Sexism, Social Outcomes, and the Gender Wage Gap

Owen, Ann L. and Wei, Andrew

Hamilton College, Federal Reserve Board

11 August 2020

Sexism, Social Outcomes, and the Gender Wage Gap

Ann L. Owen*

Hamilton College

Andrew Wei

Federal Reserve Board**

August 2020

Abstract

Using Google Trends data to identify hostile sexism, we find that sexism explains about 8 cents (or 41 percent) of the residual gender wage gap, the wage gap after controlling for education, occupation, industry, and age. We find evidence for a direct effect of sexism consistent with labor market discrimination and an indirect effect that works through social outcomes that reduce hours worked which itself directly affects wages. Consistent with theories of discrimination, the direct impact of sexism is greater for women who are less educated, work in less competitive industries, and work in industries with fewer female workers.

JEL Codes: J10, J31, J71

1 Introduction

We exploit geographic variation in both hostile sexism and the gender wage gap and find that sexism has both a direct and indirect effect on the gender wage gap. Hostile sexism invokes explicitly negative beliefs about women; overall, we find that about 41 percent of the gender wage gap between white men and women that exists after controlling for education, occupation, industry, and age can be explained by it. The direct effect of sexism accounts for about 29 percent of the wage gap and is consistent with labor market discrimination: the wage gap between white men and white women with the same education, age, industry, and occupation is larger in geographic areas that exhibit more sexism. We also find corroborating evidence that sexism is related to discrimination by finding larger impacts in circumstances in which discrimination would be less costly.

However, there is also a nontrivial indirect impact because sexism can influence the non-labor market choices of women, such as marriage and having children. These social outcomes are significant determinants of labor market outcomes as well, especially the number of hours worked outside the home. By instrumenting for both sexism and social outcomes in a simultaneous equation model using GMM, we show a causal role for both the direct and indirect effects of sexism on wages. The evidence from these simultaneous equation models suggests multiple ways in which sexism affects the gender wage gap: It affects it directly as well as indirectly via social outcomes which then influence hours worked and hourly wages. Importantly, our results hold when controlling for a large number of characteristics and a broad set of fixed effects. Previous studies do find some evidence suggesting the relevance of sexism in determining the gender wage gap, but the evidence thus far has been correlational and context-specific (Charles et al., 2009, 2018; Janssen et al., 2016).

Our work uses a novel measure of sexism constructed from Google Trends data. Google data provides several advantages over previous measures, constructed exclusively with the General Social Survey (GSS) (Charles et al., 2009, 2018). First, it identifies sexism at the media market level, providing a larger sample size compared to studies that examine labor market outcomes of women at the state level.[1] Second, it is based on searches that occur mostly in private. In contrast, the GSS collects data with face-to-face interviews, a method that may understate an individual's sexism due to social censoring (Bound et al., 2001; Wright, 1993; Tourangeau and Yan, 2007; Berinsky, 1999, 2002).

The direct impact of sexism on the wage gap is suggestive of discrimination and we present several findings consistent with this interpretation. The seminal work of Becker (1971) predicts that sexism may decrease a woman's relative wage by causing disutility among employers, co-workers, or customers who interact with women. However, discrimination could be less prevalent in competitive markets because employers who choose to pay to discriminate may be driven out. For example, Janssen et al. (2016) use data from Swiss referenda to assess discriminatory attitudes towards women and correlate them with the gender wage gap, finding that it has a larger impact in the high-concentration manufacturing industry. Similarly, Hirsch et al. (2010) use data at the worker-firm level and find that the gender wage gap is larger in German firms with greater monopsony power. Blau and Khan (2017) review several papers that also find evidence suggestive of wage discrimination against women that point to various market imperfections that allow it to occur.

---

[1] A media market is a collection of counties identified by Nielsen Media Research that receive the same television broadcasting. It may be similar, but not identical, to a commuting zone or MSA. There are currently 384 MSAs in the U.S. and 210 media markets. We use this as a unit of analysis because Google search data is aggregated in this way.

Consistent with the idea that the extent of discrimination that occurs is influenced by its economic costs and benefits, we find that sexism has a greater impact in less competitive industries, in less competitive labor markets, in industries in which workers are more likely to directly interact with customers, and in industries with a smaller share of female workers. Finally, we show that sexism negatively affects the wages of women in a geographic area and somewhat weaker evidence suggesting it increases the wages of men.

We argue that the indirect impact of sexism works through its promotion of social outcomes that negatively influence hours worked outside the home and female wages. Akerlof and Kranton (2000) predict that utility from social outcomes depends on the extent an outcome is consistent with identity. Norms in more sexist areas may reinforce a more traditional female identity and women in these areas may feel compelled to pursue social outcomes traditionally associated with being female. Consistent with this idea, we use an instrumental variables strategy to find that in more sexist areas, women are more likely to marry, have children, and spend more time at home.

In finding causal evidence for the role of sexism in determining social outcomes that influence the wage gap, we link two separate strands in the literature. One strand finds evidence that marital status, children, and home hours contribute to a wider wage gap. For example, Hersch and Stratton (2002) find that married women who spend more time doing housework have lower wages. Similarly, Lundberg and Rose (2000) show that married women who have their labor supply interrupted by childbirth earn lower wages.[2] A second strand of related literature finds evidence that these outcomes are influenced by norms and attitudes. For

---

[2] See also Hersch and Stratton (1997), Korenman and Neumark (1992), and Becker (1985).

4

example, Bertrand et al. (2015) establish that gender identity within a household influences marriage formation, the wife's labor force participation, and the division of labor within the household. In a similar finding, Fernandez and Fogli (2009) find that for second-generation American women the culture of their ancestors influences their fertility and labor force participation. Other authors who have studied a link between culture and social outcomes that could influence labor force behavior include Alesina et al., 2013; Carlana, 2019; and Bursztyn et al., 2017. Our work merges these two approaches by finding evidence for the indirect impact of sexism on the wage gap via social outcomes.

Because we exploit geographic variation in the wage gap, social outcomes, and sexism within the U.S. to obtain our results, our work is most closely related to the strand of the literature that examines how social and economic outcomes of women vary geographically within the U.S. Many have studied the evolution of social and economic outcomes of women through time (e.g., Bailey et al., 2014; Hsieh et al., 2019; Black and Spitz-Oener, 2010). In contrast, relatively few examine how social and economic outcomes of women vary geographically within the U.S. Beaudry and Lewis (2014) study the change in the gender wage gap over the 1980 – 2000 time period using cross-city differences, attributing a decline in the wage gap over this time period to the adoption of computers. Black et al. (2014) explain the cross-city difference in the labor supply of married women with variation in commuting times.

The work most closely related to ours is Charles et al. (2018) who study sexism, the gender wage gap, and social outcomes of women across states, measuring attitudes about women with survey responses from the General Social Survey (GSS). However, we approach the same issue with a different method for measuring sexism that relies on Google search data. This allows us to study sexism and the wage gap in a smaller geographic area than a state. This is

appealing intuitively because a media market more closely resembles a local labor market than a state. A media market is a collection of counties identified by Nielsen Media Research that receives the same television broadcasting. Especially in large states, there can be significant variation in wage gaps and culture across media markets in the same state. For example, in Odessa, TX, we estimate that the average woman with the same education, occupation, and industry earns 76 cents for every dollar the average man earns. In Austin, the average woman earns 82 cents for every dollar earned by a comparable man. By measuring sexism in a different way in smaller geographic areas, we are able to obtain more robust results and evidence of causality.

We proceed by first explaining the construction of the measures of sexism and the wage gap in Section 2. Section 3 explains our methods and Section 4 presents results for the direct effect of sexism on the wage gap. We then move on to exploring the indirect effects of sexism on the wage gap via social outcomes in Section 5 and discuss the determination of male and female wages as well as alternative explanations in Section 6 before offering a conclusion in Section 7.

2 Measuring Sexism and the Wage Gap

Our general approach is to explore the role that sexism plays in explaining the variation in the gender wage gap across different geographic areas in the U.S. To do so, we develop a measure of sexism and the wage gap at the same geographic level. In this section, we explain the procedures for developing these measures.

2.1 Measuring Sexism

Note: The following sections describe offensive language. We refer to these words using coded language, shown in Appendix A.

Constructing a Sexism Dictionary

We focus exclusively on hostile sexism. Hostile sexism, identified by Glick and Fiske (1996, 1997), involves explicitly negative attitudes about women. It reflects beliefs that women are inferior to men in both competence and character. In contrast, benevolent sexism is more paternalistic and can even idealize women.[3] To capture how hostile sexism is expressed in language, we rely on Hall and Bucholtz (2012). They identify three themes that are reflected in most expressions of sexism: 1) women as sexual objects, 2) women as dumb and emotional, and 3) women as rude and evil. Throughout this paper, unless otherwise noted, we use the terms "sexism" and "hostile sexism" interchangeably. Although our focus on hostility towards women may cause us to underestimate the sexism of an area, an advantage of this approach is that there is less ambiguity in identifying words or phrases that clearly objectify or demean women, and we are less likely to identify trends that are related to other cultural attitudes or beliefs.[4]

Because sexism is multidimensional, it is difficult to capture it with only one word or phrase.[5] As a result, we proceed by constructing a sexism dictionary; we use Hall and Bucholtz's (2012) three themes to serve as the basis for this exercise. Their list of 112 sexist

---

[3] Charles et al. (2009, 2018) construct a measure of sexism that focuses on beliefs about the role of women in society. In contrast, our measure of sexism only captures negative and demeaning attitudes.

[4] For example, religion may be intertwined with benevolent sexism in a complicated way.

5 Stephens-Davidowitz (2014) is able to effectively capture racism with searches for one word and variants of it.

phrases provide a starting point for our dictionary.[6] Because their list includes regional slang and phrases with multiple meanings, we employ a filtering procedure. First, we select any clearly-sexist phrases, adding them to our dictionary (words 1-4, Table 1). Then, we employ a rudimentary search tree algorithm from Owen and Wei (2020). The algorithm enters each of Hall and Bucholtz's (2012) 112 phrases into thesaurus.com, adding any sexist phrases that appear in the results. By repeating this process for new phrases we find until no more sexist words come up in the results, we build a tree of related words for each entry in the original list of 112 phrases. We conclude our search with an expanded list of 266 sexist phrases.

