gendered job advertising has been well documented in historical labor-market studies (e.g., Goldin (1990)) and provides a useful opportunity to compare the male and female labor markets in the first half of the twentieth century.[10]

To extract gender information from headers, we combine the template-matching procedure of Section 2.2.2 with OCR. This hybrid approach is motivated by a practical consideration: template matching alone struggles with the small fonts often used to denote gender within section headers, and the dividers between headers and ads are frequently faint, causing the first ad to bleed into the header image. We therefore proceed in two steps. We first run OCR once per classified page using Google Cloud Vision, which provides the best results for our application of the tools we tested.[11] The OCR output returns bounding boxes for

[10] The mix of advertised occupations also differs by gender. Goldin (1990), p. 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.

[11] We tested Amazon Textract, Google Cloud Vision, and the open-source Tesseract. Correia and Luck (2023) reach a similar conclusion regarding Google Cloud Vision. In addition, we apply OCR to each

Figure 6: Accuracy of help-wanted ad counts



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

every letter and word on the page. We then assign each OCR-detected word to a specific ad using the ad bounding boxes recorded during the column- and ad-splitting step (Section 2.2). For each header identified by template matching, we reclassify it as for men or for women based on whether gender-specific words (e.g., *male*, *men*, *female*, *women*) appear in the OCR text within the header's bounding box. Appendix D.7 provides further details.

3 Help-Wanted Time Series: City Level Measures

In this section, we present city-level measures of help-wanted advertising. We first report total ad counts, constructed as monthly averages of Sunday editions, and then provide breakdowns by gender. In addition, we leverage variation in posting costs across cities and over time to estimate their effect on firm advertising behavior.

3.1 City-level Counts

We construct monthly series of help-wanted ads for each city by averaging the Sunday counts within a month. We focus on Sundays as they carried substantially more classified postings than other days of the week. Bezanson (1929) notes that in the early twentieth century, help-wanted advertisements displayed notable seasonal patterns in certain occupations, for instance, demand for domestic workers spiked in the summer months. We observe seasonal patterns throughout our sample. To account for these, we seasonally adjust each city-level measure and take a 3-month moving average of the resulting series.[12] The measures for each city are shown in Figure 7, with NBER recession dates shaded in grey. For the *New York Times*, we report ad counts prior to January 1919 with a dashed line to reflect the paper's low circulation in those earlier years.[13]

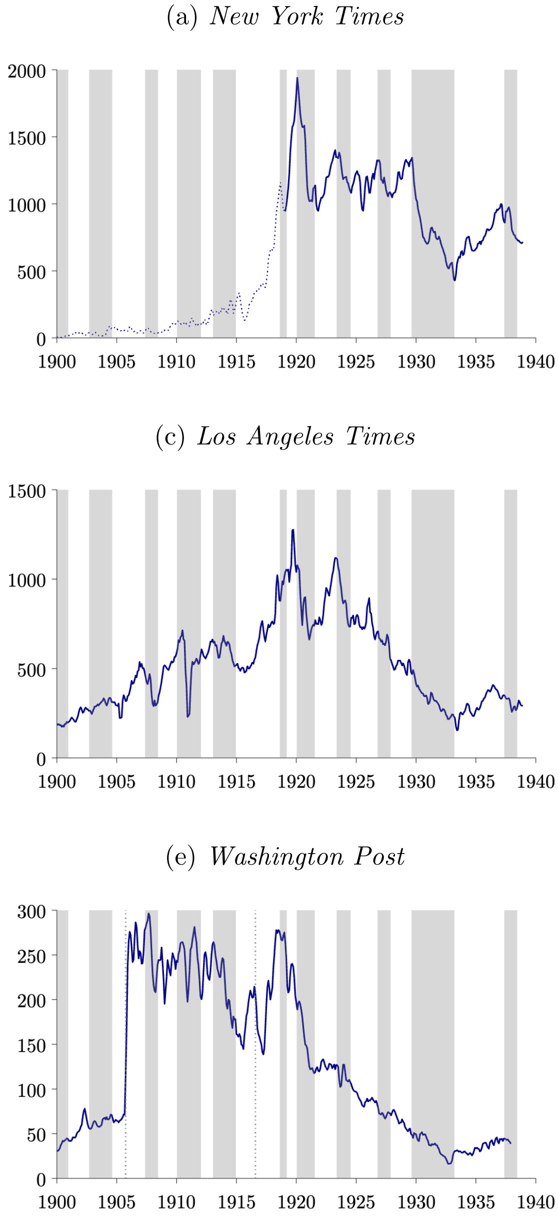
Several newspaper-specific patterns stand out for New York, Los Angeles, and Washington. As noted, the *New York Times* had comparatively low circulation in the city before 1919 and hence captured a low volume of help-wanted advertising. The *Los Angeles Times* shows a moderate upward trend during the first seven years of the sample, likely reflecting the rapid growth of the city's population. The *Washington Post* exhibits a large spike in late 1905 driven by the paper's decision to make labor advertisements free starting that November, with the goal of boosting circulation.

newspaper page rather than each ad, as providers price by file rather than by file size and our newspapers contain hundreds of classified ads per day.

[12] Appendix E compares our monthly counts with and without seasonal adjustment.

[13] In 1900, the *New York Times* had a circulation of around 40 thousand, far below the city's leading papers, which were close to 180 thousand. Circulation rose sharply between 1912 and 1916, from 175 thousand to 320 thousand, narrowing the gap with the leading newspapers in the city.

Figure 7: City-Level Help-Wanted ads, Jan. 1900 to Dec. 1938



Note: Seasonally adjusted monthly averages. NBER recession dates in grey. *New York Times*: the ad counts are reported from 1919 onward due to low circulation in prior years. *Washington Post*: posting ads was free between November 1905 and July 1916, as highlighted by the vertical dotted lines.

Table 2: Pairwise Correlations of Help-Wanted Ads with the *Chicago Tribune*

	Levels		Per Capita (1927=100)	
	Full	Post-WWI	Full	Post-WWI
New York Times	0.40	0.92	0.95	0.95
Los Angeles Times	0.79	0.91	0.83	0.92
Atlanta Constitution	0.72	0.89	0.72	0.91
Washington Post	0.46	0.87	0.53	0.89

Notes. Pairwise correlations of seasonally adjusted 3-month moving averages with the *Chicago Tribune*. "Levels" uses seasonally adjusted 3-month moving averages. "Per Capita" divides counts by city population (linearly interpolated between decennial censuses) and indexes to the 1927 annual average = 100. "Full" uses all available months for each city pair; "Post-WWI" restricts to January 1919 onward. For the *New York Times* we beging the adjusted series in January 1919.

Abstracting from idiosyncratic city-level movements, help-wanted counts display similar cyclical patterns across cities between 1900 and 1938: firm advertising activity was highly correlated across cities over the first decades of the twentieth century. The first two columns of Table 2 report pairwise correlations with the *Chicago Tribune*, which we use as the reference city because of the consistent quality of its help-wanted data over the full sample. The last two columns report the same correlations after adjusting for city populations, based on interpolating population estimates in decadal censuses, and rescaling each city measure to 100 in 1927. In each case, we report the moments first for the full sample and then for the post-World War I (WWI) subsample.

The data also share a common set of major trends. Beginning in mid-1918, the number of help-wanted ads rose very rapidly in every city, peaking around 1920. This surge reflected a significant labor shortage tied to the conscription of young men into military service during World War I, together with business disruptions from non-pharmaceutical interventions adopted during the 1918 flu (Correia *et al.*, 2022). The boom came to an end during the recession of 1920–21, when the number of help-wanted ads declined sharply across all cities. After the recession, the series stabilized in New York City, Chicago, and Los Angeles, while a downward trend persisted past 1920 in the two smaller cities, Atlanta and Washington, D.C.

All cities experienced a sharp decline in help-wanted ads during the Great Depression, with some regional variation. New York, Chicago, and Los Angeles share a common pattern: large declines at the onset of the Depression, notable recoveries through the mid-1930s, and a renewed downturn in 1936–37. The fall in labor demand at the start of the Depression was sharpest in New York and Chicago, while declines in Los Angeles and Washington were more gradual. The initial decline in the *Atlanta Constitution* was more severe but briefly recovered to its pre-Depression level in the summer of 1931.[14] These differences are consistent with the annual employment patterns documented by Wallis (1989): the South experienced milder employment declines and a faster recovery than the rest of the U.S.

Comparisons with Other City-level Measures Our city-level measures provide high-frequency, geographically disaggregated indicators of labor demand for a period in which such data are scarce. To validate them, we compare our measures to the few city-level indicators of economic and labor-market conditions available over part of our sample.

First, monthly city-level help-wanted registrations at public employment agencies between January 1924 and January 1932 provide an alternative measure of local labor demand. Figure F2 compares the two series for the five cities. They largely share similar trends and cyclical movements: both decline during recessions and recover during expansions. This suggests that the help-wanted measures from classified ads capture similar information about local labor demand.

Second, *Bradstreet's* weekly "Trade at a Glance" tables provide city-level one-word summaries of business conditions, a potentially comparable indicator of local economic activity. We follow Correia *et al.* (2022) to construct business condition indicators, extending their measures over a broader time frame. To do so, we categorize the one-word summaries, provide a numerical value to each category, and then aggregate the measure into a monthly frequency. Figure F3 compares the *Bradstreet* indicator to our help-wanted measure for New York City, Chicago, Los Angeles, and Atlanta.[15] The two series share similar cyclical fluctuations, although the *Bradstreet* indicator is notably more volatile. For New York City, the *Bradstreet* indicator displays a steady decline after 1923 that misses several major cyclical fluctuations.

3.2 City-level Counts by Gender

Labor ads were grouped by gender in the early twentieth century. Figure 8 reports the city-level help-wanted indices by gender. Counts of help-wanted ads for men are generally higher

[14] Relative to its level at the onset of the Depression (August 1929), the Atlanta series bottomed out at 33% in January 1931.

[15] *Bradstreet's* did not provide information for Washington, D.C. Appendix F.4 provides further details on the indicator's construction.

Table 3: Help-Wanted Volatility by Gender

	Std. dev.			Corr. with U	
	Men	Women	Corr(Men, Women)	Men	Women
New York	0.18	0.26	0.59	-0.47	-0.71
Chicago	0.22	0.35	0.92	-0.53	-0.44
Los Angeles	0.19	0.21	0.63	-0.44	-0.53
Atlanta	0.31	0.31	0.22	-0.11	-0.52
Washington D.C.	0.35	0.35	0.86	-0.49	-0.37

Notes. Monthly series are converted to quarterly averages, expressed as percentage deviations from the sample mean, and filtered using a band-pass filter retaining fluctuations between 2 and 50 quarters. Standard deviations and correlations are computed on the resulting cyclical components. U is the national civilian unemployment rate. New York data begin in January 1919.

