
A Time-Series Analysis on the Behaviors and Attitudes Towards Women as Economic Agents

Meredith Mante

Lucia Yu

Motivation and Background References

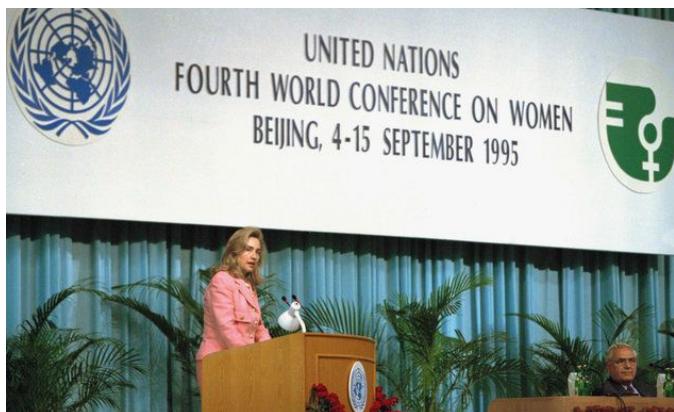
how much progress has really been made for women as economic agents in the United States?

Corporate-talent pipeline by gender



Source: Data for 2012 are from *Unlocking the full potential of women at work*, in which McKinsey examined the employee pipeline of 60 US corporations. Data for 2015 are from *Women in the Workplace*, in which LeanIn.Org and McKinsey examined the employee pipeline of 118 US corporations.

McKinsey&Company



Don't Buy Into The Gender Pay Gap Myth



Karin Agness, [USA30 CONTRIBUTOR](#)

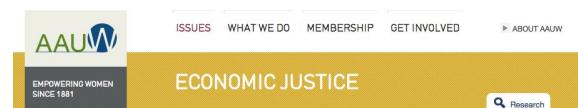
I cover the intersection of women, policy and politics.

[FULL BIO](#) ▾

Opinions expressed by Forbes Contributors are their own.



Senate Appropriations Committee Chair Barbara Mikulski, D-Md., at podium, accompanied by fellow Senate Democrats, speaks during a news conference on Capitol Hill in Washington, Wednesday, Sept. 10, 2014, to discuss the Paycheck Fairness Act. (AP Photo/Cliff Owen)



The Simple Truth about the Gender Pay Gap (Fall 2016)



What is the Problem?

In this project, we examine how attitudes and beliefs about women in the workforce change over time, and how that may correlate with changes in circumstances of women's role in the workforce.

Using data from respondents from the General Social Survey (GSS) by the US Census Bureau from 1974 to 2014, we regress metrics on:

- 1) the overall percent change inflation-adjusted income of women in the workforce for each year,
- 2) average job prestige of women-held occupations per year,
- 3) the percent change in support of women in the labor force and affirmative action measures from all respondents, and
- 4) the percent change of job satisfaction for female respondents in the workforce on the percent of women in the workforce.

Describing the Data

the General Social Survey (GSS)



We used historical survey data from the GSS, conducted by the US Census Bureau from 1972-2014.

- The GSS is a 90-minute in-person interview that uses a full-probability, representative sampling of the English-speaking US adult population, with a response rate of 70% more than other social surveys [1]
- Not all questions are asked all years; only the core questions are covered for all years
- Until 1994, it was conducted almost annually (due to funding limitations, there were no GSSs in 1979, 1981, or 1992). Since 1994, the GSS has been conducted in even numbered years

We used the GSS Data Explorer to export the variables to be examined

- The GSS Data Explorer developed by the University of Chicago [2]