The next step of the procedure extracts how people express sexism on Google. We enter the expanded list of 266 words into Google Correlate, identifying the 100 most correlated queries for each word at the state level. This step gives a list of 15,460 potentially sexist queries made by Google users. To extract queries referring specifically to females, we filter this list by the keywords "girl(s)" and "woman(en)". We identify any negative adjectives that are used to refer to females, finding six: "fat", "ugly", "stupid", "dumb", "emotional", and "cheating". We construct the final dictionary by generating all possible permutations between these six adjectives and the four words: "girl", "girls", "woman", and "women." Table 1 shows the final dictionary. It includes 16 phrases, representing all three themes from Hall and Bucholtz (2012).

The sexism dictionary includes phrases that users search frequently. The most frequently searched word, "[word 2](es)", has a search volume of over 130,000 per month. This, for context, is about 7 times more than searches for "[word 5](s)", 2.8 times more than searches for

---

6 Hall and Bucholtz (2012) construct a list of 112 sexist phrases by surveying a diverse sample of 365 Berkeley undergraduate students over the 1991-1992 period, asking each student to provide 10 expressions of sexism.

"Economics", and about equal to searches for "Sweater."[7] Frequency analysis on top searches indicates that phrases in our dictionary are often used to find sexist material.[8] The most popular topics are "lyrics", followed by "meme", "song", and "pictures." A cursory search on Google shows that these topics, used in conjunction with our dictionary, returns sexist content. For example, the second result of "[word 3] lyrics" is about the desire to sexually exploit an attractive woman. The first result of "[word 2](es) meme" is a website of GIFs that objectify women. Women as objects is one of the three themes of hostile sexism identified by Hall and Bucholz (2012).

The Sexism Index

We use Google Trends to obtain the relative search volume of phrases in our sexism dictionary over the 2004-2019 time period, apply principal components analysis to these search volumes, and generate *Sexism Index,* a variable that is normalized so that it has a mean of zero and standard deviation of one. The final variable measures the number of standard deviations an area's sexist search volume is form the mean. Additional technical details are in Appendix A.

Sexism Index assigns weight in an intuitive manner. Ranking phrases in the sexism dictionary by their correlation to *Sexism Index* reveals which phrases receive the heaviest weight. Four of the five most correlated phrases— "ugly woman(en)", "[word 4](s)", "cheating girl(s)", "[word 3](s)", and "fat girl(s)"—all portray women as objects, suggesting that the first theme in Hall and Bucholtz (2012) receives the heaviest weight. An important point is that *Sexism Index*

---

[7] A potential concern is that a few of our words have low search volumes below 1,000 per month. We show that our results are robust to dropping these words from the dictionary.
[8] To perform frequency analysis, we scrape the top searches of each word in our sexism dictionary, obtaining a list of 330 queries. From the list, we manually remove spurious entries and words in our sexism dictionary. We rank the remaining words by the frequency a word appears across all queries.

varies widely by geography: The least sexist media markets are San Francisco and Portland; the most are Myrtle Beach and Columbus. For example, the level of sexism we identify in Myrtle Beach is 3.04 standard deviations above the mean, while the level of sexism in San Francisco is 4.41 standard deviations below the mean.

To further explore the interpretation of the index, we correlate its average over 2004-2015 with responses to two questions asked in the 2016 American National Election Studies (ANES) survey. The first question is, "Do women often consider innocent remarks sexist?" and the second question is, "Do women often fail to appreciate what men do for them?" Affirmative responses to both of these questions indicate a negative evaluation of women. We use this survey data to construct a measure of sexist attitudes at the state level and recalculate the *Sexism Index* at the state level for comparison. We find a positive correlation either using the average response to the questions or by using a ranking of the states based on the average score (See Figure 1).[9] Although low numbers of respondents in some media markets prohibits us from using the ANES data in our analysis, the positive correlation between these survey responses and *Sexism Index* provides some validation that the index we created is capturing sexist attitudes.

2.2 Constructing Wage Gaps

Our main dependent variable is the residual gender wage gap of a media market. The raw gender wage gap is defined as the average log wage of men minus the average log wage of women. We calculate a residual wage gap by estimating this raw wage gap while controlling for age, occupation, industry, and education as described in this section.

---

[9] Specifically, we code the survey responses so that higher values indicate stronger sexist attitudes and then normalize the results by subtracting the mean and dividing by the standard deviation.

To calculate the wage gap in a geographic area, we start with individual level data from the American Community Survey (ACS). We impose several sample restrictions on the ACS data to avoid conflating issues with race, age, and employment: we restrict the sample to whites between the ages of 25 and 64 who are wage earners, but not self-employed. We do not restrict our analysis to only full-time, year-round workers to allow us to consider the ways in which sexism and social outcomes might affect this aspect of labor force attachment. However, 95% of the individuals in our data work more than 46 weeks/year and 93% work 30 or more hours/week. Although we do not report these results in detail, we find that our conclusions are robust to focusing on full-time (> 30 hours/week), year-round workers.[10] We use the log of real hourly wages in our estimations which are defined as the log of annual wage and salary divided by the product between hours worked in a week and weeks worked in a year, deflated by the CPI.[11]

To construct the wage gap at the media market level, we estimate an OLS regression of the form

$$(1) \qquad \ln W_i = a_d^m + a_d^f + \gamma_t + \varphi\, V_i + \varepsilon_i,$$

where the dependent variable is the log hourly wage of individual i. a_d^m and a_d^f are a set of male and female indicators for each media market, γ_t is a set of year fixed effects, and V_i is a vector of individual characteristics that capture an individual's age, occupation, education, and industry.[12]

---

[10] Any results discussed in the text but not reported are available from the authors upon request.

[11] For some years, the ACS lacks a continuous variable for weeks worked in a year. We impute weeks worked using the midpoint of the categorical variable for these years. We only use the top two categories: "worked 50-52 weeks" and "worked 48-50 week". The other categories span a longer set of weeks, making imputation less accurate.

[12] Specifically, V_i includes *Age quartic, Schooling FE, Industry FE,* and *Occupation Type FE*. Appendix B defines these variables in detail. These variables are common to the labor economics literature. We specify them in the most restrictive manner. It is possible that sexism influences some of these characteristics, leading us to underestimate the impact of sexism.

We construct the residual wage gap of media market d by calculating $a_d^m - a_d^f$, the difference

between the coefficients on the male and female indicators of media market d. Our wage gap

compares the wages of men and women from the same media market. This makes it independent

from media market characteristics that have an equal impact on male and female wage.

In our main estimations, we pool annual data from the 2004 to 2019 ACS and use that in

cross-sectional estimations.[13] In some supplementary estimations, we also present results from

panel data and construct a second wage gap that is time-variant and compares the wages of men

and women who live in a media market in a specific year. It is constructed by entering the male

and female media market fixed effects separately by year such that the wage gap of media

market d and year t is defined as $a_{dt}^m - a_{dt}^f$. We use the panel data to explore if our conclusions

are subject to omitted variables, however, we do not use these data for our main results because

examining annual changes in sexism and the wage gap will not allow us to identify the longer

term effects of culture.

Our main estimations rely on data from the American Community Survey (ACS).   An

alternative source of wage data is the Current Population Survey (CPS). We choose to use the

ACS because it provides two main advantages over the CPS. It provides more observations and it

identifies an individual's birthplace at the state level, which we use as part of our identification

strategy described below. In results not reported, we obtain similar results with CPS data.

---

[13] The average individual in the data earns $28.86 per hour, is 44 years old, and has 15 years of schooling.

2.3 Preliminary Data Patterns

The wage gap varies considerably across geography (Figure 2). Despite the considerable geographic variation, all wage gaps are positive and statistically significant. This implies that in every media market, the average man continues to earn more than the average woman with the same education, age, occupation, and industry.

The magnitudes of the wage gaps are large. The mean wage ratio across all media markets during this 15-year time span is 80.09. This implies that in the average media market, women earn 80 cents for each dollar earned by men who are similar in education, age, occupation, and industry.

How do women fare in the media markets with the largest and smallest wage gaps? The media market with the largest wage gap is Lake Charles, Louisiana, where women earn 68 cents for each dollar earned by men. The media market with the smallest wage gap is Gainesville, Florida, where women earn 86 cents for each dollar earned by men. Appendix C provides more detail on the media markets with the largest and smallest wage gaps.

3 Methods

The first step of our estimation process is to obtain a residual wage gap and a measure of sexism by media market as described above. In the second step, described below, we use that residual wage gap as the dependent variable and explore the ways in which sexism is related to it.

We use four different estimation techniques in the second step: cross-sectional estimation, instrumental variables, simultaneous equations models, and panel estimation. In this section, we explain the cross-sectional specification. As we proceed through the analysis, we then explain

how we modify our initial approach to present more convincing evidence of a causal role for sexism in determining the wage gap.

The first specification is a cross-sectional OLS regression of the form:

$$(2) \qquad \text{Wage Gap}_d = \alpha + \beta_1 \text{ Sexism Index}_d + \beta_2 \text{Social outcome}_d + \delta Z_d + \mu_d.$$

The dependent variable is the wage gap of media market d and the main explanatory variable is the Google-based $Sexism\ Index_d$. The variable $Social\ outcome_d$ represents any one of the following three social outcomes: *Women Married, Children*, and *Home Hours Gap. Women Married* is the proportion of white women between the ages of 25 and 64 that are currently married, *Children* is the proportion of white women between the ages of 25 and 40 with at least one child at home, and *Home Hours Gap* is the average weekly home hours (non-labor force hours) of white women between the ages of 25 and 64 minus the average weekly home hours of white men between the ages of 25 and 64.

The cross-sectional specification includes a vector of control variables, Z_d. This vector includes 5 socioeconomic characteristics specific to each media market: *%Employed, %Female Employed, LFP, Female LFP,* and *College Gap. %Employed* is the proportion of white workers employed in a media market, *%Female Employed* is the proportion of white women currently employed, *LFP* is the labor force participation rate of whites, *Female LFP* is the labor force participation rate of white women, and *College Gap* is the proportion of white men with a college education minus the proportion of white women in that media market with a college education. Arguably, some of our "control" variables might also be considered social outcomes, especially *Female LFP* and the *College Gap.* We do not focus on these variables in our analysis, however, because we study the residual wage gap computed using only men and

women who earn wages and controlling for education of the individual. This allows us to more directly compare to previous literature that attempts to close the wage gap with individual characteristics (Blau and Kahn, 2017). All of the control variables are calculated using individuals who are between the ages of 25 and 64.

The errors of the cross-sectional specification are heteroscedastic due to our method of estimating wage gaps. As a result, in this specification and all additional estimations, we weight the regression by the inverse of the standard error of the wage gap and we make its standard errors robust to arbitrary heteroscedasticity and within-state error correlation.