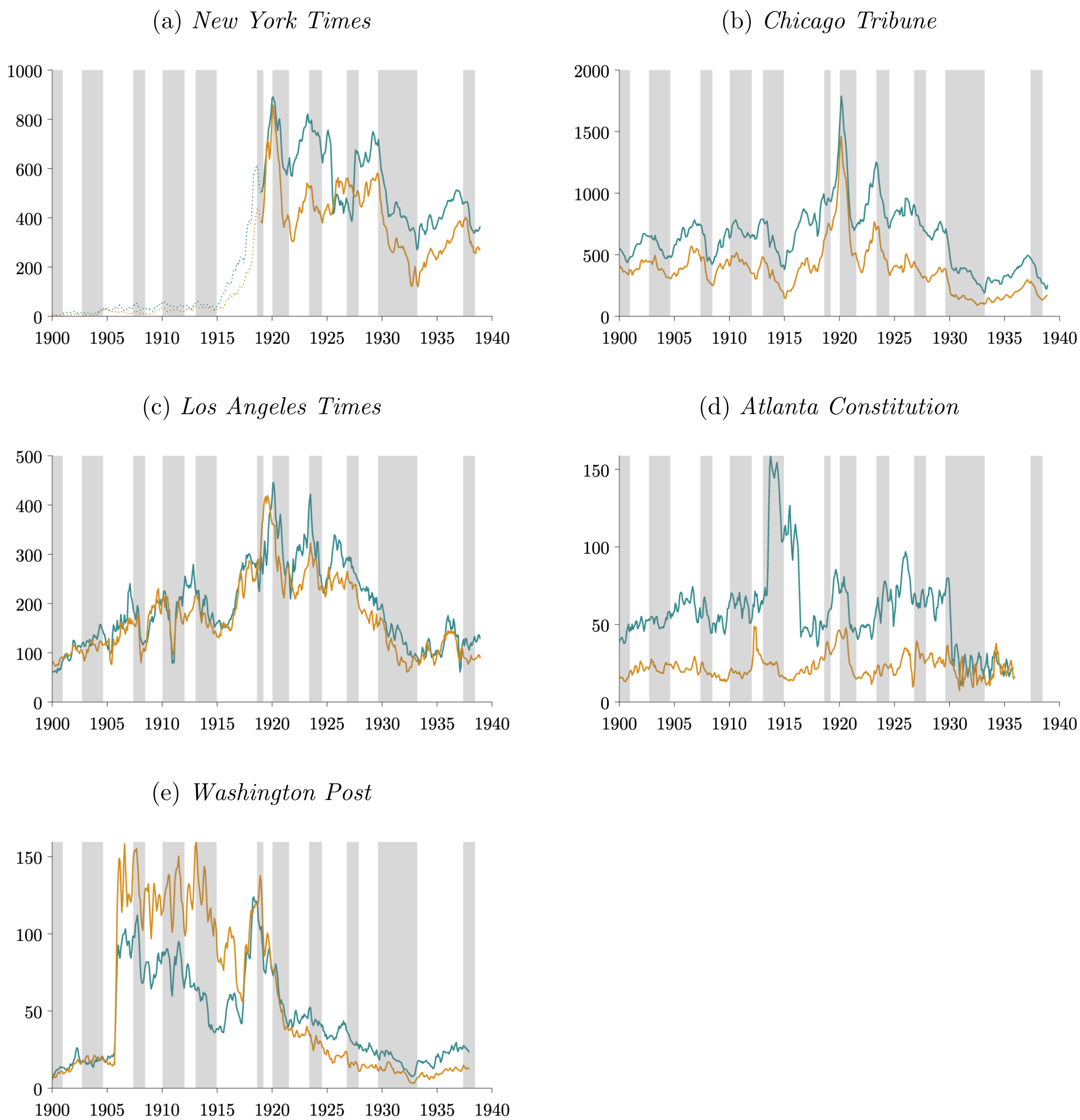
than those for women. In the *Chicago Tribune*, for example, the average daily number is 643 for men and 380 for women. The gap is even more pronounced for the *Atlanta Constitution*, where ads for women are particularly scarce relative to other cities. Patterns vary regionally, however. In Los Angeles, ad counts for women and men are much closer to each other. In Washington, the number of ads for women exceeded the number for men during the period when classifieds were free, but had fallen below those for men by 1921.

Beyond level differences, the gender split of help-wanted ads diverges sharply from the gender composition of employment. Table F1 compares the average Sunday ratio of female to male ads in each city-year with the corresponding ratio of gainful workers from the U.S. Census. On average, the ratio of women to men in employment is around 0.25 – one woman for every four male workers – while the ratio in help-wanted ads is closer to 0.75.

In interpreting these high female ratios, an important caveat is potential sectoral bias in newspaper coverage. Many male-dominated occupations, including in manufacturing and mining, historically relied on alternative recruitment methods such as union referrals, direct hiring at the work site, or informal networks. Appendix B reports the distribution of occupations listed in the *New York Times*, *Chicago Tribune*, and *Los Angeles Times* on May 11, 1930. Compared to the U.S. Census, the occupational structure of newspaper ads likely over-represents sectors in which women were concentrated and under-represents blue-collar and industrial male employment.

We next turn to the cyclical properties of help-wanted ads by gender. Table 3 reports, for

Figure 8: City-Level Help-Wanted ads by Gender



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

each city-gender pair, the standard deviation of the cyclical component of help-wanted ads, the correlation between the cyclical components of male and female help-wanted, and the correlation of each gender series with the national unemployment rate. Female help-wanted ads are generally more volatile than male ads, and the two series tend to move closely together. In Chicago, for example, the volatility of female help-wanted is nearly 60 percent greater than the volatility for men, and the two series have a contemporaneous correlation of 0.92. Atlanta is the largest outlier from this pattern: volatilities for men and women are similar, and the two series have a low correlation of 0.22.

These gender differences are visible in specific episodes. During the 1920 recession, the number of help-wanted ads in the *New York Times* fell from a peak above 800 to 600 for men, while the decline was more pronounced for women, from 800 to under 400. In the *Los Angeles Times*, help-wanted ads for women rose more rapidly than those for men following the Spanish flu and WWI.

In addition, potential demand for female labor could have been higher than paid advertising captured. The *Washington Post*'s period of free classified ads (1905–1916) saw a surge in advertisements for female workers relative to men, suggesting that the cost of posting ads likely discouraged employers, particularly households and employers hiring for lower-wage, female-dominated roles, from advertising regularly.[16] The surge in female postings once that barrier was removed points to a latent demand for female workers that was not previously reflected in paid advertising. In the next subsection, we examine more directly how advertising costs influenced firm behavior.

3.3 Help-wanted and Ad Posting Costs

Three newspapers in our sample—the *New York Times*, the *Los Angeles Times*, and the *Washington Post*—systematically published information on advertising costs, as discussed in Section 2.[17] We use this information to estimate how firms respond to changes in posting costs through the following panel regression:

$$\ln(HW_{s,t}) = \alpha_t + \eta_s + \beta \ln(Ad_{s,t}) + \gamma X_{s,t} + \varepsilon_{s,t}, \tag{2}$$

[16] Wage gaps in this period are well documented by Goldin (1990) and Kessler-Harris (1982).

[17] Systematic data on posting costs for U.S. newspapers are only available starting in 1951; Zagorsky (1993) studies how help-wanted ads have varied with posting costs since that date.

Table 4: Elasticity of Help-Wanted Ads to Posting Costs

	All		*Men*		*Women*	
		+ U.		*+ U.*		*+ U.*
β	-0.09** (0.03)	-0.10** (0.03)	-0.12 (0.04)	-0.12 (0.05)	-0.12** (0.03)	-0.12** (0.03)
β_U		-0.005*** (0.001)		-0.005 (0.003)		-0.006*** (0.000)
Obs.	825	825	825	825	825	825
Adj. R^2	0.90	0.90	0.84	0.85	0.87	0.87

Notes: Log-log regressions of $\ln(\mathrm{HW}_{s,t})$ on $\ln(\mathrm{Ad}_{s,t})$, the contemporaneous log posting cost. The "+ U." columns add the national unemployment rate. All specifications include city fixed effects, a time trend and its square, and two lags of the log dependent variable. Standard errors clustered by city in parentheses. Data sample is January 1915 to December 1938; Washington Post free-ads months excluded. Statistical significance indicated by ** ($p < 0.10$) and ***($p < 0.05$).

where $HW_{s,t}$ is the monthly average of Sunday help-wanted counts in city s in month t, prior to seasonal adjustment and 3-month averaging. $Ad_{s,t}$ is the corresponding labor-ad posting cost.[18] The coefficient β is the elasticity of help-wanted advertising with respect to posting costs. The term α_t includes a linear and quadratic time trend, controlling for factors common across cities in a given month; η_s is a city fixed effect that controls for city-specific time-invariant factors; and $X_{s,t}$ contains lags of the help-wanted count. $\varepsilon_{s,t}$ is an idiosyncratic error term. Because posting costs changed infrequently before 1915, we focus on the period from 1915 to 1938. Appendix Figure C1 displays the dates and values of price changes for each newspaper.

Table 4 reports the results. The first column presents the estimated elasticity for all help-wanted ads, which is -0.09, statistically significant at the 10 percent level but economically small: a 10 percent increase in the cost of posting is associated with about a 1 percent decline in help-wanted ads. The third and fifth columns report separate estimates for male and female ads. The elasticity is -0.12 for both, though the estimate is statistically significant at the 10 percent level for women but not for men. The second, fourth, and sixth columns add the national unemployment rate as a control. The ad rate elasticities are virtually unchanged, providing reassurance that the estimated response to posting costs is not driven by business-cycle comovement between ad rates and labor demand. The unemploy-

[18] We use a log-log specification because a level-on-level specification is inappropriate given the large differences in ad counts across cities.

ment rate itself is strongly significant for women but not for men, consistent with greater cyclical sensitivity of female labor demand. Overall, the demand for help-wanted advertising is inelastic, with an estimated elasticity of -0.09.

4 National Help-Wanted Index

In this section, we construct a national help-wanted index by aggregating labor ad counts across five cities and discuss the cyclical dynamics of the labor market between 1900 and 1938.

4.1 Constructing the National Index

We construct the national help-wanted index as:

$$NHWI_t = 100 \times \frac{\sum_{c=1}^{C} HW_{c,t}\omega_{c,t}}{\sum_{c=1}^{C} HW_{c,t=1927:1}\omega_{c,t=1927:1}} \times \frac{LF_{t=1927:1}}{LF_t}, \tag{3}$$

where C is our set of metropolitan areas; $HW_{c,t}$ is the seasonally adjusted and smoothed monthly ad count from area c; $\omega_{c,t}$ is a weight for area c equal to its share of the five-city population in period t, obtained by interpolating between decennial censuses; and LF_t is an estimate of the overall labor force. The index is normalized to 100 in January 1927. The resulting series are shown in Figure 9.[19] Panel (a) displays the aggregate index; panel (b) shows the breakdown by gender.

Our aggregate help-wanted series, combined with existing proxies for job vacancies in the remaining decades of the twentieth century, yields 125 years of monthly labor-demand indicators.[20] Panel (c) of Figure 9 shows this combined series, linking our national index with the job vacancy rate from Petrosky-Nadeau and Zhang (2021) to cover the period from 1938 to December 2024.

4.1.1 External Validation

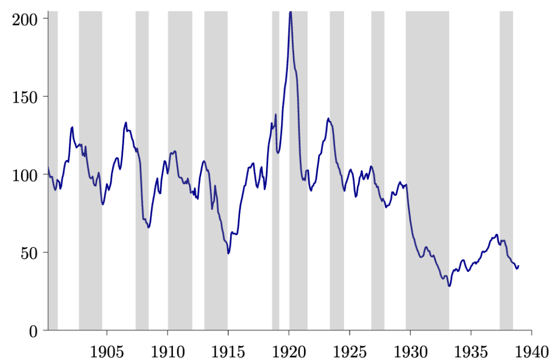
Our approach of building the national help-wanted index mirrors MetLife's methodology. The MetLife index was initiated in 1927, with past issues

[19] When constructing the aggregate index, we use *New York Times* counts only from January 1919 onward, given the paper's low earlier circulation. We also drop Atlanta after December 1935 and Washington after December 1937 for lack of help-wanted data. See Section 3 for further details.