[1] https://www.nsf.gov/pubs/2007/nsf0748/nsf0748_3.pdf

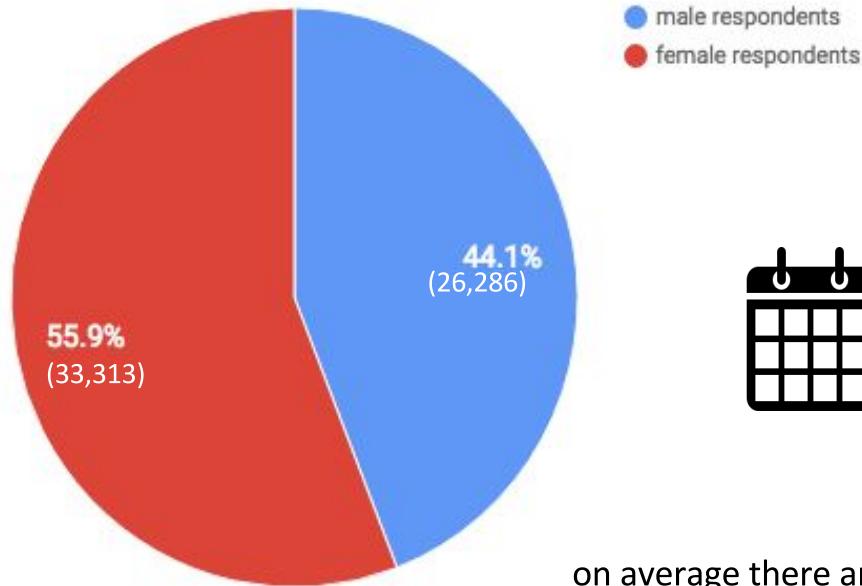
[2] <https://gssdataexplorer.norc.org/>

Describing the Data

general statistics

Over the span of 40 years of GSS data for the survey questions that we selected,

Total Number of Respondents for All Years: **58,599**



on average there are about **1,923** respondents per year



we examined years **1974 to 2014** of GSS data



72% of male respondents surveyed in workforce

53% of female respondents surveyed in workforce



Describing the Data

overview of variables

	DATA TYPE	DATA SCALE	DESCRIPTION
year	discrete	1972-2014	missing for 1974, 1979, 1981, 1992, 2001, 2003, 2005, 2007, 2009, 2011, 2013
gender	binary	female, male	
job category	nominal	n/a	job categories expand in 1980 and 2010
real income	continuous	\$0 - \$500,000	adjusted for inflation; examined women and men
prestige score	discrete	0-97	developed by Robert W. Hodge, Paul S. Siegel, and Peter H. Rossi; a standardized, subjective aggregate score of occupations based on relative social standing; score to be examined for jobs held by women; missing for 2010, 2012, 2014
"On the whole, how satisfied are you with the work you do?"	ordinal	very satisfied, mod. satisfied, a little dissatisfied, very dissatisfied; don't know, no answer, n/a	respondent's perception of current job satisfaction; will be analyzed as a percentage of total women participating in workforce; very satisfied and moderately satisfied will be in bucket 'satisfied,' little dissatisfied and very dissatisfied will be in bucket 'dissatisfied,' and no answer, n/a and don't know will be not included