An advantage of the cross-sectional specification is that it allows us to study the impact of sexism over a long time period because we have estimated a residual wage gap (with year fixed effects) and the sexism of a media market over a 15-year period.[14] The longer time frame is better suited to capture the influence of social outcomes like *Women Married* and *Children* that would not respond to contemporaneous changes in sexism.

In sum, in all of our estimation strategies, we estimate a residual wage gap for each media market and then use that estimate as the dependent variable in a second step of our estimation process.[15] That second step may involve cross-sectional, IV, GMM, or panel data methods. Our approach is similar to that employed by Charles and Guryan (2008) and Beaudry and Lewis (2014). The approach allows us to focus on the appropriate unit of analysis, the wage gap for a geographic area. An alternative way of proceeding would be to do a one-step estimation process

---

[14] See also Charles and Guryan (2008) and Charles et al. (2018) for a similar long-term focus.
[15] The advantage to using the residual wage gap as the dependent variable is that it allows us to control for a large number of individual characteristics. If we used the raw wage gap instead, we would sacrifice many degrees of freedom when controlling for these characteristics in the main specification.

with the individual-level wage data and use the sexism of an individual's place of residence as an independent variable. However, if we estimated the main specification at the individual level, we would duplicate *Sexism Index* across individuals from the same area. We would underestimate the standard error of the impact of sexism, even if we clustered errors geographically. That said, we confirm that our conclusions are robust to this alternative estimation procedure and provide individual level results in Appendix D.

4 Direct Effect of Sexism on the Wage Gap

4.1 Cross-sectional Estimates

Table 3 reports the estimation results of Equation 2 using cross-sectional data. Across all five specifications, a media market's *Sexism Index* is a positive and significant predictor of its wage gap. At the bottom of Table 3, we provide an estimate of the cost of sexism in cents based on the results in each column. In other words, this estimate is how much smaller the residual hourly wage gap would be if every media market had the lowest level of sexism. For example, the results in Column 1 indicate that the residual wage gap would close by 3.15 cents if every media market had the lowest level of sexism. Controlling for female economic outcomes in the media market makes the impact of sexism slightly larger (Column 2), suggesting that the baseline specification underestimates the impact of sexism.

In Columns 3 through 5 of Table 3, we add social outcomes as explanatory variables. As expected, media markets with a larger percentage of women who are married (Column 3) and have children at home (Column 4) have larger gender wage gaps. In addition, the greater the difference in home hours between men and women, the larger the wage gap (Column 5). The

16

results in Column 3 indicate that a 10 percentage point increase in the percent of women who are married increases the wage gap by 3.8 percent. Column 4 indicates that a 10 percentage point increase in the percent of women with a child increases the wage gap by 1.4 percent. Finally, Column 5 indicates that a 10 hour increase a woman's relative home hours per week increases in the wage gap by 14.5 percent.

Examining the magnitude of the coefficients on sexism in Columns 3 through 5 reveals that controlling for these social outcomes makes the impact of sexism smaller. T-tests for the difference of coefficients indicate that the decrease is statistically significant. For example, the bottom of Column 3 indicates that controlling for *Women Married* makes the impact of sexism 47 percent smaller ($p < 0.01$), the bottom of Column 4 indicates that controlling for *Children* makes the impact of sexism 38 percent smaller ($p < 0.05$), and the bottom of Column 5 indicates that controlling for *Home Hours Gap* makes the impact of sexism 33 percent smaller ($p < 0.01$).

Because the estimations in Table 3 control for a number of individual and media market characteristics that might affect productivity of women relative to men, one interpretation of the consistently significant correlation between sexism and the gender wage gap is discrimination. Although these results do not provide direct evidence on this, in Table 4, we present results that are consistent with the effect of sexism working through discrimination by examining differential effects of sexism for workers under different circumstances. In these estimations, we estimate wage gaps separately for different groups of individuals in the same media market so each media market has more than one wage gap associated with it.

For example, in Column 1 of Table 4, we estimate a residual wage gap for two groups of people in the same media market: those with college educations and those without. In other words, even though the residual wage gap is calculated controlling for the years of education of

17

individuals, this specification allows us to determine if there is a systematic difference in the way that sexism affects the residual wage gap for two groups: those with college educations and those without. The negative coefficient on the interaction term between *Sexism Index* and *College Educated* implies that the impact of sexism on the wage gap of college educated workers is smaller. In fact, when we examine the net effect of sexism on the wage gap of college educated workers, we find that it is small and statistically insignificant. In other words, our evidence suggests that sexism affects the wage gap for workers without college educations, but we find no evidence that it affects workers with college educations. Given that workers with college educations are in a more competitive labor market, this would be consistent with discrimination by employers. It occurs in the markets in which employers pay a lower price to discriminate.

The estimations in Columns 2 and 3 also investigate differential circumstances for the role of sexism. In Column 2, we present results that suggest that sexism has a larger effect on the wage gap in manufacturing and commerce industries. This result is similar to that found in Janssen et al. (2016) who argue that discrimination is less costly in manufacturing because it is a less competitive industry and is more beneficial in commerce industries because they require more interactions between workers and customers who may be biased.[16] The fact that sexism has a larger influence in the less competitive manufacturing industry is suggestive of discrimination by employers, but the greater impact of sexism in the commerce industries suggest that discrimination by customers is also at play.

---

[16] Autor et al. (2020) calculate Herfindahl-Hirschman Indices for several different industries and find the manufacturing industry to be most concentrated.

An additional result consistent with discrimination by employers appears in Column 3 where we show that in industries that have more women, sexism has a smaller effect on the wage gap. This too is consistent with sexism generating discriminatory behavior when discrimination is less costly. Black (1995) predicts that increases in gender diversity decrease the part of the gender wage gap attributable to discrimination. A larger fraction of women in the labor force increases the cost of discrimination caused by sexism because it requires even higher wages paid to men, who are relatively scarce. This would drive discriminators out of the market. These results are also consistent with discrimination by employers.

Finally, in the last column in Table 4, we present evidence that explores an interaction between race and sexism. Although in our main results we focus on the gender wage gap among white workers, we are also able to calculate a residual gender wage gap among Black workers. When we combine that with the wage gaps in each media market for white workers and estimate the impact of sexism for each race, we find that sexism widens the gender wage gap even more for Black workers than white.[17] This is shown in the positive coefficient on the interaction of *Sexism Index* and *Black* in the last column of Table 4. This too would be consistent with discrimination by employers if Black workers were in less competitive labor markets, allowing discrimination against Black women to be less costly.

Taken together, the results in Table 4 suggest that sexism has a larger effect on the wage gap in circumstances in which discrimination is less costly or potentially more beneficial to the firm. While this is not direct proof that the sexism of a media market generates discrimination in

---

[17] We do not find evidence of a similar interaction for the broader category of non-white workers.

the labor market, these findings are consistent with discrimination by both employers and customers.

4.2 Instrumental Variables Estimation

While the estimates in Table 4 are suggestive of sexism causing discriminatory behavior and exacerbating wage gaps, they do not present clear evidence of a causal link. One issue of concern is that women who are more productive may move to areas with less sexism and higher wages for women. This would generate a bias in the coefficient of *Sexism Index* in our estimations.

To address this issue, we employ an instrumental variables strategy, using two instruments for the sexism of a media market: the sexism of the closest media market and the mean sexism of all remaining media markets, weighted by the inverse of the distance to the original media market. Both of these instruments should be related to the sexism in a media market, but should be independent of the wages of women in that media market. We present results of this estimation in the first two columns of Table 5. Both the Sargan-Hansen p-values and the F-statistics for the first stage suggest that these are valid instruments. (First stage results are available in Appendix E.) The coefficients of *Sexism Index* remain positive and significant in the second stage results presented in Columns 1 and 2 of Table 5, confirming the conclusions that we drew with the cross-sectional study. Furthermore, it is interesting to note that the coefficients increased in size in the IV estimations, reducing concerns about potential bias discussed above in the cross-sectional results.

Finally, although the estimations in the first two columns of Table 5 pass the usual tests for instrument validity, we still consider a potential threat to validity that may result from

spillover of economic conditions between media markets that might affect wages. We note that it is more likely for spillover effects to cause spatial correlation in the level of wages rather than spatial correlation in the wage gap. Nevertheless, we address this concern in two ways. First, in the estimations reported in Columns 3 and 4 of Table 5, we exclude from the instrument set the sexism of any media market within 100 miles in the first stage estimation. Although the resulting coefficients are similar to those in Columns 1 and 2, the first stage F statistics are lower, suggesting that this set of instruments is weak. As a result of that concern, we adopt a second strategy by returning to the baseline model in Columns 1 and 2 but explicitly controlling for the wage gap of nearby media markets in the second stage (Columns 5 and 6). These results are also consistent with our previous results, suggesting a stronger causal role for sexism in determining the gender wage gaps that exist at the media market level.

4.3 Panel Data Estimation

Another concern about cross-sectional estimates is that there may be omitted characteristics of the media markets, correlated with sexism, that are responsible for the results we obtain. To address that concern, we present our main results estimated with a fixed effects model using panel data in Table 6.

These estimations include both a year and a media market fixed effect. We use the same set of control variables as used in the cross-section estimation, but in the panel, all data is at the annual frequency. The interpretation of these results is slightly different because they represent the impact of annual changes in sexism on annual changes in the wage gap within a media market. Thus, while the cross-sectional results may capture longer term impacts, these results capture the way in which annual changes are correlated contemporaneously. As such, the coefficients in all six panel specifications are smaller than the cross-sectional results, but are still

positive and statistically significant. Overall, we conclude from this exercise that the relationship between sexism and the wage gap is robust to the inclusion of media market and year fixed effects, reducing the concern that it is being driven by omitted variables.

5 Social Outcomes and Indirect Effects of Sexism

While the results presented earlier show a significant direct effect of sexism on the wage gap, we also note that we found that several social outcomes (*Women Married*, *Children*, and *Home Hours Gap*) also are positively correlated with the wage gap as well. When we control for these social outcomes in our main estimations, the magnitude of the direct effect of sexism gets smaller, indicating that sexism is correlated with these social outcomes. This suggests that social outcomes may be an indirect channel for sexism to impact the wage gap. In this section, we explore that channel and ultimately allow for simultaneous estimation of the gender wage gap and multiple social outcomes.