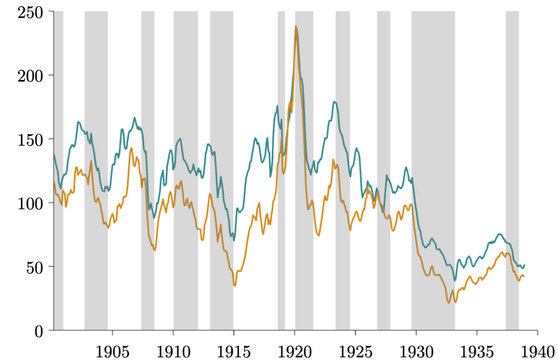
[20] The vacancy rate for 1900–1927 is constructed by rescaling our help-wanted index to match the level of the national vacancy rate series constructed by Petrosky-Nadeau and Zhang (2021) at December 1927.

Figure 9: Aggregate Help-Wanted Index

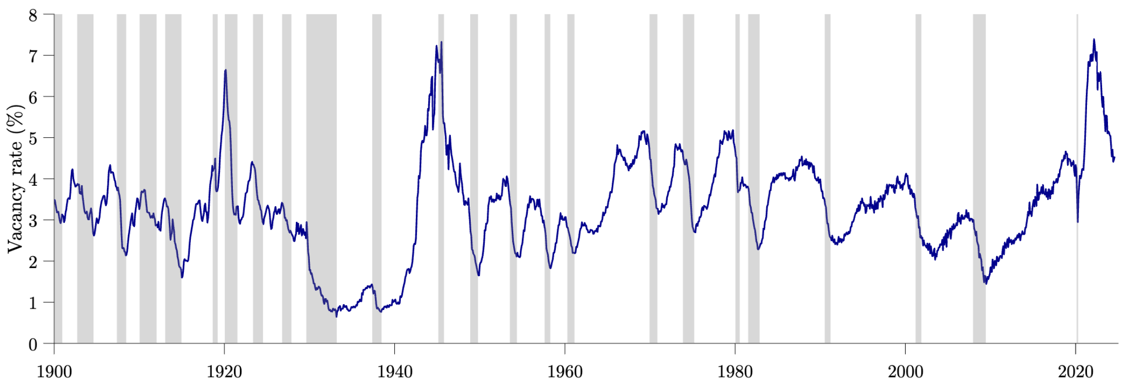
(a) Help-Wanted Index



(b) Help-Wanted Index by Gender



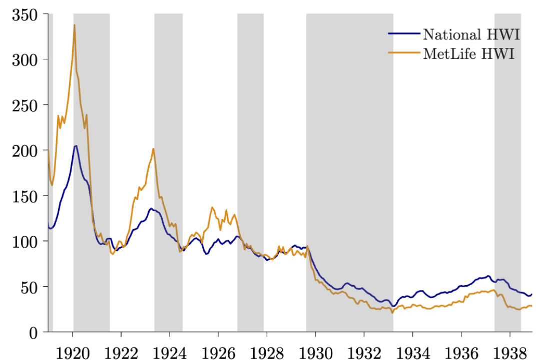
(c) 125 years of help-wanted



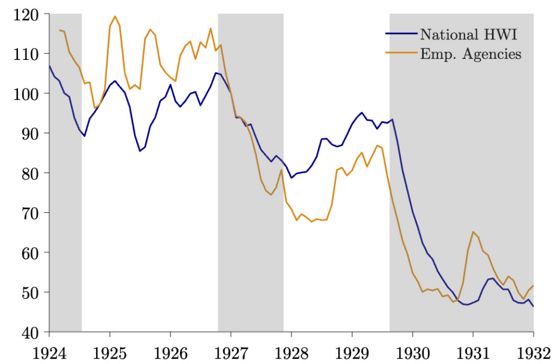
Note: The help-wanted indices are normalized to Jan. 1927=100. NBER recession dates in grey. Panel (b): yellow line is for female labor ads and green line is for male labor ads.

Figure 10: External Validation

(a) Comparison with MetLife



(b) Comparison with Public Employment Agencies



Note: Panel (a) The city-weighted help-wanted Index is defined in equation (3). Both series are normalized to Jan. 1927=100 for comparability. NBER recession dates in grey. Panel (b) Red lines are the city-weighted public employment agency series, and the blue lines are the city-weighted newspaper labor indices as defined in equation (3). They include New York City, Chicago, Los Angeles, Atlanta, and Washington DC. Both series are normalized to Jan. 1927=100 for comparability.

of print newspapers collected to extend the series back to 1919.[21] The shared methodology allows us to validate our national index by comparing it to MetLife's during the period of overlap. Relative to the MetLife program, data limitations restrict us to a subset of newspapers, but our five cities still provided substantial national coverage, more than 33 percent of all U.S. city populations according to the 1930 Census.

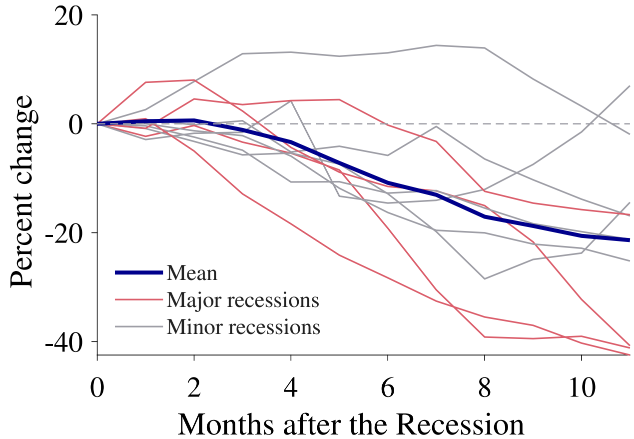
As shown in Figure 10a, our help-wanted index tracks the MetLife series closely during the overlapping period, with a correlation of 0.98 between the two. Our measure is less volatile during 1919 - 1927 period, in which the accuracy of the MetLife data has been questioned (Zagorsky, 1998). Specifically, our measure mirrors the MetLife business-cycle troughs but not the elevated peaks during expansions, which had previously been cited as a reason for concern with the reliability of the MetLife index prior to 1927.

As a second external check, we collect data from public employment agencies. From January 1924 through January 1932, the U.S. Department of Labor, in cooperation with state and local employment services, produced monthly reports of the number of help-wanted registrations from local businesses.[22] We digitized these reports to construct an alterna-

[21]This 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 used in the vacancy index were incorporated in 1927 (Berridge, 1929).

[22]According to Lee (2007), about four percent of the U.S. labor force searched for jobs through public employment offices during this period; users were mainly unskilled workers.

Figure 11: Dynamics of Help-Wanted Index Following the Start of a Recession



Note: Percent change in the help-wanted index from the start of a recession. Major recessions, following the classification of Bordo and Haubrich (2017), include 1907, 1920, 1929, and 1937.

tive measure of firm search. The key difference between the two channels is that posting with a public employment agency was free, while posting an ad in a newspaper required a fee.

Figure 10b compares our newspaper-based help-wanted index with the corresponding series from public employment agencies, both derived from the same five cities in our sample. The yellow line represents the employment agency data, while the blue line depicts our index based on newspaper advertisements. The two series track each other very closely throughout the period of overlap.

4.2 Dynamics of Firm Search

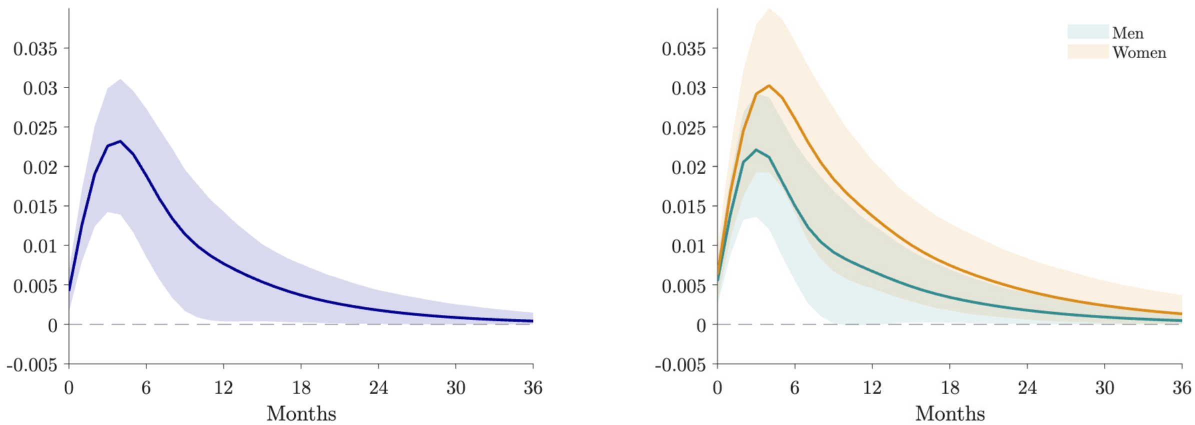
The aggregate help-wanted index displays strong procyclical dynamics. Panel (a) of Figure 9 shows that the index declines rapidly at the onset of a recession and recovers during the subsequent expansion, a pattern similar to the behavior of job-posting series in the post-war period. The expansion beginning in March 1919, after the most severe wave of the Spanish flu in late 1918 and the end of WWI, was accompanied by an extraordinary surge in help-wanted advertising. The index peaked on the eve of the 1920 recession, drawing parallels with the spike in job openings during the recovery from the COVID-19 pandemic in 2020.

Figure 11 further illustrates these points by plotting the help-wanted index relative to its level at the start of each recession. The blue line plots the mean measure across recessions; red lines denote the "major recessions" identified by Bordo and Haubrich (2017), and gray

Figure 12: VAR Impulse Response of Help-Wanted Index to Productivity Shock

(a) National Help-Wanted Index

(b) Men vs Women



Note: Impulse response of the help-wanted index to a labor productivity shock from bivariate VAR(4) models on band-pass filtered cyclical components, 1900–1938. Panel (a): national help-wanted index. Panel (b): separate models on the men's and women's indices. Structural identification via Cholesky decomposition with productivity ordered first. Shaded areas denote 90% bootstrap confidence intervals.

lines mark the remaining recessions. The four major recessions are: (i) the 1907 recession, following a stock market crash and banking panic; (ii) the 1920 recession, 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 1937 recession, the third-worst downturn of the twentieth century and often attributed to monetary policy. The figure shows that recessions look similar at their onset, but severe recessions are followed by sharper declines and slower recoveries in help-wanted advertising.

We complement this descriptive evidence with a structural VAR to characterize the dynamic response of help-wanted advertising to productivity shocks. Specifically, we estimate bivariate vector autoregressions (VARs) on the cyclical components of labor productivity and the national help-wanted index. The cyclical components are extracted by applying a band-pass filter to the log of each series and retaining fluctuations with periods between 2 and 72 months. We identify the structural shock to labor productivity using a Cholesky decomposition with productivity ordered first, and include four lags. Panel (a) of Figure 12 presents the impulse response of the aggregate help-wanted index. Following a positive productivity shock, firm search activity rises, peaking after about 3 months and then gradually returning to trend. The response is statistically significant for the first 11 months, consistent with the procyclical nature of help-wanted advertising documented above.

The response of help-wanted advertising for women to a productivity shock is roughly twice as large as for men. Panel (b) reports the same exercise estimated separately on the men's and women's help-wanted indices. Both indices rise following a positive productivity shock, but the responses differ notably: the women's help-wanted index response remains statistically significant for approximately the entire horizon, while the men's index shows a more muted peak and wider confidence intervals, losing signifance after 8 months.