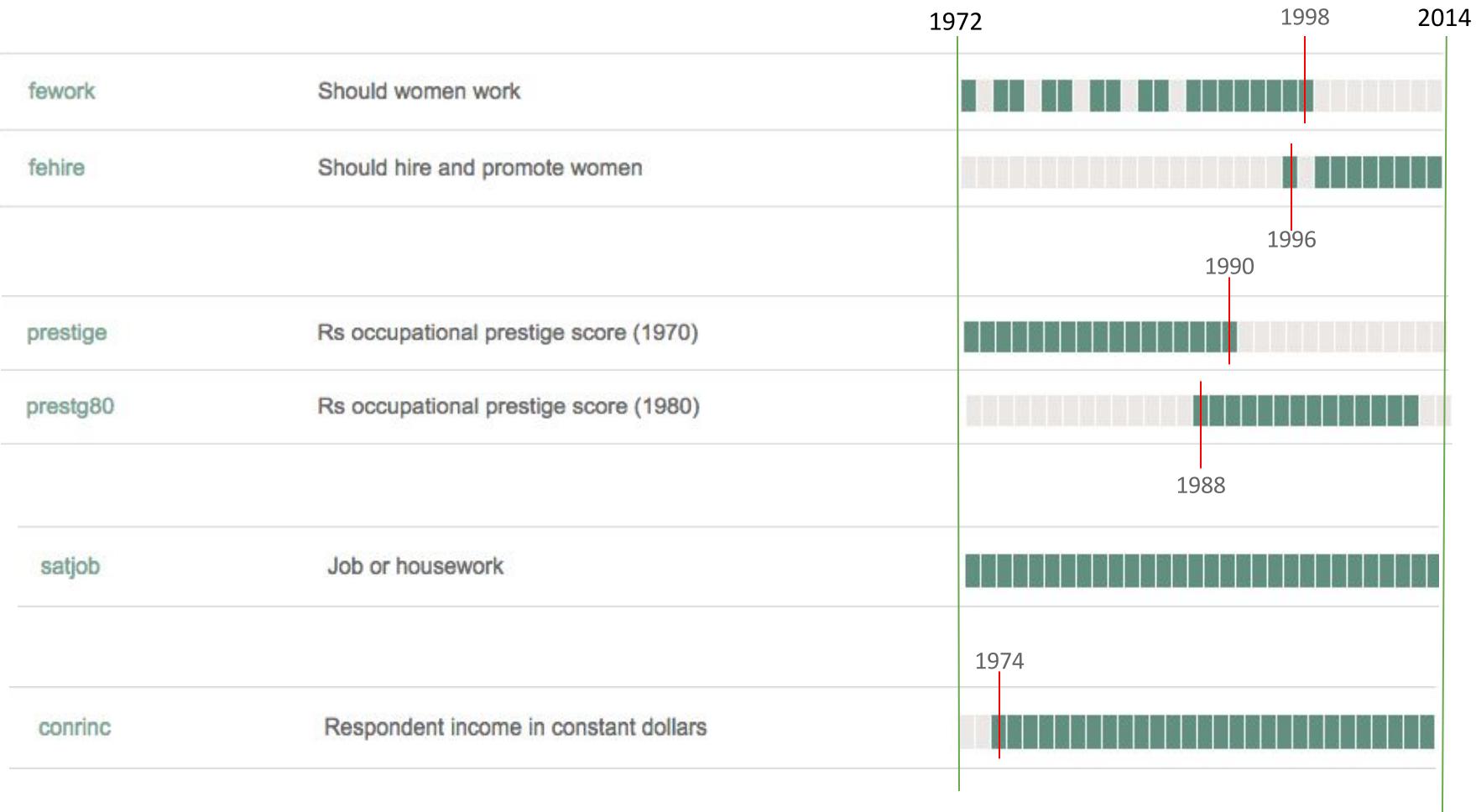
Describing the Data

overview of variables

	DATA TYPE	DATA SCALE	DESCRIPTION
"Do you approve or disapprove of a married woman earning money in business or industry if she has a husband capable of supporting her?"	binary	approve, disapprove	in GSS 1970-1998; this question was discontinued after 1980; will be analyzed as a percentage of the total of those who are in all job fields
"Because of past discrimination, employers should make special efforts to hire and promote qualified women."	ordinal	strongly agree, agree, neither, disagree, strongly disagree	in GSS 1996-2014; this question first appeared in 1980; strongly agree and agree will be in bucket 'approve,' disagree and strongly disagree will be in 'disapprove' and neither will be neutral; buckets will be analyzed as percent of both men and women in all fields in conjunction with the approval and disapproval rates in previous question