5.1 IV Estimation

As we explore the relationship between the sexism of an area and the gendered social outcomes of these areas, we are aware that these relationships may also be endogenous and adopt an instrumental variables strategy for this estimation as well. Women with different social outcomes may sort into areas with higher or lower wage gaps or different levels of sexism. In our previous estimations, we used an instrument based on spatial correlation of sexism. While we also were able to obtain consistent statistically significant results with an instrument based on spatial correlation for social outcomes, the instrument turns out to be relatively weak, and we pursue an alternative strategy.

As the instrument for social outcomes, we use the average social outcomes of the birth state. The first stage results are in Appendix E and show that this is a relatively strong instrument for social outcomes at the birth state-media market level. The logic driving the birth state as an instrument is that it is independent from sorting because individuals do not choose their birthplace. A concern, however, is that outcomes in the birth state directly influence the wages of women still living in the state. We argue that the threat posed by this possibility is small, but we address remaining concerns by controlling for both the average wage gap and average sexism of the birth state.

We are able to include the averages of the birth state because we estimate the wage gap at the birth state-media market level. A final point is that this specification is no longer at the media market level. We construct a dataset where, instead of each observation representing a media market, each observation represents all individuals born in a state b and currently living in a media market d. We use the average social outcomes of the state b to instrument for the social outcomes of women born in b and currently living in d.

Results of the second stage estimation of the wage gaps and social outcomes via the IV strategy appear in Table 7. In all these estimations, we instrument for both sexism and the social outcome. The p-values on the Sargan-Hansen test and the F-statistics for the first stage estimations suggest that our IV strategy is valid. Focusing first on the first two columns of Table 7, we see that sexism has a positive impact on the social outcome of marriage (column 2) and that the proportion of women that are married has a positive impact on the wage gap. The estimations in these two columns show both a direct and indirect effect of sexism. Specifically, the calculations at the bottom of Table 7 reveal that the indirect effect of sexism through its

impact on the proportion of women married is 1.74 cents.[18] In other words, if the average media

market had the lowest level of sexism, fewer women would be married and the effect of that

lower marriage rate would close the gender wage gap by 1.74 cents. The largest magnitude of an

indirect effect of sexism is via its impact on the *Home Hours Gap* (columns 5 and 6 of Table 7).

If the average media market had the lowest level of sexism, the effect on the *Home Hours Gap*

would close the gender wage gap by 3.21 cents. We do not find similarly strong effects for

Children (columns 3 and 4 of Table 7). When the instrumental variables strategy is used,

Children is no longer significantly related to the wage gap and we do not pursue further analysis

with it.

5.2 Simultaneous Equation Models

The results in Table 7 are from relationships estimated separately. But, the underlying

mechanisms are occurring simultaneously and it is most appropriate to estimate relationships

between sexism, social outcomes, and wage gaps simultaneously. In other words, we estimate a

model in which sexism has a direct effect on the wage gap and an indirect effect via its effect on

the social outcomes, which themselves affect the wage gap. These relationships are described

below.

$$\text{Wage Gap}_{bd} = \beta_0 + \beta_1 \text{Sexism Index}_d + \beta_2 \text{Women Married}_{bd} + \beta_3 \text{Home Hours Gap}_{bd} + \beta_4 Z_{bd} + \mu_{bd}$$

(3) $$\text{Women Married}_{bd} = \gamma_0 + \gamma_1 \text{Sexism Index}_d + \gamma_2 \text{Wage Gap}_{bd} + \gamma 3 Z_{bd} + \varepsilon_{bd}$$

$$\text{Home Hours Gap}_{bd} = \theta_0 + \theta_1 \text{Sexism Index}_d + \theta_2 \text{Wage Gap}_{bd} + \theta_3 \text{Women Married}_{bd} + \theta_4 Z_{bd} + \zeta_{bd}$$

---

[18] Using the coefficients reported in columns 1 and 2, the relevant calculation is
$e^{-.218+4.408(.187)(.026)} - e^{-.218}$, multiplied by 100.

As before, Z represents socioeconomic characteristics of the media market used as control variables. The subscript *b* denotes birth state and *d* the media market of residence.

We estimate this via the iterative GMM estimator, treating sexism and social outcomes as endogenous variables and using as instruments the same variables from our earlier estimations, an instrument for sexism based on spatial correlation and an instrument for social outcomes based on the social outcomes of the birth state. As mentioned above, we no longer include *Children* in our estimations because we did not find that it is robustly related to the wage gap.

This procedure yields interesting conclusions presented in Table 8. The positive and significant coefficient on *Sexism Index* in all three estimations indicates that the sexism of a geographic area is directly related to the wage gap and to the social outcomes. In addition, the fact that the proportion of women married is positively related to the *Home Hours Gap* (column 3) indicates that marriage influences the amount of hours women work outside the home. Thus, when we consider all the ways in which sexism can affect the wage gap, our results suggest that one channel is via the social outcome of *Women Married*, which affects the *Home Hours Gap*, which impacts the gender wage gap. Once we estimate these relationships simultaneously, we no longer find a direct effect of the proportion women who are married on the wage gap. The proportion of women who are married only affects the wage gap via its impact on the *Home Hours Gap*.

The bottom of Table 8 tabulates the direct and indirect effects of sexism. The direct effect still remains substantial at 5.68 cents. But the indirect effect is also non-trivial. Adding all the direct and indirect effects of sexism yields a cost of sexism of 8.12 cents on the hourly wage

gap, equivalent to 41.42 percent of the wage gap. In other words, if the average media market had the lowest level of sexism, the residual wage gap in that media market would close by 8.12 cents.

6 Additional interpretation

6.1 The impact of sexism on male and female wage

We have argued that sexism affects wages of women directly via discrimination and indirectly via influencing social outcomes of women which then impact labor market behavior. If we are correct, then the impact of sexism should make female wages lower and should either have no effect on male wages or make them higher. To explore this idea, we estimate the impact of sexism on both parts of the wage gap, the coefficient on the female media market fixed effect (female wages) and the coefficient on the male media market fixed effect (male wages). In each estimation, we include the coefficient on the media market fixed effect of the other gender as an independent variable to control for the fact that wages overall in some media markets are higher.

The results appear in Table 9. Panel A shows the results in which the coefficient on the female media market effect is the dependent variable and Panel B shows the results in which the coefficient on the male media market fixed effect is the dependent variable. Essentially, this approach allows us to understand better the ways in which sexism of an area affects the wage gap by examining separately the two components of the wage gap. Examining first the results in Panel A, we see that, as expected, the male wage is highly positively correlated with the female wage in the same media market, but the sexism of that media market is also a significant and negative predictor of the female wage in all specifications.

Interestingly, when we examine the male wage in Panel B, to the extent that we find any statistically significant effect, it is positive. The coefficient on sexism is not statistically significant in all estimations, but when it is, it suggests that sexism increases the wages of men. In total, these results are consistent with our interpretations above that the sexism of an area increases the gender wage gap.

6.2 Considering alternative explanations

In this section, we briefly consider two diametrically opposed alternative explanations for our results. First, we note that some religious beliefs are associated with sexism, though, arguably a more benevolent sexism than what we identify in our index. However, to determine if our index actually proxies for religion, we constructed another Google search index from Stephens-Davidowitz (2014), based on the searches for "God." When we enter that as an additional control variable in our estimations, we find that it does not substantially change the coefficient of *Sexism Index* and does not enter the estimations in a statistically significant way, suggesting that our index of hostile sexism does not proxy for religious beliefs. (See Columns 1 and 2 of Appendix Table F.)

Another possibility is that because our index measures hostility towards women, we consider that the index may be picking up hostility in general. In other words, in some media markets, individuals may simply express themselves on the Internet with greater hostility to both men and women. To determine if this is driving our results, we construct two indices that might capture general "hostility searches." The first one is a male sexism index. In other words, we remove the four original words from our sexism dictionary and replace "girl(s)" with "boy(s)" and "woman(en)" with "man(en)." Although all the words in our dictionary do not have obvious male corollaries and the search volumes of the remaining male counterparts are significantly

lower, we are still able to follow the same procedures and construct a male sexism index. We enter that into our estimations as a control variable and find that it does not enter significantly nor does it substantially affect the results for the original *Sexism Index*. (See Columns 3 and 4 of Appendix Table F.) Similarly, we construct a different index that is based solely on Google searches for definitions and meanings of the words in our dictionary. That also enters our estimations insignificantly and does not alter our main results. (See Columns 5 and 6 of Appendix Table F.) Taken together, these results suggest that variation in general hostility expressed on the Internet is not responsible for the association between sexism and the wage gap that we document in this paper.

7 Conclusion

We find strong causal evidence that sexism affects the gender wage gap in a meaningful way. It has a direct effect on the wage gap, which is consistent with labor market discrimination. In addition, we identify a channel through which it has an indirect effect by influencing social outcomes that impact labor market behavior. Interestingly, we find that sexism has a smaller effect in circumstances that would make discrimination more costly. In particular, we do not find evidence that sexism affects the wages of college-educated workers.

The magnitude of the impact is important: our estimates suggest that the average media market has an hourly wage gap that is 8.12 cents higher that it would be if it had the level of sexism of the least sexist area. A reduction of sexism of this magnitude would reduce the residual wage gap by 41 percent.

Our work suggests multiple channels through which cultural attributes of an area can affect economic outcomes. Necessarily, our study focuses only on one type of attitude, hostile

sexism, and a limited number of economic outcomes. Expanding both the types of cultural

attributes and the economic outcomes they may influence is a fruitful area for further research.

References

Akerlof, George A., and Rachel E. Kranton. 2000. "Economics and Identity." *Quarterly Journal of Economics*, *115* (3): 715-753.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn. 2013. "On the Origins of Gender Roles: Women and the Plough." *Quarterly Journal of Economics*, *128* (2): 469-530.

Autor, David, David Dorn, Lawrence F. Katz, Christina Patterson, and John Van Reenen. 2020. "The Fall of the Labor Share and the Rise of Superstar Firms." *Quarterly Journal of Economics*, *135* (2): 645-709.

Bailey, Martha J., Melanie E. Guldi, and Brad J. Hershbein. 2014. "Is There a Case for a 'Second Demographic Transition'? Three Distinctive Features of the Post-1960 US Fertility Decline." In *Human Capital in History: The American Record*, edited by Leah Platt Boustan, Carola Frydman, and Robert A. Margo, 273-312, University of Chicago Press.

Beaudry, Paul, and Ethan Lewis. 2014. "Do Male-Female Wage Differentials Reflect Differences in the Return to Skill? Cross-City Evidence from 1980-2000." *American Economic Journal: Applied Economics*, *6*(2): 178-94.

Becker, Gary. S. 1971. *The Economics of Discrimination*, Second edition. Chicago and London: University of Chicago press, 1957.