4.3 Business Cycle Moments

We report standard business-cycle moments for the early twentieth-century labor market. We remove longer-run trends from each time series at a quarterly frequency and focus on second moments of the resulting cyclical components.[23] The moments of interest—standard deviations and contemporaneous cross-correlations—are computed as proportional deviations from trend. Table 5 reports these moments for the full sample period (panel A) and for the subsample from January 1900 to October 1929, which excludes the Great Depression (panel B). Panel C reports the same moments for the post-World War II sample from 1955 to the end of 2024; details on the post-WWII data are in Appendix F.

Unemployment and job postings are volatile over the business cycle in our full pre-1938 sample. The standard deviation of the cyclical component of unemployment is 0.44, more than twice the 0.19 standard deviation of help-wanted ads. The contemporaneous correlation between unemployment and vacancies, which captures the slope of the Beveridge curve, is close to -0.6 as shown in Table 5. Labor market tightness is strongly procyclical, with a correlation of 0.40 with labor productivity.

The Great Depression, with its large departures from normal business-cycle fluctuations, could distort the characterization of the cyclical dynamics of the labor market. Panel B of Table 5 addresses this issue by reporting the same set of moments for the 1900–1929 subperiod. The patterns are markedly similar to the full sample, though the volatility of unemployment and labor tightness is lower when we exclude the Great Depression, and the slope of the Beveridge curve is somewhat steeper.

For comparison, panel C of Table 5 presents the moments of variables in the post-World War II period (1955-2024). Consistent with the literature, the volatility of all variables is

[23] We extract the cyclical component by retaining fluctuations at horizons between 2 and 50 quarters using a band-pass filter.

Table 5: Empirical Business Cycle Moments, Quarterly

Panel A: 1900:I to 1938:IV

		U	HWI	θ	X
Std. dev.		0.44	0.19	0.62	0.06
Corr. matrix	U	1.00	-0.59	-0.68	-0.37
	HWI		1.00	0.70	0.36
	θ			1.00	0.40

Panel B: 1900:I to 1929:III

		U	HWI	θ	X
Std. dev.		0.37	0.18	0.54	0.05
Corr. matrix	U	1.00	-0.64	-0.87	-0.38
	HWI		1.00	0.66	0.36
	θ			1.00	0.45

Panel C: Jan. 1955 to Dec. 2024

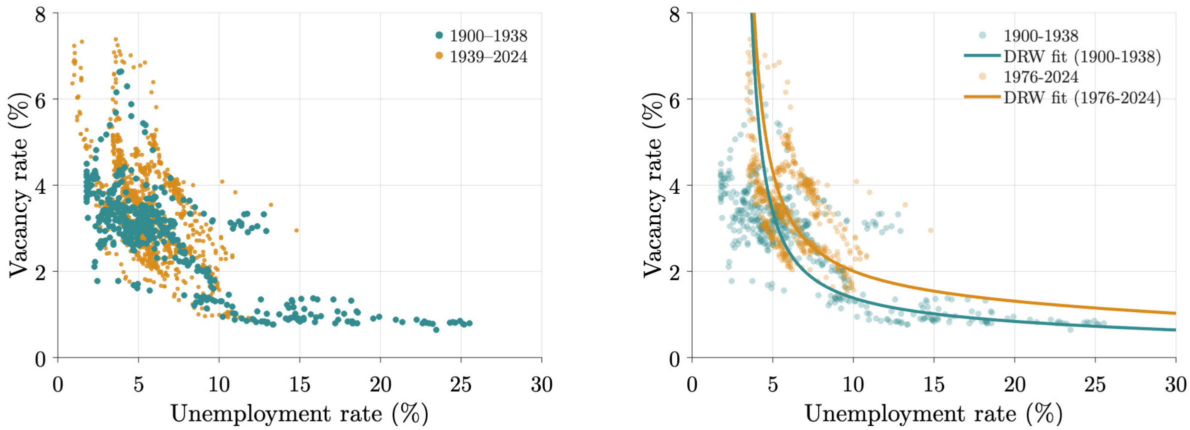
		U	HWI	θ	X
Std. dev.		0.20	0.17	0.32	0.02
Corr. matrix	U	1.00	-0.74	-0.84	0.19
	HWI		1.00	0.90	0.24
	θ			1.00	-0.04

Notes. Monthly series are converted to quarterly averages, expressed as percentage deviations from the sample mean, and filtered using a Baxter-King band-pass filter retaining fluctuations between 2 and 50 quarters. U is the civilian unemployment rate; HWI is the national help-wanted index in panels A and B and the JOLTS vacancy rate in panel C; $\theta = HWI/U$ is labor market tightness; X is labor productivity. Panel A covers the full sample (1900:I–1938:IV), panel B excludes the Great Depression (1900:I–1929:III), and panel C uses post-war data (1955:I–2024:IV).

Figure 13: The U.S. Beveridge curve and matching function: A long run view

(a) Empirical Beverdge curve

(b) Fitted matching function



Note: Panel (a) plots unemployment and vacancy rates between Jan. 1900 and Dec. 2024. Panel (b) plots the Beveridge curves implied from estimating a matching function over pre- and post-WWII samples.

considerably lower in the modern era. The variables nonetheless display correlation patterns that are similar to those of the early twentieth century. Notably, the volatility of labor-market tightness relative to that of unemployment is similar across sample periods.

4.4 Beveridge Curve and A Labor Market Matching Function

Our new data provide a longer historical view of the Beveridge curve. Figure 13a plots the vacancy rate against the unemployment rate from January 1900 to December 2024. The scatter plot in panel (a) contrasts our new data series from January 1900 to December 1938, shown in teal dots, with data covering the end of the Depression and the post-war period between January 1939 and December 2024, shown in amber.

Consistent with more recent periods, the scatter for the early twentieth century displays a downward-sloping, convex relationship between the unemployment rate and the vacancy rate. The 1900–1918 portion overlaps remarkably with the post-WWII experience of movements in vacancies and unemployment. The 1919–1938 portion, by contrast, was marked by extraordinary events – World War I, the Spanish flu, and the Great Depression – and exhibits notable shifts in the Beveridge curve that are not typical for business cycle movements. The high unemployment and low vacancy rates of the Great Depression, for example, trace a particularly flat portion of the curve, as noted in Petrosky-Nadeau and Zhang (2021). Our novel data from 1900 to 1918 thus help reveal a remarkable stability in the U.S. Beveridge

Table 6: Matching Function Estimates

	Pre-WWII 1900–1938	*Post-WWII 1976–2024*
χ (level)	3.97	2.25
ι (curvature)	0.64	0.72
δ	0.027	0.027
Obs.	468	584
Mean η_U	0.57	0.55
Median η_U	0.63	0.55

Notes. Nonlinear least squares estimates of the den Haan *et al.* (2003) matching function. δ is the monthly separation rate, set to the JOLTS average (2000–2024). $\eta_U = \chi^\iota/(\chi^\iota + \theta^{-\iota})$ is the elasticity of matching with respect to unemployment.

curve over the past 125 years, at least as measured by the estimated matching elasticities, a stability that would be obscured by relying only on the existing MetLife vacancy series.

We investigate this stability of the Beveridge curve further by estimating a labor market matching function on pre- and post-WWII data. Starting from the steady-state relationship between the unemployment rate and the flows into and out of unemployment:

$$U = \frac{\delta}{\delta + M(V,U)/U} \tag{4}$$

where δ is the flow rate into unemployment; and $M(V,U)/U$ is the flow rate out of unemployment, determined by the matching of unemployed workers U with vacant jobs V through the matching function $M(V,U)$. Following den Haan *et al.* (2003), we parameterize the matching function as $M(V,U) = \chi V U \left((\chi V)^\iota + U^{-\iota}\right)^{-1/\iota}$, where χ is a matching level shifter and ι is a curvature parameter. This functional form has the desirable property that matching probabilities are bounded between zero and one. We fit equation (4) to data by nonlinear least squares. The separation rate δ is set to the average monthly total-separations rate from JOLTS over 2000–2024, $\delta = 0.027$, for both samples. Table 6 reports the estimates and panel (b) of Figure 13a displays the implied Beveridge curves.

To compare the implied labor-market dynamics across eras more directly, we compute the elasticity of matching with respect to unemployment. Under our chosen functional form, the elasticity is given by $\eta_U = \chi^\iota/(\chi^\iota + \theta^{-\iota})$, where $\theta \equiv V/U$ is labor-market tightness. The elasticity η_U measures the responsiveness of the job-finding rate to unemployment, holding vacancies fixed. The mean elasticities are remarkably similar across eras (0.57 versus 0.55). This similarity suggests that the underlying process of matching workers to jobs has been broadly stable for over a century, despite substantial changes in labor-market institutions and search technologies. The larger gap between the mean and median elasticities in our data sample reflects the influence of the Great Depression: very low levels of labor-market tightness generate low matching elasticities that pull the pre-war mean below its median. Taken together, the stability of the Beveridge curve and the similarity of matching elasticities across eras imply that the fundamental process of matching workers to jobs has remained broadly similar across 125 years of U.S. labor-market history.

5 Conclusion

This paper provides new monthly, city-level and national measures of firm search based on the help-wanted advertisements in early-twentieth century U.S. newspapers. We extract employer-placed advertisements using image-processing techniques and use them to construct help-wanted indices for 1900 to 1938. We validate the methodology with extensive manual checks and by comparison with national and city-level indicators of help-wanted activity available for subperiods of our sample.

Help-wanted advertisements vary substantially over our sample period. Firm search effort is procyclical, with vacancy postings declining sharply at the onset of recessions. We also document a striking stability in the U.S. Beveridge curve over the past 125 years. Disaggregated patterns highlight the importance of regional and demographic context: there is meaningful variation across cities and greater cyclical sensitivity in advertisements placed for women.

Our methodology is adaptable to other newspapers, scalable to additional dimensions of the classified advertisements, and can be enriched with textual analysis to extract deeper economic insights from historical listings. We hope that it encourages further research using archived newspapers to study historical economic dynamics.

References

Alexopoulos, M. and Cohen, J. (2016). The Medium Is the Measure: Technical Change and Employment, 1909-1949. *The Review of Economics and Statistics*, **98** (4), 792–810.

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.

Bezanson, A. (1929). *Help-Wanted Advertising as an Indicator of the Demand for Labor*. University of Pennsylvania Press.

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.

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.

den Haan, W. J., Ramey, G. and Watson, J. (2003). Liquidity flows and fragility of business enterprises. *Journal of Monetary Economics*, **50** (6), 1215–1241.

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

Fishback, P. V. (1992). *Soft Coal, Hard Choices: The Economic Welfare of Bituminous Coal Miners, 1890-1930*. Oxford University Press.

— (1998). Operations of 'Unfettered' Labor Markets: Exit and Voice in American Labor Markets at the Turn of the Century. *Journal of Economic Literature*, **36** (2), 722–765.

Fujita, S. and Ramey, G. (2009). The cyclicality of separation and job finding rates. *International Economic Review*, **50** (2), 415–430.

Goldin, C. and Katz, L. F. (1998). The Origins of Technology-Skill Complementarity. *The Quarterly Journal of Economics*, **113** (3), 693–732.

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

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

Kessler-Harris, A. (1982). *Out to Work: A History of Wage-Earning Women in the United States*. Oxford University Press.

Kirchhoff, S. M. (2011). The u.s. newspaper industry in transition. CRS Report No. R40700.