Becker, Gary. S. 1985. "Human Capital, Effort, and the Sexual Division of Labor." *Journal of Labor Economics*, 3 (1, Part 2): S33-S58.

Berinsky, Adam. 1999. "The Two Faces of Public Opinion." *American Journal of Political Science*, 43 (October): 1209-1230.

Berinsky, Adam. J. 2002. "Political Context and the Survey Response: The Dynamics of Racial Policy Opinion." *Journal of Politics*, *64*(2): 567-584.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender Identity and Relative Income Within Households." *Quarterly Journal of Economics*, *130*(2): 571-614.

Black, Dan A. 1995. "Discrimination in an Equilibrium Search Model." *Journal of labor Economics*, *13* (2): 309-334.

Black, Sandra E., and Alexandra Spitz-Oener. 2010. "Explaining Women's Success: Technological Change and the Skill Content of Women's Work." *Review of Economics and Statistics*, *92*(1): 187-194.

Black, Dan. A., Natalia Kolesnikova, and Lowell J. Taylor. 2014. "Why do so few women work in New York (and so many in Minneapolis)? Labor supply of married women across US cities." *Journal of Urban Economics*, *79*: 59-71.

Blau, Francine. D., and Lawrence M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature*, *55*(3): 789-865.

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Measurement Error in Survey Data." In *Handbook of Econometrics* (Vol. 5, pp. 3705-3843). Elsevier.

Bursztyn, Leonardo, Thomas Fujiwara, and Amanda Pallais. 2017. "'Acting Wife': Marriage Market Incentives and Labor Market Investments." *American Economic Review*, *107*(11): 3288-3319.

Carlana, Michela. 2019. "Implicit Stereotypes: Evidence from Teachers' Gender Bias." *Quarterly Journal of Economics*, *134*(3): 1163-1224.

Charles, Kerwin K., and Jonathan Guryan. 2008. "Prejudice and Wages: An Empirical Assessment of Becker's The Economics of Discrimination." *Journal of Political Economy*, *116*(5): 773-809.

Charles, Kerwin K., Jonathan Guryan, and Jessica Pan. 2009. "Sexism and Women's Labor Market Outcomes." *Unpublished manuscript, Booth School of Business, University of Chicago*.

Charles, Kerwin K., Jonathan Guryan, and Jessica Pan. 2018. "The Effects of Sexism on American Women: The Role of Norms vs. Discrimination" (No. w24904). National Bureau of Economic Research.

Fernandez, Raquel, and Alessandra Fogli. 2009. "Culture: An Empirical Investigation of Beliefs, Work, and Fertility." *American Economic Journal: Macroeconomics*, *1*(1): 146-77.

Glick, Peter, and Susan T. Fiske. 1996. "The Ambivalent Sexism Inventory: Differentiating Hostile and Benevolent Sexism." *Journal of personality and social psychology*, *70*(3): 491.

Glick, Peter, and Susan T. Fiske. 1997. "Hostile and Benevolent Sexism: Measuring Ambivalent Sexist Attitudes Toward Women." *Psychology of Women Quarterly*, *21*(1): 119-135.

Hall, Kira, and Mary Bucholtz. 2012. *Gender articulated: Language and the socially constructed self*. Routledge.

Hersch, Joni, and Leslie S. Stratton. 1997. "Housework, Fixed Effects, and Wages of Married Workers." *Journal of Human Resources*, 285-307.

Hersch, Joni, and Leslie S. Stratton. 2002. "Housework and Wages." *The Journal of Human Resources*, 217-229.

Hirsch, Boris, Thorsten Schank, and Claus Schnabel. 2010. "Differences in Labor Supply to Monopsonistic Firms and the Gender Pay Gap: An Empirical Analysis Using Linked Employer-Employee Data from Germany." *Journal of Labor Economics* 28(2): 291-330.

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow. 2019. "The Allocation of Talent and U.S. Economic Growth." *Econometrica*, *87*(5): 1439-1474.

Janssen, Simon, Simone Tuor Sartore, and Uschi Backes-Gellner. 2016. "Discriminatory social attitudes and varying gender pay gaps within firms." *ILR Review*, *69*(1): 253-279.

Neumark, David, and Sanders Korenman. 1992. "Sources of bias in women's wage equations: results using sibling data" (No. w4019). National Bureau of Economic Research.

Lundberg, Shelly, and Elaina Rose. 2000. "Parenthood and the earnings of married men and women." *Labour Economics*, *7*(6): 689-710.

Owen, Ann. L., and Wei, Andrew. 2020. "Inequality and Bias in the Demand for and Supply of News." *Social Science Quarterly*, *101*(1), 91-106.

Stephens-Davidowitz, Seth. 2014. "The cost of racial animus on a black candidate: Evidence using Google search data." *Journal of Public Economics*, *118*, 26-40.

Tourangeau, Roger, and Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological bulletin*, *133*(5): 859.

Wright, Gerald, C. 1993. "Errors in Measuring Vote Choice in the National Election Studies, 1952-88." *American Journal of Political Science*, 291-316.

Table 1 Sexism Dictionary

Women as objects	*Women as dumb and emotional*
[word 1](s)	stupid girl(s)
[word 3](s)	stupid woman(en)
[word 4](s)	dumb girl(s)
fat girl(s)	dumb woman(en)
fat woman(en)	emotional girl(s)
ugly girl(s)	emotional woman(en)
ugly woman(en)	

Women as rude and evil
[word 2](es)
cheating woman(en)
cheating girl(s)

Figure 1A

(a) Absolute

$y = 0.0064 + 0.99\,x$

$t = 1.78^*$

Do women often fail to appreciate what men do for them?

(b) Rank

$y = 20 + 0.24\,x$

$t = 1.70^*$

Do women often fail to appreciate what men do for them?

34

Figure 1B

(a) Absolute

$y = -0.072 + 1.6\,x$

t = 2.53**

Do women often consider innocent remarks sexist?

(b) Rank

$y = 16 + 0.39\,x$

t = 2.97***

Do women often consider innocent remarks sexist?
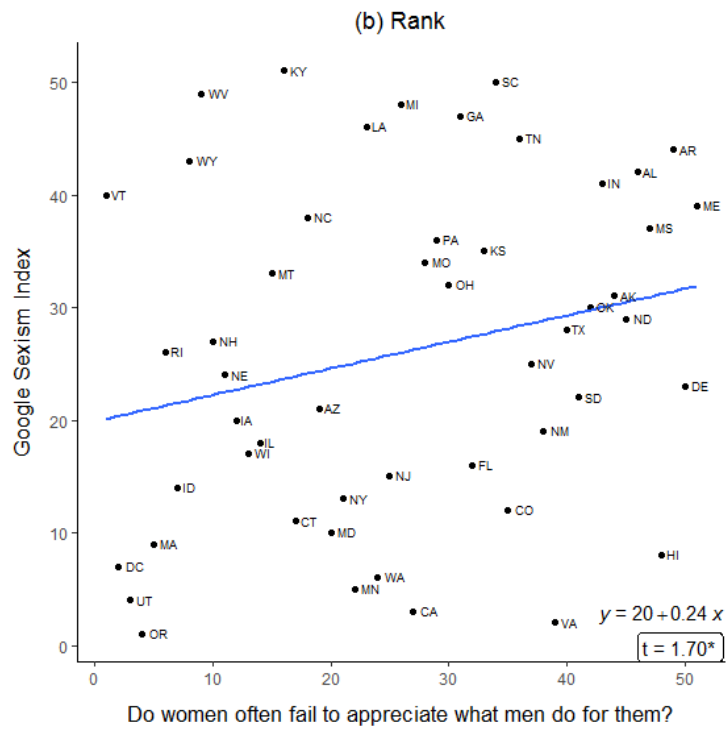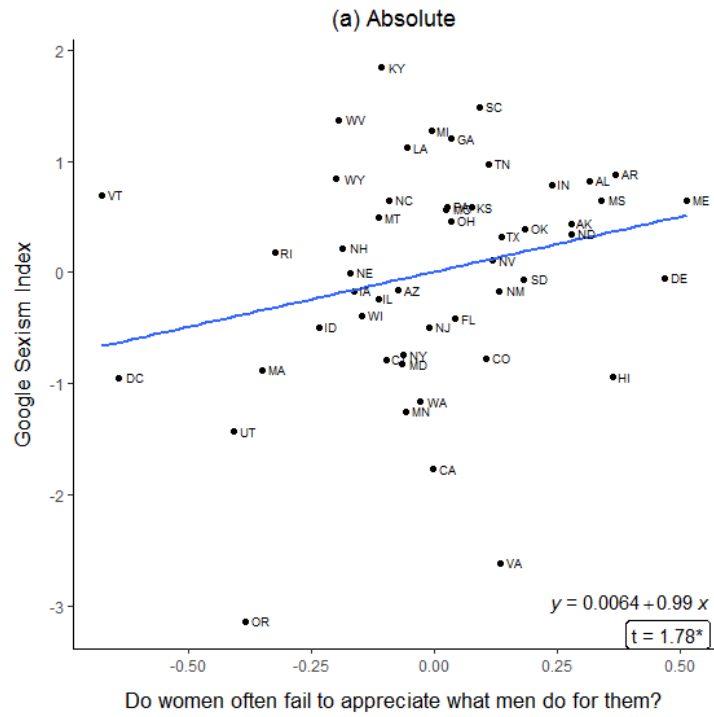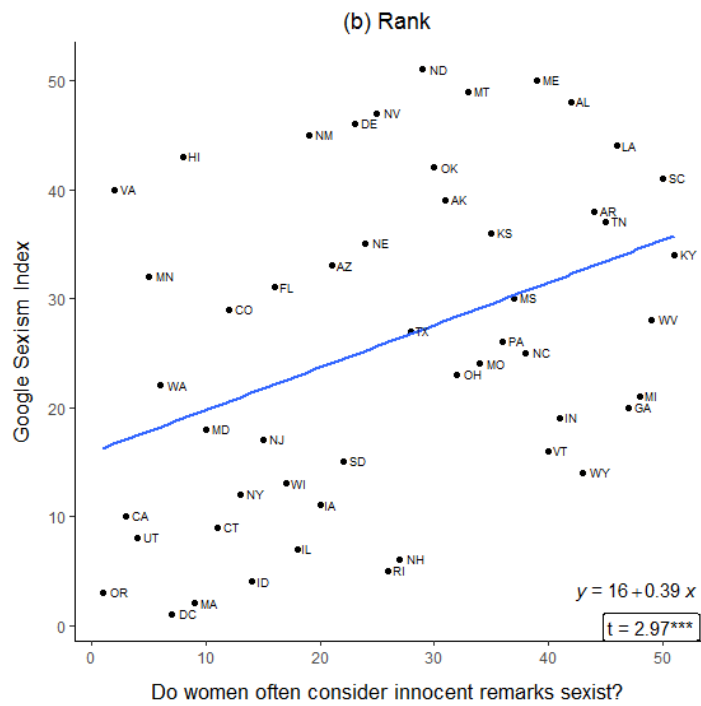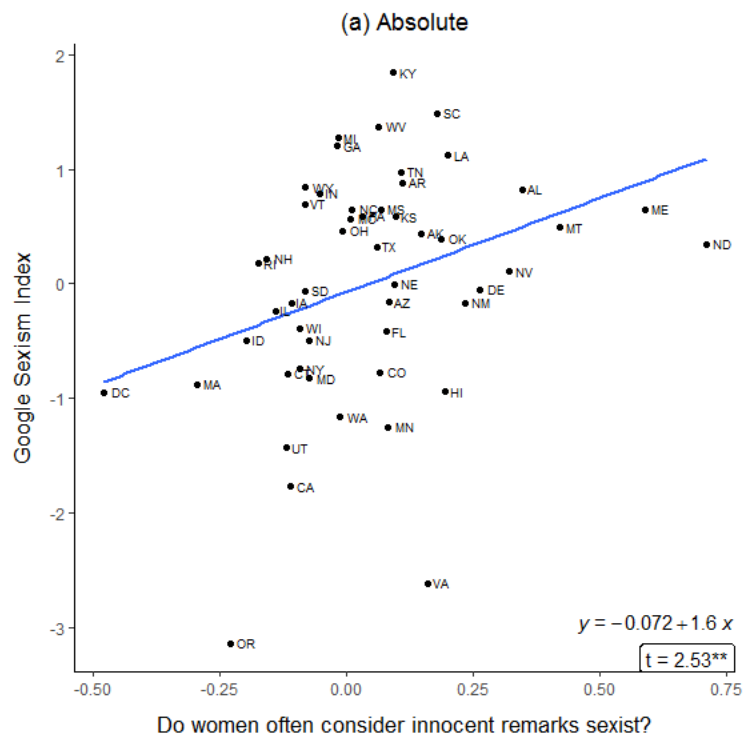
Table 2, Panel A: Cross-sectional Summary Statistics

VARIABLES	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
Dependent Variable					
Wage Gap	174	0.222	0.0344	0.148	0.387
Explanatory Variable					
Sexism Index	174	0.000	1.000	-4.408	3.037
%Women Married	174	0.659	0.0331	0.578	0.767
Children	174	0.647	0.0648	0.425	0.767
Home Hours Gap	174	6.258	0.985	3.926	10.27
Control Variables					
%Employed	174	0.750	0.0406	0.646	0.858
%Female Employed	174	0.692	0.0493	0.559	0.826
LFP	174	0.787	0.0356	0.702	0.876
Female LFP	174	0.725	0.0452	0.624	0.838
College Gap	174	-0.0211	0.0200	-0.104	0.0544

Notes: Each observation is a media market

Table 2, Panel B: Panel Summary Statistics

VARIABLES	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
Dependent Variable					
Wage Gap	2,144	0.220	0.0470	0.0668	0.482
Explanatory Variable					
Sexism Index	2,118	0.000	1.000	-2.744	4.318
%Women Married	2,144	0.660	0.0397	0.519	0.794
Children	2,144	0.648	0.0726	0.382	0.869
Home Hours Gap	2,144	6.259	1.244	1.409	12.87
Control Variables					
%Employed	2,144	0.749	0.0435	0.593	0.883
%Female Employed	2,144	0.690	0.0530	0.511	0.871
LFP	2,144	0.788	0.0375	0.658	0.897
Female LFP	2,144	0.724	0.0485	0.572	0.871
College Gap	2,144	-0.0190	0.0308	-0.145	0.0807

Notes: Each observation is a media market in a year.

Table 2, Panel C: Summary Statistics for Instrumental Variables Estimation

VARIABLES	(1) N	(2) Mean	(3) SD	(4) Min	(5) Max
Dependent Variable					
Wage Gap	7,433	0.218	0.270	-2.579	4.176
Instruments					
Sexism of closest media market	7,433	-0.0162	1.085	-4.408	3.037
Sexism of remaining (weighted)	7,433	0.0581	0.136	-0.307	0.317
%Women Married (birth)	7,433	0.643	0.0288	0.576	0.726
Children (birth)	7,433	0.573	0.0422	0.471	0.706
Home Hours Gap (birth)	7,433	6.000	0.572	4.516	8.845
Explanatory Variable					
Sexism Index	7,433	-0.0920	1.012	-4.408	3.037
%Women Married	7,433	0.664	0.129	0	1
Children	7,219	0.595	0.210	0	1
Home Hours Gap	7,433	6.447	5.069	-60	60.50
Control Variables					
%Employed	7,433	0.756	0.0872	0.222	1
%Female Employed	7,433	0.693	0.122	0	1
LFP	7,433	0.791	0.0820	0.222	1
Female LFP	7,433	0.724	0.118	0	1
College Gap	7,433	0.00209	0.179	-1	1

Notes: Each observation is a pair composed of a birth state and a media market of residence.

Figure 2: Geographic Variation in the Residual Wage Gap
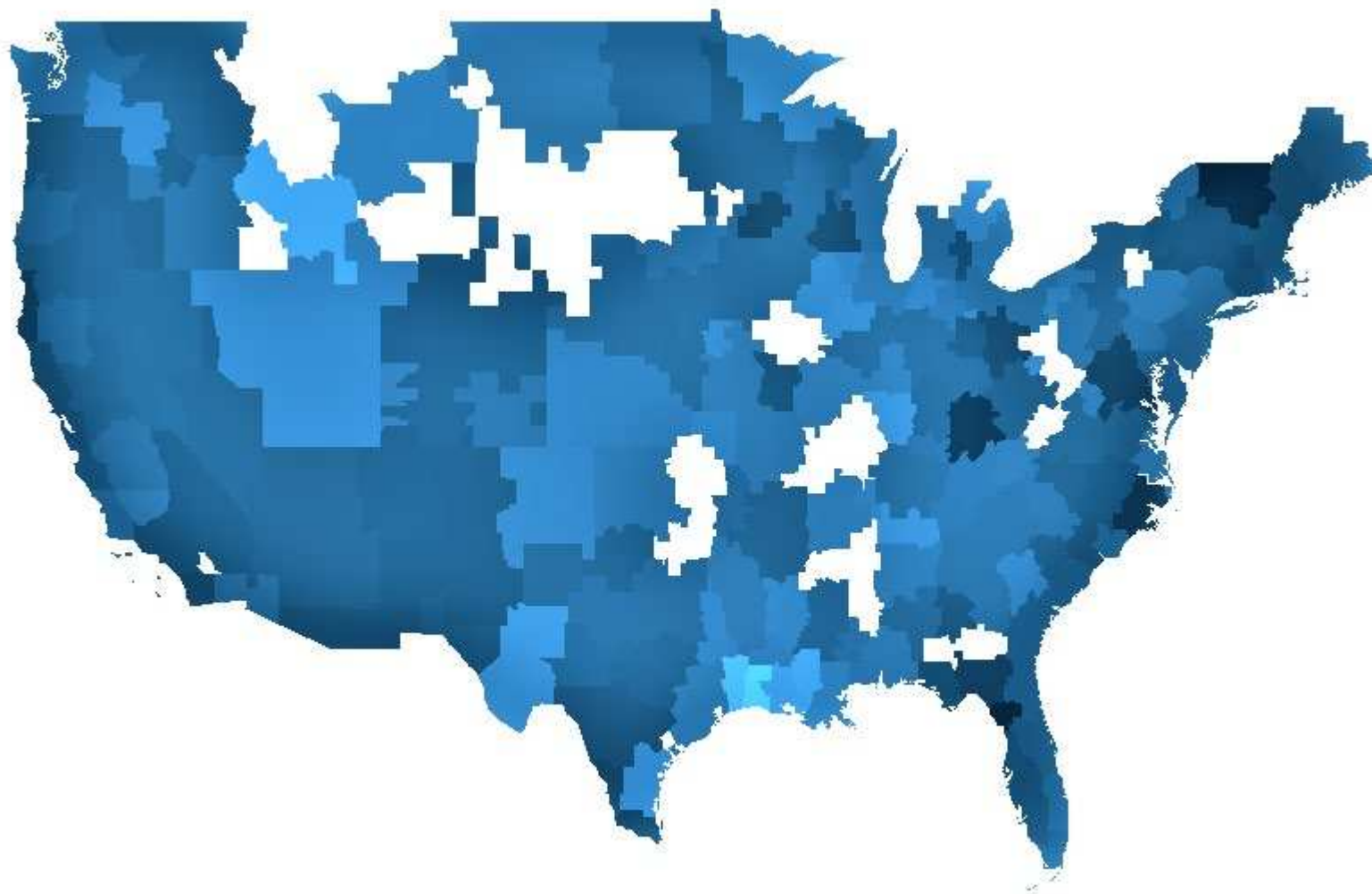
Notes: Darker areas signify a wider wage gap. White areas signify areas without data.

Table 3. Cross-sectional Estimates of the Impact of Sexism.

VARIABLES	Dependent Variable: Residual Wage Gap				
	(1)	(2)	(3)	(4)	(5)
Sexism Index	0.00875***	0.00981***	0.00524**	0.00611*	0.00658***
	(0.00244)	(0.00229)	(0.00220)	(0.00313)	(0.00196)
Women Married			0.380***		
			(0.0854)		
Children				0.140***	
				(0.0490)	
Home Hours Gap					0.0145***
					(0.00241)
%Employed	0.733*	-3.154***	-2.445***	-2.778***	-1.549
	(0.424)	(0.934)	(0.882)	(0.881)	(0.947)
LFP	-0.836	4.441***	3.267***	4.037***	2.137*
	(0.535)	(1.149)	(1.093)	(1.115)	(1.193)
%Female Employ.		3.942***	2.524**	3.307***	2.024**
		(0.933)	(1.066)	(0.908)	(0.994)
Female LFP		-5.134***	-3.373***	-4.378***	-2.634**
		(1.112)	(1.252)	(1.097)	(1.192)
College Gap	0.250	-0.275	-0.273	-0.251	-0.0553
	(0.151)	(0.213)	(0.222)	(0.214)	(0.202)
Observations	174	174	174	174	174
R-squared	0.172	0.319	0.410	0.372	0.447
Change in the Impact of Sexism					
Δ in coefficient from Col. 2			-0.00458***	-0.00370**	-0.00324***
			(0.000988)	(0.00162)	(0.00102)
% change in magnitude			46.63	37.75	32.97
Costs of Sexism					
In cents	3.15	3.54	1.87	2.19	2.36
As % of wage gap	15.85	17.82	9.41	11	11.86

Notes All specifications are weighted by the inverse of the standard error of the wage gap. Standard errors are robust to arbitrary heteroscedasticity and within-state error correlation. Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 4. The Impact of Sexism by Group.