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.

Merz, M. (1995). Search in the labor market and the real business cycle. *Journal of Monetary Economics*, **36** (2), 269–300.

Pedriana, N. and Abraham, A. (2006). Now you see them, now you don't: The legal field and newspaper desegregation of sex-segregated help wanted ads 1965–75. *Law & Social Inquiry*, **31** (4), 905–938.

Petrongolo, B. and Pissarides, C. A. (2001). Looking into the black box: A survey of the matching function. *Journal of Economic Literature*, **39** (2), 390–431.

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

Ramey, V. A. and Francis, N. (2009). A century of work and leisure. *American Economic Journal: Macroeconomics*, **1** (2), 189–224.

Rosenbloom, J. L. (1994). Looking for Work, Searching for Workers: U.S. Labor Markets after the Civil War. *Social Science History*, **18** (3), 377–403.

Schweinitz, D. D. (1932). *How Workers Find Jobs: A Study of Four Thousand Hosiery Workers in Philadelphia*. University of Pennsylvania Press.

Shen, Z., Zhang, R., Dell, M., Lee, B. C. G., Carlson, J. and Li, W. (2021). Layoutparser: A unified toolkit for deep learning based document image analysis. In *International Conference on Document Analysis and Recognition (ICDAR)*, pp. 131–146.

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American Economic Review*, **95** (1), 25–49.

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.

Wallis, J. J. (1989). Employment in the Great Depression: New data and hypotheses. *Explorations in Economic History*, **26** (1), 45–72.

Zagorsky, J. L. (1993). Understanding the economic factors affecting help-wanted advertising. *Journal of Advertising*, **22** (3), 75–84.

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

Online Appendix for: Firm Search in the Labor Market: Evidence from Help-Wanted Advertisements

Huixin Bi, Nicolas Petrosky-Nadeau, Nora Traum, and Greg Woodward

May 15, 2026

A Acknowledgments

We are deeply grateful to the many extraordinary individuals whose combined efforts made this project possible.

We would like to acknowledge Brandyn Bok, Shane Boyle, Erin Crust, Joël Cuerrier, Owen Engbretson, August Hays-Ekeland, Sandrine Jean, Rhiwoo Kim, and Nish Sinha for their excellent contributions to exploring image processing and textual analysis methods as well as developing newspaper templates.

We also acknowledge the following contributors who conducted extensive manual data collection to verify our methods and provide supplemental data for our analysis. We thank undergraduate research assistants at HEC Montréal: Malik Tazi, Samy Djallil Harkat, and Yusra Negm Sarwat; graduate research assistants at HEC Montréal: Christine El-Rami, Aaron Israel, Oussama Marhraoui, and John Ndanwuzwe; research assistants/interns at the Federal Reserve Bank of Kansas City: Phillip An and Eric Lachman; and research assistants at the Federal Reserve Bank of San Francisco: Shane Boyle, Gracie Ellsworth, Mollie Pepper and Addie New-Schmidt.

B Occupations

Table B1 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." 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 impossible to classify the occupation. We count all these types of occupations as in " no category."

The table shows that help-wanted ads covered a broad spectrum of occupations. Ads tend to be dominated by trade occupations, including by advertising, banking, retail jobs, managers, and salesmen.[1] For comparison, the last column of Table B1 provides the distribution of gainfully employed individuals at the state level in the 1930 Census.[2]

C Classified Ad Rates

The cost for posting labor ads varied over time and across newspapers. Many newspapers regularly printed the current rates in their Sunday publications. We manually collected the posting costs for the *New York Times*, the *Los Angeles Times*, and *Washington Post*. Information for the *Chicago Tribune* and *Atlanta Constitution* were not printed in the newspapers. 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. For consistency, we use the one-time posting cost. In addition, costs are typically higher on Sundays compared to other days of the week. We use the Sunday costs as the labor indicators are constructed using Sunday newspapers.

In addition, the specification of posting charges also varied across newspapers. Over the period, the *Washington Post* provided a charge rate either by word or by line, and additionally specified a minimum charge per ad that was above these individual charges. For consistency, we use the minimum charge per ad throughout the sample. The *Los Angeles Times* initially charged ads by the number of words with a minimum fee: 1 cent per word with a minimum charge of 15 cents between 1900 and 1903, and 1 cent per word with a minimum charge of 25 cents between 1903 to 1918. Therefore, we use the minimum charge for that period. After 1918, we change to the per line cost following the information the *Los Angeles Times* published. For the *New York Times*, we use the per line cost throughout the sample.

Figure 2 presents the nominal and real costs over time of posting a help-wanted ad in the three newspapers. To compute the real cost, we use the historical consumer price index (CPI) values. Prior to 1913, we use the annual CPI constructed by Albert Rees and maintained by the Minneapolis Federal Reserve.[3] We assume the price values remained the same within

[1] We categorize all ads for a chauffeur as domestic and personal services, since these ads often are associated with domestic work. 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.

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

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

a year. Since 1913, monthly CPI has been computed by the Bureau of Labor Statistics in a consistent manner.

Figure C1 displays the monthly help-wanted ad counts alongside the dates at which posting rates changed for each newspaper. Red vertical lines indicate rate increases; green lines indicate rate decreases. The annotation on each line shows the magnitude of the change in cents. For the *Washington Post*, the shaded area marks the period during which classified ads were free (November 1905 to July 1916).

D 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.

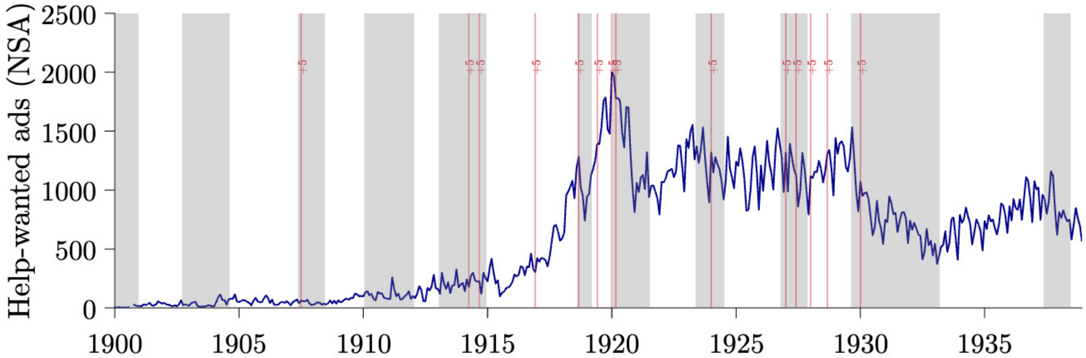
D.1 Image Processing

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.

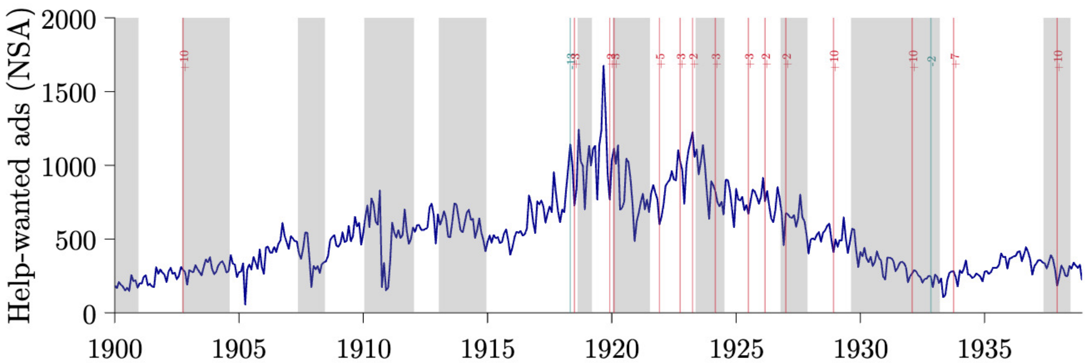
Image Alignment During the digitization process, text documents are often not oriented properly while scanned. Ensuring column/ad dividers are as vertical/horizontal as possible improves accuracy. We use the Separator Net Post Processor from NewsEye (2020) to identify any skew or slant to images and then use the warpAffine method from OpenCV to rotate images accordingly.

Figure C1: Help-Wanted Ad Counts and Ad Rate Changes

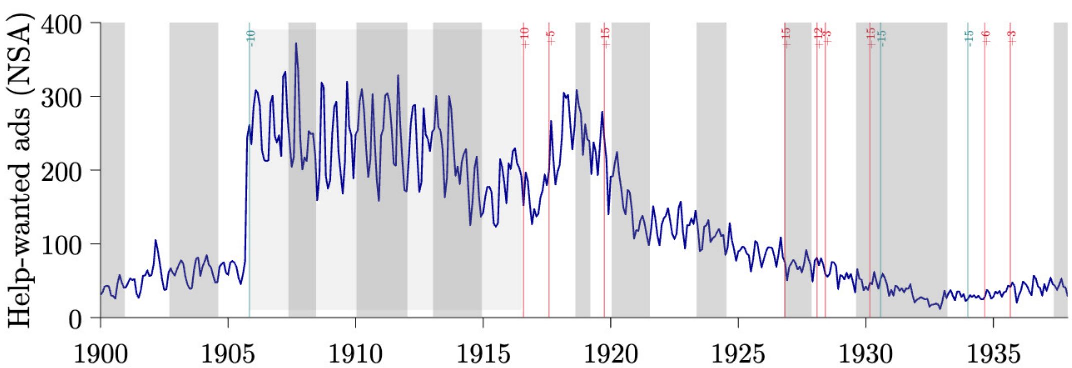
(a) *New York Times*



(b) *Los Angeles Times*



(c) *Washington Post*



Notes: Monthly average of Sunday help-wanted ad counts (not seasonally adjusted). Red vertical lines mark ad rate increases; green lines mark decreases, with annotations showing the change in cents. Shaded green area in panel (c) denotes the Washington Post free-ads period (November 1905 to July 1916). Grey bars are NBER recession dates.

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. 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 D1b, the image colors are inverted, from black to white or from white to black. Inverting the image colors simplifies the separating process by giving the foreground values greater than zero.

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 D1c.

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 D1d. 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 D1d.

   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 a histogram, see the blue line in Figure D2. 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 D1: Identifying ad dividers

(a) Templates (b) Inverted (c) Text Removed (d) Row Means

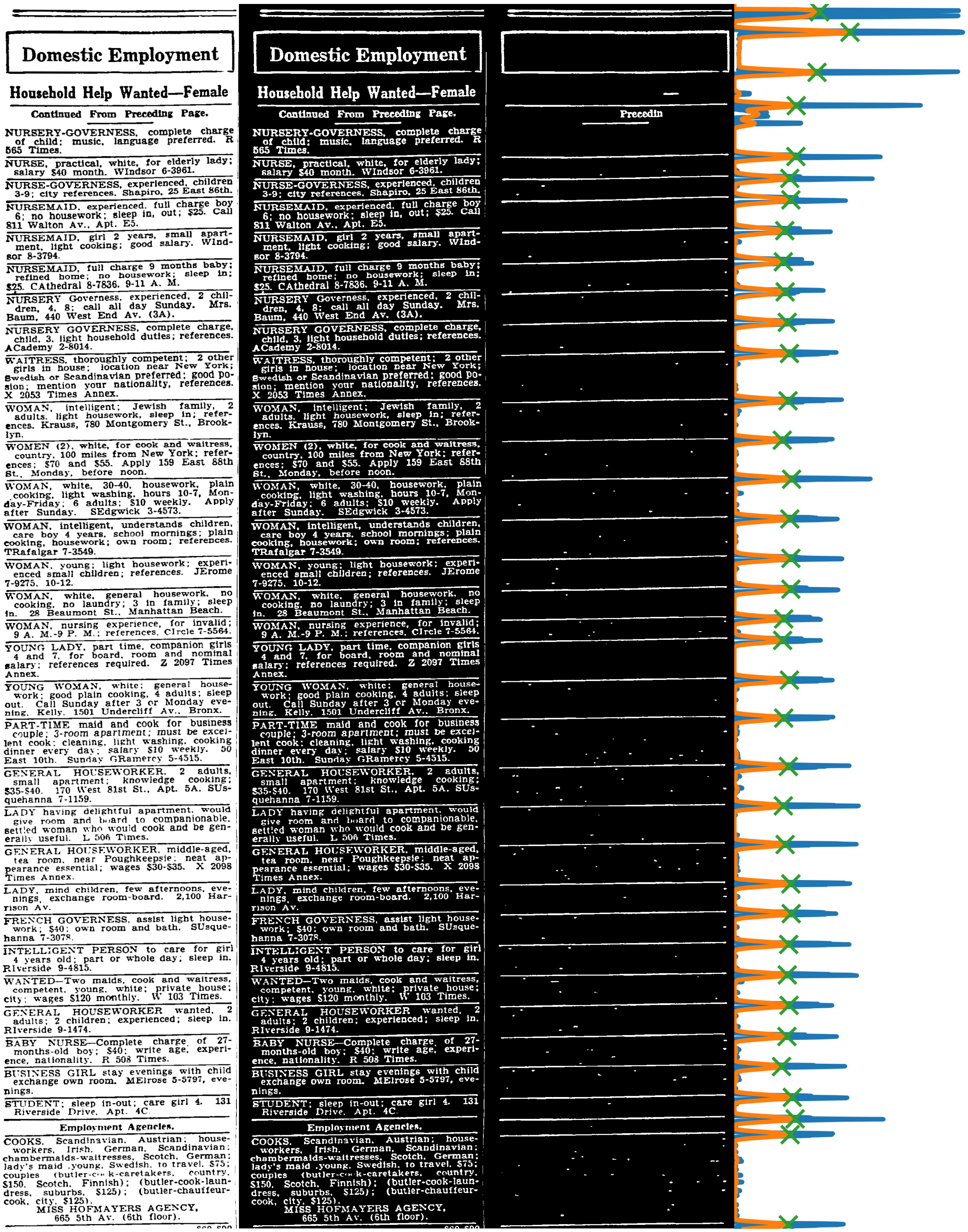


Figure D2: Prominence Threshold Calculation

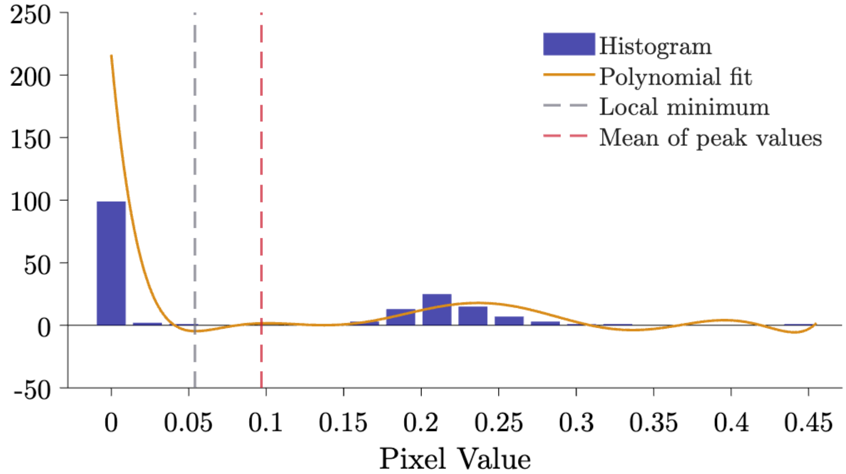


Figure D2. Finally, peaks with a prominence above this threshold are used to divide the column into individual ads, as shown by green cross signs in Figure D1d.

D.2 Additional Details on Template Matching

Newspaper publishers changed the wording and typesetting for section headers over time. Table D1 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. In total, we created 947 templates for headers across newspapers over our entire time frame.

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. In practice, a template is moved across the targeted image, and a match scoring is constructed by measuring the correspondence of the black-and-white pixel locations, 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 D3 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.

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.

2. For each period shown in Table D1, we run the template matching procedure on all

Figure D3: Template size and matching score

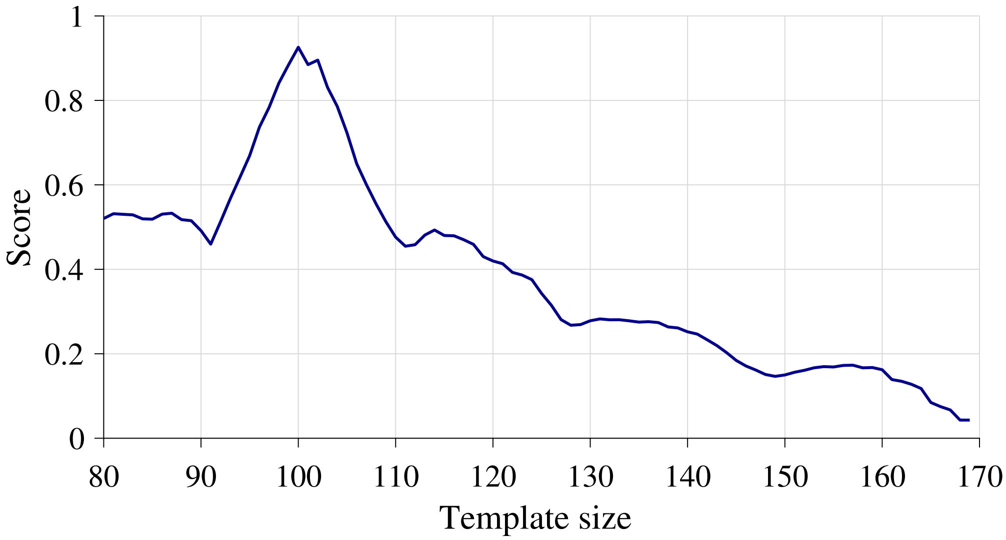


image files within that time period and generate a maximum match score for each pair of image and template.

The match scores usually have a bimodal distribution with two peaks. In these cases, the match score value at the trough is chosen as the new threshold.

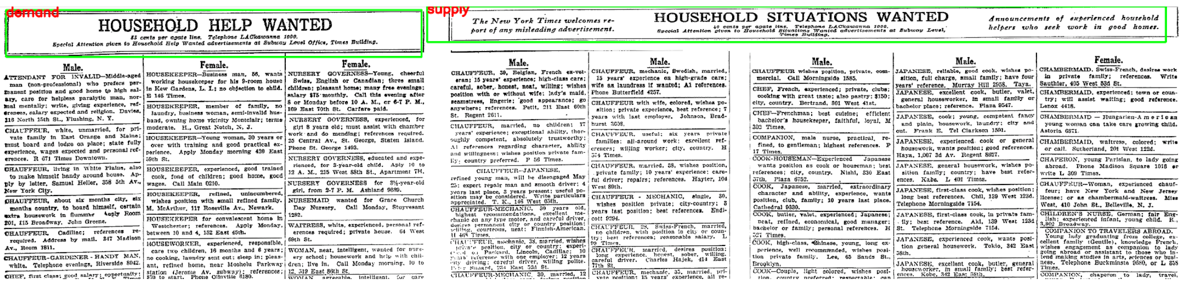
3. We repeat the previous step by using the updated thresholds. Finally, we also fine-tune the thresholds by comparing template matching results with manual counts.

D.3 Multi-Column Header Identification

The *New York Times* employs a more complex layout with categorical headers spanning multiple columns, see Figure D4. These multi-column headers varied in size and kerning, 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 different classes: background, help-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.

Figure D4: Template Matching with Multi-column Headers: example from the *New York Times*.



D.4 Scope of image processing work

Table D2 provides a summary on the scope of image processing work for our five newspapers. In total, we have processed close to 620,000 image files, 73,000 of which contain labor ads. The morphological operations are used to split more than 11 million ads, about 8 millions of which are labor ads.

D.5 Additional Comparisons with Manual Counts

The print quality varies not only across newspapers, as shown in Table 1, but also across sub-periods for the same newspapers. The quality is particularly poor after 1929 for the *Atlanta Constitution* and after 1926 for the *Washington Post*. Table D3 shows the error measures for periods with better print quality, which are notably better than the baseline case in Table 1.

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 for every month for 1) the *New York Times* between 1900 and 1938, 2) for the *Los Angeles Times* and the *Chicago Tribune* between 1925 and 1938, and 3) for the *Atlanta Constitution* between 1930 and 1935.

D.6 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.[4] The implementation of the procedure involves the following steps:

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

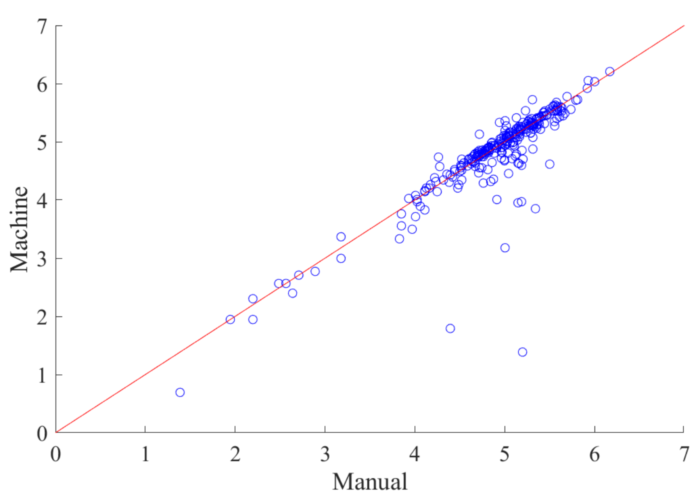
1. We collect more than 9,000 images between 1925 and 1938 and then manually classify those images to the following categories: help-wanted headers (female), help-wanted headers (male), other headers, as well as non-header images. Table D4 provides a summary of the image files. We split the period into two sub-periods, 1925-30 and 1931-38, since the wordings for help-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.

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-wanted ads based on the classification of each image file.

Figure D5 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 is comparable to the template matching method in terms of accuracy. Since a large amount of manual classification is needed to achieve high accuracy, the machine learning method is more costly and, therefore, we use the template matching method in the baseline case.

Figure D5: Comparison between machine learning and manual counts in the *Los Angeles Times*.



by Facebook.

D.7 Optical Character Recognition

Applying OCR on each labor ad becomes prohibitively expensive, as state-of-the-art OCR tools charge by the number of images, instead of image sizes. As explained in Section 2.4, we first store the bounding box of each ad when the image is segmented into columns and 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 by at least 50% is classified as text of that ad.

Multi-column headers are treated similarly. Any word that overlaps with the bounding box of a multi-column header by at least 50% is included in the text for that ad. To simplify further analysis, the multi-column header is recorded as a header in each column of the OCR output.

After extracting the text from the images with OCR, textual analysis can be used to extract gender information.

1. First, spelling check is applied to the text to remove common spelling errors caused by the OCR process.

2. Then the text is cleaned. Words and characters that add no value are removed.

3. Regular expressions are used to search for gender keywords, such as female, male, men, women, in section headers that are identified after the Template matching procedure.