VARIABLES	Dependent Variable: Residual Wage Gap			
	(1)	(2)	(3)	(4)
Sexism Index	0.0152***	0.00567**	0.0147***	0.00785***
	(0.00291)	(0.00222)	(0.00245)	(0.00227)
Sexism * College Educated	-0.0154***			
	(0.00349)			
College Educated	-0.0714***			
	(0.00704)			
Sexism * Manufacturing Industry		0.00981**		
		(0.00370)		
Sexism * Commerce Industry		0.00558***		
		(0.00195)		
Manufacturing Industry		0.0398***		
		(0.00424)		
Commerce Industry		0.0239***		
		(0.00344)		
Sexism * Diverse Industry			-0.00830***	
			(0.00270)	
Diverse Industry			-0.0330***	
			(0.00497)	
Sexism * Black				0.00855***
				(0.00293)
Black				-0.0823***
				(0.00385)
Observations	348	522	348	348
R-squared	0.571	0.365	0.379	0.625

Notes: All specifications control for *%Employed*, *LFP*, *%Women employed*, *Female LFP*, and *College Gap*. The wage gaps for each education or industry group are estimated separately and then stacked at the media market level. The variable *Diverse Industry* represents industries with a *%Female* above the median. All specifications are weighted by the inverse of the standard error of the wage gap. Standard errors are robust to arbitrary heteroscedasticity and within-state error correlation. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Cross-sectional, IV Estimates of the Impact of Sexism on the Wage Gap

VARIABLES	Dependent Variable: Residual Wage Gap					
	(1)	(2)	(3)	(4)	(5)	(6)
Sexism Index	0.0155***	0.0185***	0.0161**	0.0224***	0.00858**	0.0123***
	(0.00574)	(0.00496)	(0.00791)	(0.00864)	(0.00416)	(0.00410)
Specification	Baseline		Exclude areas within 100 mi		Control for nearby wage gaps	
Hansen P-value	.98	.43	.86	.37	.16	.12
F-statistic (1st stage)	14.31	12.17	6.74	5.71	17.57	13.39
Observations	174	174	174	174	174	174
R-squared	0.126	0.244	0.118	0.161	0.336	0.381
Controls						
College Gap	X	X	X	X	X	X
Econ. Outcomes	X	X	X	X	X	X
Fem. Econ. Outcomes		X		X		X

Notes: Economic outcomes include *%Employed* and *LFP*. Female economic outcomes include *%Female Employed* and *Female* LFP. As two instruments for a media market's sexism, we use the sexism of the closest media market and the average sexism of all other media markets, weighted by the inverse of the distance. Columns 3 and 4 account for potential spillover by excluding media markets within 100 miles from the instruments. Columns 5 and 6 control for the wage gap of the closest media market and for the average wage gap of all other media markets, weighted by the inverse of the distance. All specifications are weighted by the inverse of the standard error of the wage gap. Standard errors are clustered at the state level. Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 6: Panel Specification

VARIABLES	Dependent Variable: Residual Wage Gap				
	(1)	(2)	(3)	(4)	(5)
Sexism	0.00307*	0.00282*	0.00275*	0.00281*	0.00280*
	(0.00162)	(0.00153)	(0.00156)	(0.00151)	(0.00152)
Women married			0.0788		
			(0.0600)		
Children				0.00431	
				(0.0365)	
Home Hours Gap					0.00131
					(0.00145)
%Employed	0.222**	0.00846	-0.00361	0.00807	-0.0216
	(0.0833)	(0.170)	(0.166)	(0.171)	(0.174)
LFP	-0.152	0.546**	0.532**	0.546**	0.566**
	(0.106)	(0.206)	(0.205)	(0.206)	(0.210)
Female LFP		-0.626***	-0.609***	-0.626***	-0.635***
		(0.167)	(0.169)	(0.167)	(0.170)
%Women employed		0.222	0.226	0.223	0.244
		(0.164)	(0.163)	(0.169)	(0.170)
College Gap	-0.0128	-0.0259	-0.0230	-0.0260	-0.0280
	(0.0372)	(0.0358)	(0.0357)	(0.0359)	(0.0355)
Media market & Year FE	X	X	X	X	X
Observations	2,118	2,118	2,118	2,118	2,118
R-squared	0.561	0.567	0.567	0.567	0.567

Notes: Each observation is a media market in a year. All specifications are weighted by the inverse of the standard error of the wage gap. Standard errors are robust to arbitrary heteroscedasticity and error correlation within a state. Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 7: 2SLS Estimates of the Indirect Effect of Sexism

Dependent variables:	Resid. Wage gap (1)	Women Married (2)	Resid. Wage Gap (3)	Children (4)	Resid. Wage Gap (5)	Home Hours Gap (6)
Sexism Index	0.0201***	0.0260***	0.0244***	0.0307*	0.0163***	0.458***
	(0.00580)	(0.00761)	(0.00612)	(0.0181)	(0.00428)	(0.164)
Women Married	0.187***					
	(0.0725)					
Children			0.00811			
			(0.0766)			
Home Hours Gap					0.0194***	
					(0.00693)	
Observations	7,433	7,433	7,219	7,219	7,433	7,433
R-squared	0.008	0.058	0.002	0.170	-0.129	0.044
Sargan-Hansen P-value	.77	.5	.86	.99	.99	.82
F-statistic (1st stage)	100.03	35.38	23.46	35.32	18.53	35.38
			Indirect costs of Sexism			
Via Women Married	1.74 cents		Responsible for 19 percent of the total impact of sexism			
Via Children	.09 cents		Responsible for 0.96 percent of the total impact of sexism			
Via Home Hours Gap	3.21 cents		Responsible for 35 percent of the total impact of sexism			

Notes: All specifications control for *%Employed*, *LFP*, *%Women employed*, *Female LFP*, and *College Gap*. The instruments for sexism are the sexism of the closest media market and the average sexism of all other media markets, weighted by the inverse distance. The instrument for a social outcome is the average social outcome of the birth state. All specifications are weighted by the inverse of the standard error of the wage gap. Standard errors are robust to arbitrary heteroscedasticity and within-state error correlation. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table 8: GMM Estimates. Marriage and Home Hours Difference Together.

Dependent variable	Residual wage gap	Women Married	Home Hours Gap
	(1)	(2)	(3)
Sexism Index	0.0155***	0.0209***	0.279**
	(0.00398)	(0.00615)	(0.128)
Women Married	0.0668		7.826***
	(0.0752)		(1.770)
Home Hours Gap	0.0154**		
	(0.00597)		
Residual Wage Gap		0.0206**	-0.513*
		(0.00836)	(0.274)
Origin Wage Gap	0.371***	-0.00312	2.248***
	(0.0596)	(0.0348)	(0.864)
%Employed	-0.126	0.455***	7.799***
	(0.179)	(0.106)	(2.700)
LFP	0.237	0.00207	3.827
	(0.208)	(0.103)	(2.599)
%Female Employ.	0.125	0.110*	-7.252***
	(0.116)	(0.0661)	(1.464)
Female LFP	-0.163	-0.493***	-1.819
	(0.124)	(0.0754)	(1.723)
College Gap	-0.00261	-0.00180	2.174***
	(0.0239)	(0.0128)	(0.426)
Observations	7,433	7,433	7,433
Sargan-Hansen P-value		.796	

Costs of sexism		
Direct effect of sexism	5.68 cents	29 percent
Indirect effect via Home Hours Gap	1.54 cents	7.84 percent
Indirect effect via Women Married and Hhrs.	.9 cents	4.58 percent
Total indirect effect	2.44 cents	12.42 percent
Total net effect (direct + indirect)	8.12 cents	41.42 percent
Indirect effect as a % of total effect		30 percent

Notes: We use the iterative GMM estimator. The instruments for sexism are the sexism of the closest media market and the average sexism of all other media markets, weighted by the inverse distance. The instrument for a social outcome is the average social outcome of the birth state. All specifications are weighted by the inverse of the standard error of the wage gap. Standard errors are robust to arbitrary heteroscedasticity and within-state error correlation. Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 9: Cross-sectional, OLS Estimates of the Impact of Sexism on Female and Male Wage

VARIABLES	(1)	(2)	(3)	(4)	(5)
Panel A: Impact of sexism on female wage					
Sexism Index	-0.0119***	-0.0124***	-0.00777***	-0.00700***	-0.00897***
	(0.00226)	(0.00237)	(0.00189)	(0.00214)	(0.00193)
Male wage	0.846***	0.898***	0.885***	0.794***	0.902***
	(0.0627)	(0.0563)	(0.0465)	(0.0557)	(0.0515)
Observations	173	173	173	173	173
R-squared	0.902	0.914	0.926	0.936	0.927
Panel B: Impact of sexism on male wage					
Sexism Index	0.00651*	0.00772*	0.00446	0.00573	0.00522
	(0.00362)	(0.00421)	(0.00317)	(0.00427)	(0.00351)
Female wage	0.941***	0.904***	0.936***	0.991***	0.928***
	(0.0546)	(0.0516)	(0.0471)	(0.0448)	(0.0492)
Observations	173	173	173	173	173
R-squared	0.875	0.901	0.911	0.909	0.914
Controls					
College Gap	X	X	X	X	X
Economic Outcomes	X	X	X	X	X
Female Economic Outcomes		X	X	X	X
Women Married			X		
Children				X	
Home Hours Gap					X

Notes: Economic outcomes include *%Employed* and *LFP*. Female economic outcomes include *%Female Employed* and *Female LFP*. *Female Wage* is defined as the coefficient on the female media market fixed effect. *Male wage* is defined as the coefficient on the male media market fixed effect. All specifications are weighted by the inverse of the standard error of either *Female Wage* or *Male Wage*. All standard errors are robust to arbitrary heteroscedasticity and error correlation within a state. Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

Appendix A: Construction of the Sexism Index

Coding for Offensive Language

Word	Code in text
cunt	word 1
bitch	word 2
slut	word 3
whore	word 4

Additional Technical Details

We use Google Trends to obtain the relative search volume of phrases in our sexism dictionary.

Google Trends returns an index based on the relative search volume of specific searches within a

media market such that the media market that has the highest relative search volume receives a

100. Google Trends returns a value of 0 if the absolute search volume of a word falls below an

unknown but very low privacy threshold. We employ an algorithm from Stephens-Davidowitz

(2014) to address this. For each word "x", we download 300 samples of "x", "weather", and

"weather $+ x$", dropping any values of "x" and "weather" that are either 0 or 100. Then, we

impute the missing values of "x" by regressing the non-missing values of "x" on "weather" and

"weather $+ x$". To aggregate the samples, we take the average across all 300 imputations.

Sexism Index summarizes the 2004-2019 search volumes of 16 phrases in our sexism

dictionary. To construct it, we conduct a principal components analysis on the Google Trends

index based on relative search volumes. The index is defined as the sum of all principal

components with eigenvalue over one, weighted by eigenvalue. To ease interpretation, we

normalize the resulting values of the index for each media market by subtracting the index by its

mean and dividing by its standard deviation.

Out of 16 principal components, only the first four have eigenvalues over one, explaining

83 percent of total variation. Through principal component analysis, *Sexism Index* should give

heavier weight to phrases more related to sexism, reducing the impact of variation unrelated to

sexism. The normalization of the *Sexism Index* implies that it has a mean of zero and standard deviation of one. The final variable measures the number of standard deviations an area's sexist search volume is from the mean.

Appendix B: Variable Definitions

Variable	Definition
Individual characteristics	
Age Quartic	A quartic polynomial for the current age of an individual in years.
Schooling FE	A set of dummy variables for each year of schooling.[19]
Industry FE	A set of dummy variable for each of the 300 census industry classifications in our sample.
Occupation Type FE	A set of dummy variable for the four occupation categories: management and business, science, service, and manufacturing. Constructed manually using 2010 census classifications.
Media Market Social Outcomes	
Women Married	The percent of white women between the ages of 25 and 64 currently married in a media market.
Children	The percent of white women between the ages of 25 and 40 with at least one child at home in a media market.[20]
Home Hours Gap	The average home hours of white women between the ages of 25 and 64 minus the average home hours of white men between the ages of 25 and 64 in a media market.[21]
Media Market Socioeconomic Controls	
%Employed	The percent of whites between the ages of 25 and 64 currently employed in a media market.
%Female Employed	The percent of white women between the ages of 25 and 64 currently employed in a media market.
LFP	The percent of whites between the ages of 25 and 64 participating in the labor force in a media market.
Female LFP	The percent of white women between the ages of 25 and 64 participating in the labor force in a media market.
College Gap	The percent of white men between the ages of 25 and 64 with a college education minus the percent of white women between the ages of 25 and 64 with a college education in a media market

---

[19] Since the ACS does not include a literal "years of education" variable, we impute it using the midpoint of each category.

[20] Restricting the sample to women younger than 40 maximizes the probability that a woman's eldest child is still at home.

[21] Because the ACS does not include an explicit variable for "home hours", we construct it by subtracting usual hours worked per week from 168, the total number of hours in a week.

Appendix C: Ranking of Residual Wage Ratios

Media markets with the 10 smallest gaps	Ratio	Media markets with the 10 largest gaps	Ratio
Gainesville FL	86.23872	Lake Charles LA	67.94209
Burlington VT-Plattsburgh NY	86.16167	Beaumont-Port Arthur TX	70.42979
Lexington KY	86.02189	Idaho Falls-Pocatello ID	72.61327
Harlingen-Weslaco-Brownsville-McAllen TX	85.05079	Baton Rouge LA	72.87431
Rochester MN-Mason City IA-Austin MN	84.84537	Lafayette LA	73.42238
Greenville-New Bern-Washington NC	84.81471	Corpus Christi TX	75.58092
Charlottesville VA	84.30365	Evansville IN	75.60896
Washington DC (Hagerstown MD)	84.29196	Odessa-Midland TX	75.90942
Eureka CA	84.11975	Huntsville-Decatur (Florence) AL	75.98314
Tallahassee FL-Thomasville GA	83.85233	Peoria-Bloomington IL	76.29172

Notes: Wage ratios are adjusted by individual characteristics.

Appendix D: Individual Level Results

VARIABLES	Residualized log wages	
	(1)	(2)
Female dummy	-0.164***	-0.163***
	(0.00310)	(0.00330)
Sexism	-0.0345***	-0.0267***
	(0.0110)	(0.00900)
Sexism * Female	-0.00629***	-0.00566***
	(0.00168)	(0.00190)
% Employed	0.0322***	0.0341***
	(0.00437)	(0.00475)
LFP	1.022**	3.581***
	(0.412)	(0.591)
%Women employed		-6.920***
		(0.834)
Female LFP		5.291***
		(0.946)
College gap	0.584	-0.847
	(0.641)	(0.656)
Observations	7,542,754	7,542,754
R-squared	0.028	0.034

Notes: The dependent variable is log wages, residualized by a quartic for age and fixed effects for year, years of schooling, industry, and occupation type. All specifications are weighted by the inverse of this residual. All standard errors are robust to arbitrary heteroscedasticity and within-state error correlation. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix E: First Stage Results

1st stage dependent variable	Sexism	Women Ma.	Sexism	Sexism	Children	Sexism	Sexism	Hhrs. Gap.	Sexism
2nd stage dependent variable	Wage Gap	Wage Gap	Women Ma.	Wage Gap	Wage Gap	Children	Wage Gap	Wage Gap	Hhr Gap
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Instruments									
Sexism of closest media market	0.214***	0.00302	0.212***	0.213***	0.00593	0.211***	0.214***	0.110*	0.212***
	(0.0645)	(0.00216)	(0.0647)	(0.0647)	(0.00553)	(0.0650)	(0.0647)	(0.0577)	(0.0647)
Average (weighted) sexism of remaining	2.528***	0.0721***	2.507***	2.550***	0.0811	2.526***	2.532***	1.018*	2.507***
	(0.401)	(0.0253)	(0.405)	(0.402)	(0.0766)	(0.406)	(0.398)	(0.591)	(0.405)
Women Married of birth state	0.0517	0.588***							
	(0.390)	(0.0328)							
Children of birth state				0.420	0.522***				
				(0.437)	(0.0559)				
Home hours of birth state							-0.0572**	0.475***	
							(0.0269)	(0.0711)	
Control variables									
%Employed	-2.854	0.259***	-2.862	-2.808	-0.122	-2.791	-2.771	9.974***	-2.862
	(2.097)	(0.0910)	(2.074)	(2.182)	(0.151)	(2.188)	(2.082)	(2.729)	(2.074)
LFP	-4.258**	0.0518	-4.327***	-4.535**	-0.329*	-4.621***	-4.300***	1.928	-4.327***
	(1.665)	(0.104)	(1.610)	(1.813)	(0.185)	(1.776)	(1.640)	(2.743)	(1.610)
%Female employed	-0.424	0.107	-0.428	-0.592	0.0159	-0.571	-0.397	-6.656***	-0.428
	(0.665)	(0.0674)	(0.659)	(0.738)	(0.130)	(0.723)	(0.664)	(1.682)	(0.659)
Female LFP	1.175	-0.454***	1.258	1.370	-0.254**	1.396	1.082	-4.369**	1.258
	(0.853)	(0.0752)	(0.844)	(0.937)	(0.128)	(0.946)	(0.842)	(2.046)	(0.844)
College difference	0.0250	-0.0159	0.0147	0.0271	0.0393	0.0254	0.0600	1.805***	0.0147
	(0.156)	(0.0137)	(0.156)	(0.172)	(0.0275)	(0.172)	(0.154)	(0.482)	(0.156)
Origin wage gap	-0.462	0.0842***	-0.601	-0.480	-0.0178	-0.589	-0.376	1.004	-0.601
	(0.411)	(0.0326)	(0.413)	(0.405)	(0.0534)	(0.422)	(0.390)	(0.937)	(0.413)
Observations	7,433	7,433	7,433	7,219	7,219	7,219	7,433	7,433	7,433

Notes: The results show the first stage of the 2SLS specification. The instruments for sexism are the sexism of the closest media market and the average sexism of all other media markets, weighted by the inverse distance. The instrument for a social outcome is the average social outcome of the birth state. All specifications are weighted by the inverse of the standard error of the wage gap. Standard errors are robust to arbitrary heteroscedasticity and within-state error correlation. Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Appendix F: Considering alternative explanations

VARIABLES	Dependent Variable: Residual Wage Gap					
	(1)	(2)	(3)	(4)	(5)	(6)
Sexism Index	0.00689**	0.00817***	0.0102**	0.0110***	0.00897***	0.00892***
	(0.00296)	(0.00229)	(0.00383)	(0.00358)	(0.00231)	(0.00216)
Searches for God	0.00521	0.00450				
	(0.00388)	(0.00345)				
Male Sexism Index			-0.00163	-0.00132		
			(0.00444)	(0.00377)		
Sexist Definitions Index					0.000833	-0.00380
					(0.00367)	(0.00317)
% Employed	0.467	-3.667***	0.733*	-3.168***	0.749	-3.191***
	(0.440)	(0.867)	(0.424)	(0.943)	(0.450)	(0.926)
LFP	-0.515	4.974***	-0.831	4.458***	-0.852	4.600***
	(0.540)	(1.067)	(0.525)	(1.163)	(0.562)	(1.076)
% Female Employed		4.284***		3.959***		3.847***
		(0.859)		(0.950)		(1.005)
Female LFP		-5.453***		-5.150***		-5.145***
		(1.021)		(1.128)		(1.142)
College Difference	0.188	-0.302	0.244	-0.278	0.233	-0.261
	(0.179)	(0.211)	(0.162)	(0.218)	(0.191)	(0.224)
Observations	174	174	174	174	174	174
R-squared	0.183	0.326	0.173	0.319	0.173	0.329

Notes: Searches for God is the search volume of "God", normalized by Z-score. *Male Sexism Index* takes *Sexism Index*, removes the original four words, and replaces "girl(s)" with "boy(s)" and "man(en)" with "woman(en)". *Sexist Definitions Index* adds "definition" and "meaning" to each entry in *Sexism Index*. All specifications are weighted by the inverse of the standard error of the wage gap. Standard errors are robust to arbitrary heteroscedasticity and within-state error correlation. Robust standard errors in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$.