4. The ads between headers are then categorized according to the header of the section and counted.

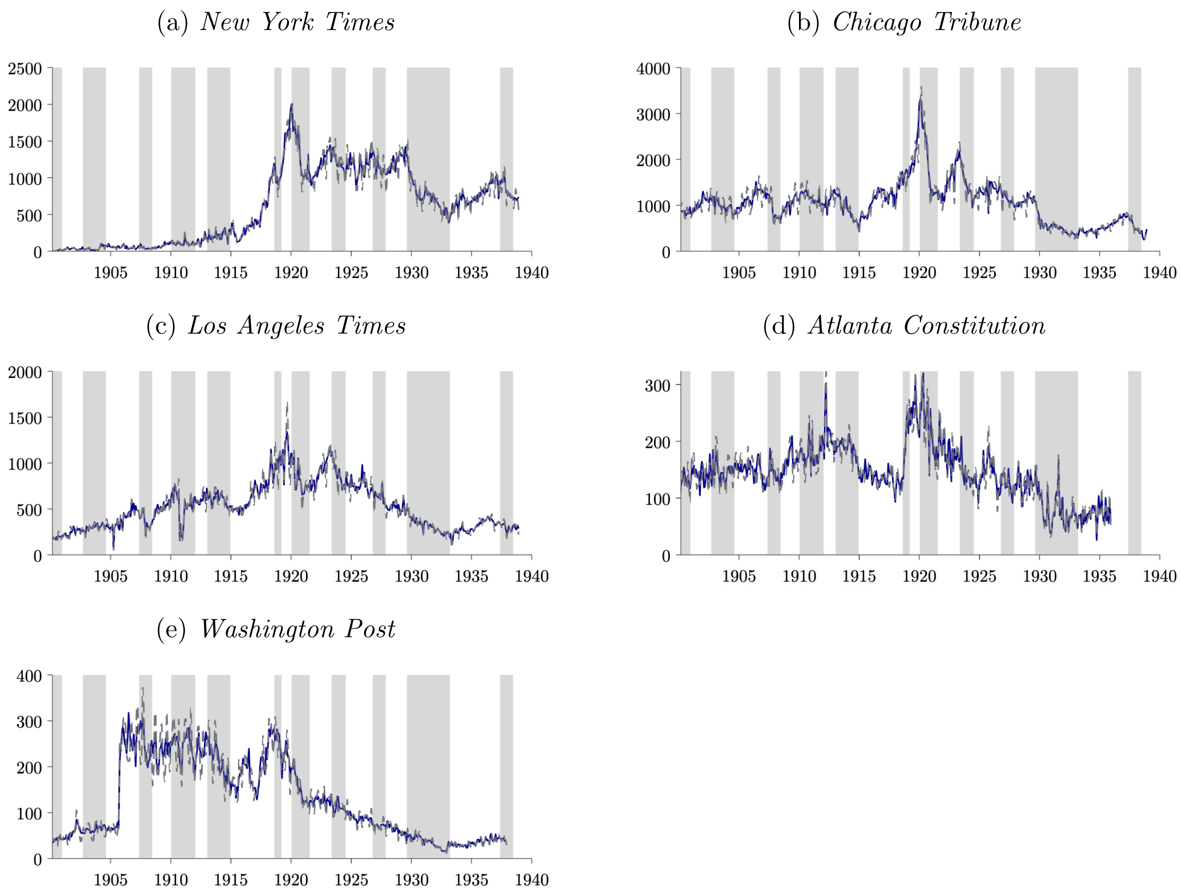
E Seasonal adjustment

Figure E1 plots a comparison of our monthly counts for help-wanted ads when the series are and are not seasonally adjusted across newspapers. In general, seasonal adjustment smooths each series and reduces the volatility.

F Additional Historical Data: Details

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

Figure E1: Comparison of seasonally adjusted and non-seasonally adjusted measures: Help Wanted

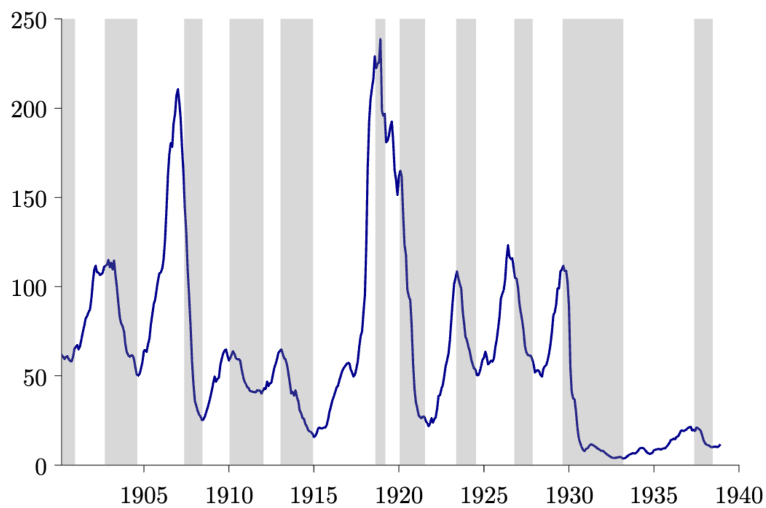


Note: The blue lines are seasonally adjusted, while the black dashed lines are not. NBER recession dates in grey.

F.1 Aggregate Labor Market Tightness

Next, we turn to labor market tightness by complementing our help-wanted index with estimates of the rate of unemployment in the first half of the 20th century.[5] The ratio of job postings to the unemployed, shown in Figure F1, 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 WWI. 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 recovered in the late 1930s.

Figure F1: Aggregate Labor Market Tightness



Note: Labor market tightness is defined as the ratio of help-wanted ads per labor force to the unemployment rate. The series is normalized to Jan. 1927=100. NBER recession dates in grey.

[5]Appendix F provides more details. Data limitations restrict the ability to use city-level measures of unemployment, see the discussion in section 3.1.

F.2 Gender Ratios: Help-wanted vs Employed

The help-wanted ads targeting women played an out-sized role during our data period. Table F1 displays the average ratio of female to male labor ads in each city in a given year, comparing to the gender ratio of gainful workers in a given state in the same year as provided by the U.S. Census. On average, the ratio of women to men employed is around 0.25, while the ratio in help-wanted ads is closer to 0.75.

F.3 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 help wanted registrations from local businesses. These reports, which we digitized, provide an alternative measure of city-level help-wanted during this period.

Figure F2 compares the two series at the city level and provides a similar message. The two series largely follow similar trends and business cycles for most cities, as both series declined during recessions. It suggests that our measures using classified ads capture similar information about local labor demand.

F.4 Bradstreet Index

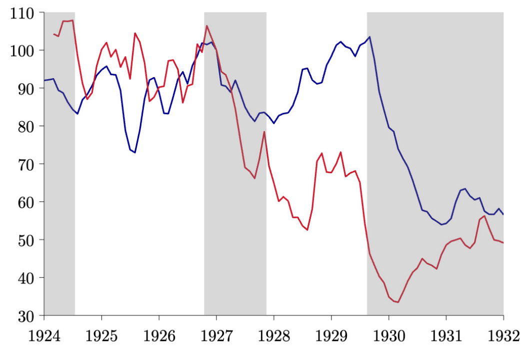
Bradstreet's Trade at a Glance summarized the status of various sectors for the main cities in the United States. Namely, the sectors referenced are wholesale trade, retail trade, manufacturing and industry, collections, and crops.

To create an index, we follow the methodology of Correia *et al.* (2022). We use their data for the years 1917-1922 and extend their approach to cover the years 1923-1932. Correia *et al.* (2022) develop a three-valued index that describes the current condition of each sector as "bad," "fair," or "good." For the classification, they allocate descriptive words from the Trade at a Glance into the three categories and denote each market in each city into a category based on what word is used to describe the market. We use their classification system and extend it to include additional descriptive words used in the additional years that we categorize. Specifically, the categories are denoted as follows:

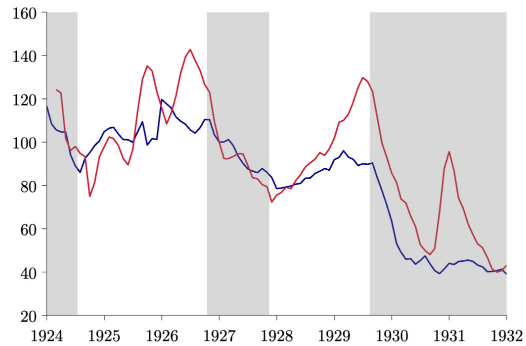
1. Good: good, brisk, excellent, active, liberal, very active, better, record, very good, steady, more active, prompt, stimulated.
2. Fair: fair, moderate, fair to good, satisfactory, close, 3/4 capacity, 60 percent, 75 percent, 75% basis, normal, fair activity, fairly active, hesitating, hesitation, only fair,

Figure F2: City-Level Help-Wanted: Newspapers vs. Public Employment Agency Data

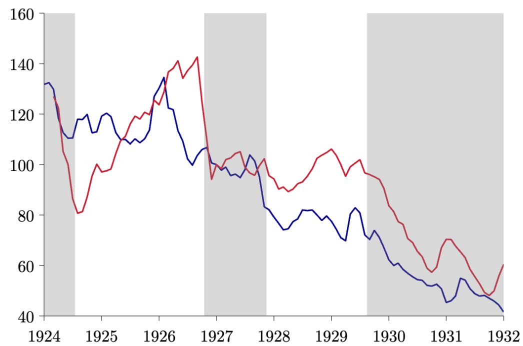
(a) *New York*



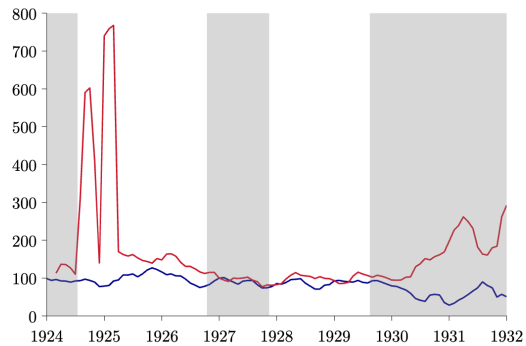
(b) *Chicago*



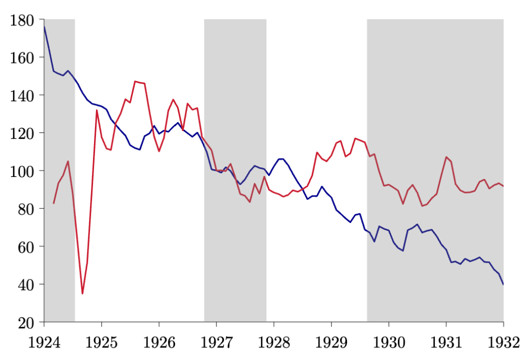
(c) *Los Angeles*



(d) *Atlanta*



(e) *Washington DC*



Note: Red lines represent public employment agency series, and blue lines are the constructed series using newspaper ads. Both series are normalized to 100 in January 1927 for comparability. NBER recession dates in grey.

slowdown, readjusting, half speed, half time, hampered, waiting, slack, uncertain, suspended, many strikes, contracted, disturbed, inactive, short time, retarded, paralyzed, irregular, unsettled, conservative.

3. Bad: quiet, dull, slow, very slow, cautious, interrupted, light, restricted, below normal,

curtailed, under normal, poor, lagging, tardy, delayed, backward, drag, spotty.

In addition, sometimes sectors are described in relative terms to the previous report, e.g., "gaining" or "slowing down." In these cases, we use the classification of Correia *et al.* (2022) and extend it to include additional descriptive words.

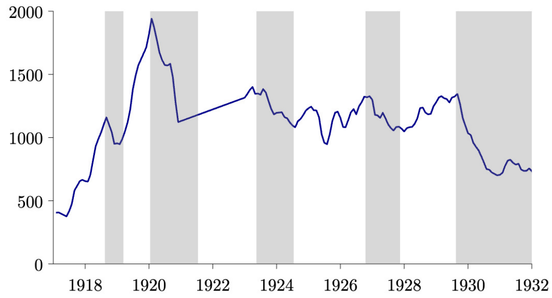
To create a numerical index, we assign a value of 0 if the rating is bad, 50 if the rating is fair, and 100 if the rating is good. We average the data in each sector at the city-month level, our frequency of interest. To create a combined index of wholesale trade, retail trade, and manufacturing, we take a simple average of the monthly index for the three sectors.

F.5 Aggregate Time Series

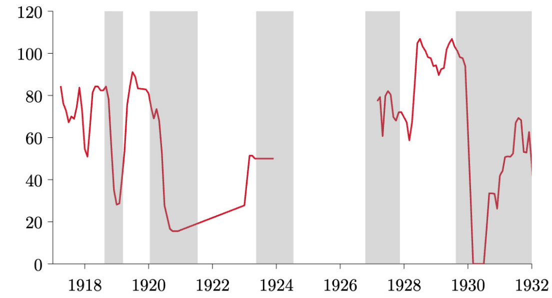
- City population estimates: U.S. Census (1900, 1910, 1920, 1930, 1940).
- Compensation of Employees: U.S. Bureau of Economic Analysis, Compensation of Employees: Wages and Salary Accruals [WASCUR], retrieved from FRED, Federal Reserve Bank of St. Louis, October 3, 2024.
  - Data Variable: Compensation of Employees: Wages and Salary Accruals [WASCUR]
  - Units: Billions of Dollars, Quarterly, Seasonally Adjusted Annual Rate
  - Time Period: 1955 to 2024
- Consumer Price Index: 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, October 3, 2024.
  - Data Variable: Consumer Price Index for All Urban Consumers: All Items in U.S. City Average [CPIAUCNS]
  - Units: Index (1982-1984 = 100), Monthly, Not Seasonally Adjusted
  - Time Period: 1955 to 2024
- Total nonfarm payroll: U.S. Bureau of Labor Statistics, All Employees, Total Nonfarm [PAYEMS], retrieved from FRED, Federal Reserve Bank of St. Louis, October 3, 2024.
  - Data Variable: All Employees, Total Nonfarm [PAYEMS]
  - Units: Thousands of Persons, Monthly, Seasonally Adjusted
  - Time Period: 1955 to 2024

Figure F3: Help-Wanted ads vs. City-Level Bradstreet Indicators, Jan. 1900 to Dec. 1938

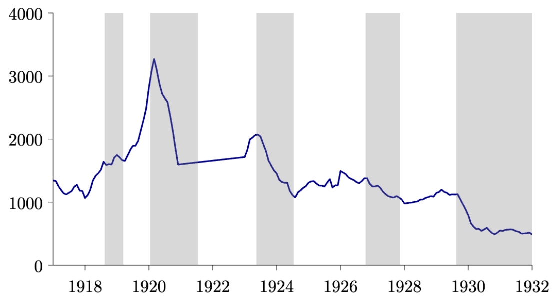
(a) Help Wanted: *New York Times*



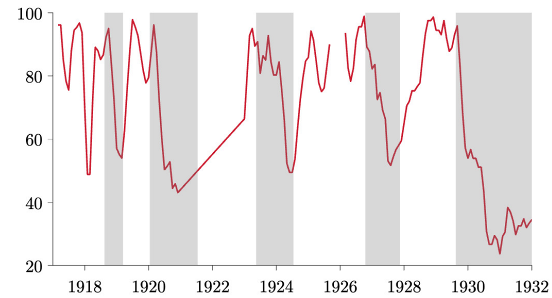
(b) Bradstreet Index: New York City



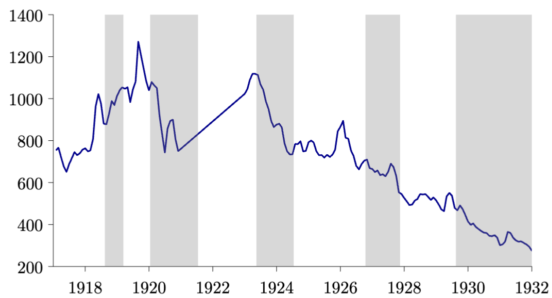
(c) Help Wanted: *Chicago Tribune*



(d) Bradstreet Index: Chicago

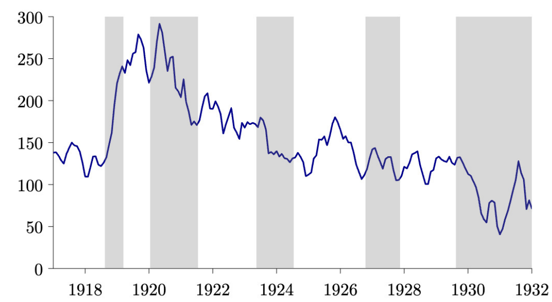


(e) Help Wanted: *Los Angeles Times*

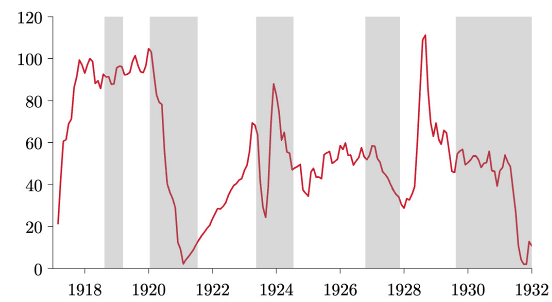


(f) Bradstreet Index: Los Angeles

(g) Help Wanted: *Atlanta Constitution*



(h) Bradstreet Index: Atlanta



- The unemployment rate and labor force are obtained from Petrosky-Nadeau and Zhang (2021).
- Labor productivity: We construct the measure of labor productivity from January 1884 to December 1940 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: Kendrick (1961), and Kendrick (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: Miron and Romer (1990) with corrections to the original supplied by the authors). 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 temporally disaggregating the annual productivity series with a Denton proportional first difference procedure and the monthly industrial production series as indicators.
- Vacancy Rate: Job Vacancy Rate (%): SF Fed Data Explorer, Federal Reserve Bank of San Francisco. Retrieved September 15, 2024; based on Petrosky-Nadeau and Zhang (2021). https://www.frbsf.org/research-and-insights/data-and-indicators/sffed-data-explorer/ located with chart code 41s022815018000022000000000000.
  - Data Variable: Job Vacancy Rate
  - Units: Percent, Monthly
  - Time Period: 1919 to 2024
- City-Level Newspaper Circulations: Gentzkow, Matthew, Jesse M. Shapiro, and Michael Sinkinson. 2014. United States Newspaper Panel, 1869-2004. Inter-university Consortium for Political and Social Research, December 10. https://doi.org/10.3886/ICPSR30261.v6.
  - Data Variable: Circulation
  - Units: Circulation (Number of Copies)
  - Time Period: 1900 to 1940

References

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

Correia, S., Luck, S. 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.

Kendrick, J. W. (1961). *Productivity Trends in the United States*. Princeton University Press.

— (1973). *Postwar Productivity Trends in the United States*. National Bureau of Economic Research.

Miron, J. A. and Romer, C. D. (1990). A new monthly index of industrial production, 1884-1940. *Journal of Economic History*, **50**, 321–35.

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.

Table B1: Frequency of Occupations on May 11, 1930 (Situation Wanted Column Removed)

Los Angeles Times, total ads = 907		
Occupation category	Help wanted	California US Census
Agriculture, fishing and forestry	<1	14
Extraction of minerals	0	2
Manufacturing and mechanical industries	5	25
Transportation and communication	<1	8
Trade	34	17
Public Service	<1	2
Professional Service	8	9
Domestic and personal service	22	11
Clerical	15	10
No category	14	-
Chicago Tribune, total ads = 837		
Occupation category	Help wanted	Illinois US Census
Agriculture, fishing and forestry	<1	11
Extraction of minerals	0	2
Manufacturing and mechanical industries	6	32
Transportation and communication	<1	9
Trade	59	15
Public Service	0	2
Professional Service	6	7
Domestic and personal service	9	10
Clerical	11	12
No category	10	-
New York Times, total ads = 2347		
Occupation category	Help wanted	New York US Census
Agriculture, fishing and forestry	<1	5
Extraction of minerals	0	<1
Manufacturing and mechanical industries	3	34
Transportation and communication	<1	9
Trade	56	16
Public Service	<1	2
Professional Service	13	8
Domestic and personal service	14	13
Clerical	13	14
No category	<1	-

Table D1: 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		

Table D2: Scope of Work

	Images	Labor Images	Ads	Labor Ads
New York Times	204,667	24,991	7,281,776	1,217,013
Chicago Tribune	42,865	8,656	11,812,803	2,128,634
Los Angeles Times	211,206	16,775	6,980,407	1,052,724
Atlanta Constitution	20,247	3,039	1,292,977	268,919
Washington Post	140,445	2,816	2,505,114	270,146

Table D3: Accuracy for Sub-periods: Comparison with Manual Counts

	1st Tertile	Median	3rd Tertile
Atlanta Constitution (1900-1929)	0.03	0.04	0.07
Washington Post (1900-1925)	0.01	0.02	0.04

Notes. Errors between machine and manual counts computed according to equation (1).

Table D4: Summary statistics of data input for the machine learning model.

Period	Category	Number of images
1925–1930	Total	4446
	Help wanted	1103
	Other headers	2153
	Non-header	1182
1931–1938	Total	4667
	Help wanted	730
	Other headers	2545
	Non-header	1202

Table F1: Female-male Ratio: Help-wanted Ads vs. Employed

		1900	1910	1920	1930
Aggregate	HW Ads	0.76	0.68	0.82	0.63
	Employed	0.22	0.27	0.26	0.28
New York	HW Ads	0.63	0.52	0.83	0.69
	Employed	0.29	0.33	0.34	0.34
Chicago	HW Ads	0.76	0.65	0.80	0.42
	Employed	0.20	0.23	0.26	0.29
Los Angeles	HW Ads	1.20	1.03	0.78	0.83
	Employed	0.16	0.19	0.23	0.29
Atlanta	HW Ads	0.37	0.29	0.57	0.57
	Employed	0.34	0.44	0.34	0.37
Washington D.C.	HW Ads	0.78	1.52	0.93	0.63
	Employed	0.48	0.50	0.65	0.57

Notes: Labor force data from the U.S. Census *Historical Statistics of the United States, Colonial Times to 1970*, Table D26-28. Census data measures "gainful workers" at state level, except for the District of Columbia. Aggregate refers to the population-weighted average of the five cities for the newspaper indices and all gainful workers in the total U.S. for the employed.

Table F2: City Population Estimates by U.S. Census

Year	New York	Chicago	Los Angeles	Atlanta	Washington D.C.
1900	3.4	1.7	0.1	0.1	0.3
1910	4.8	2.2	0.3	0.2	0.3
1920	5.6	2.7	0.6	0.2	0.4
1930	6.9	3.4	1.2	0.3	0.5
1940	7.5	3.4	1.5	0.3	0.7

Note: Decennial Census city populations, in millions