Describing the Data

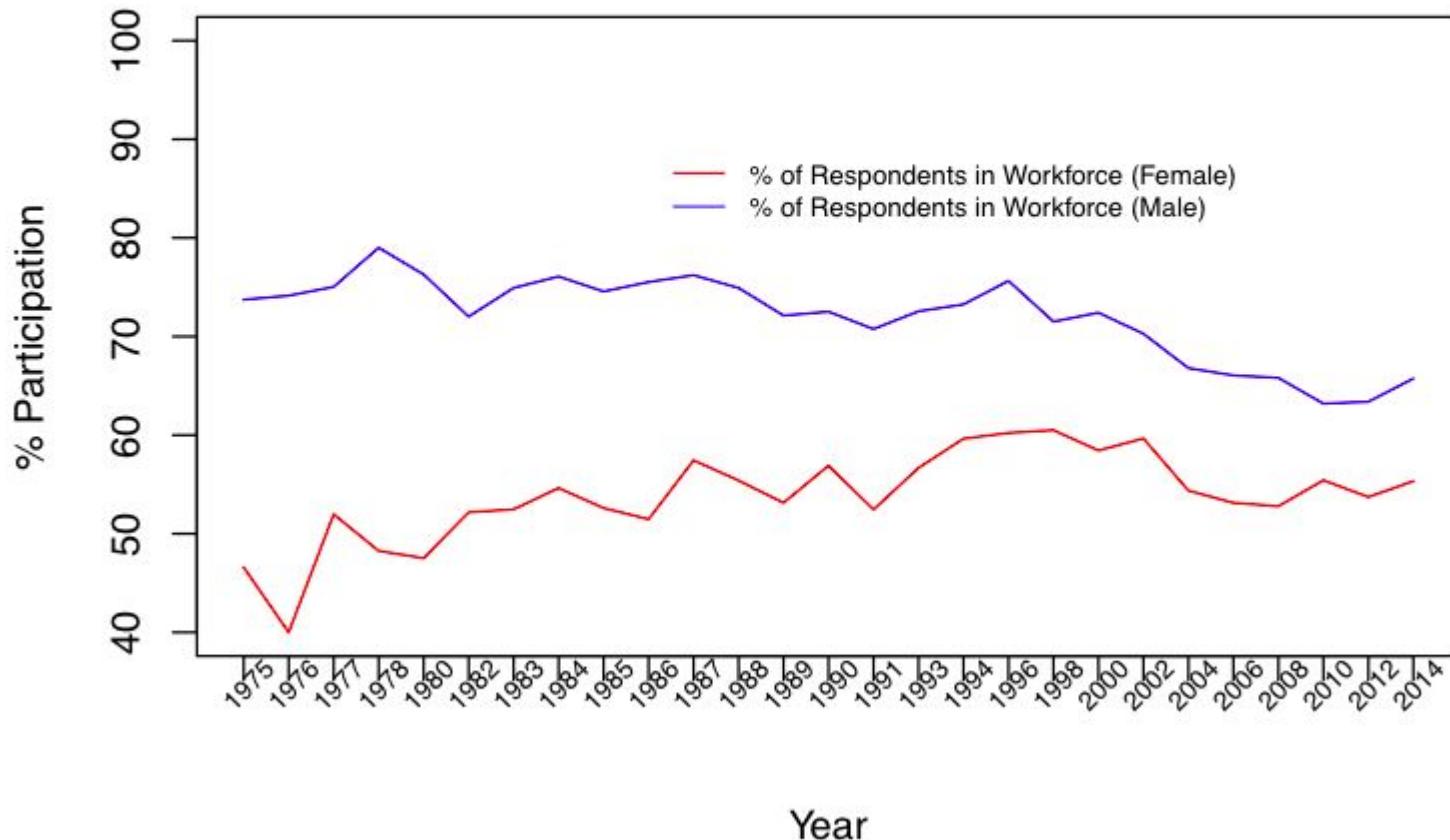
temporal span of examined variables



Describing the Data

the percent of respondents participating in the workforce

Percent Participation in Workforce (1975–2014)

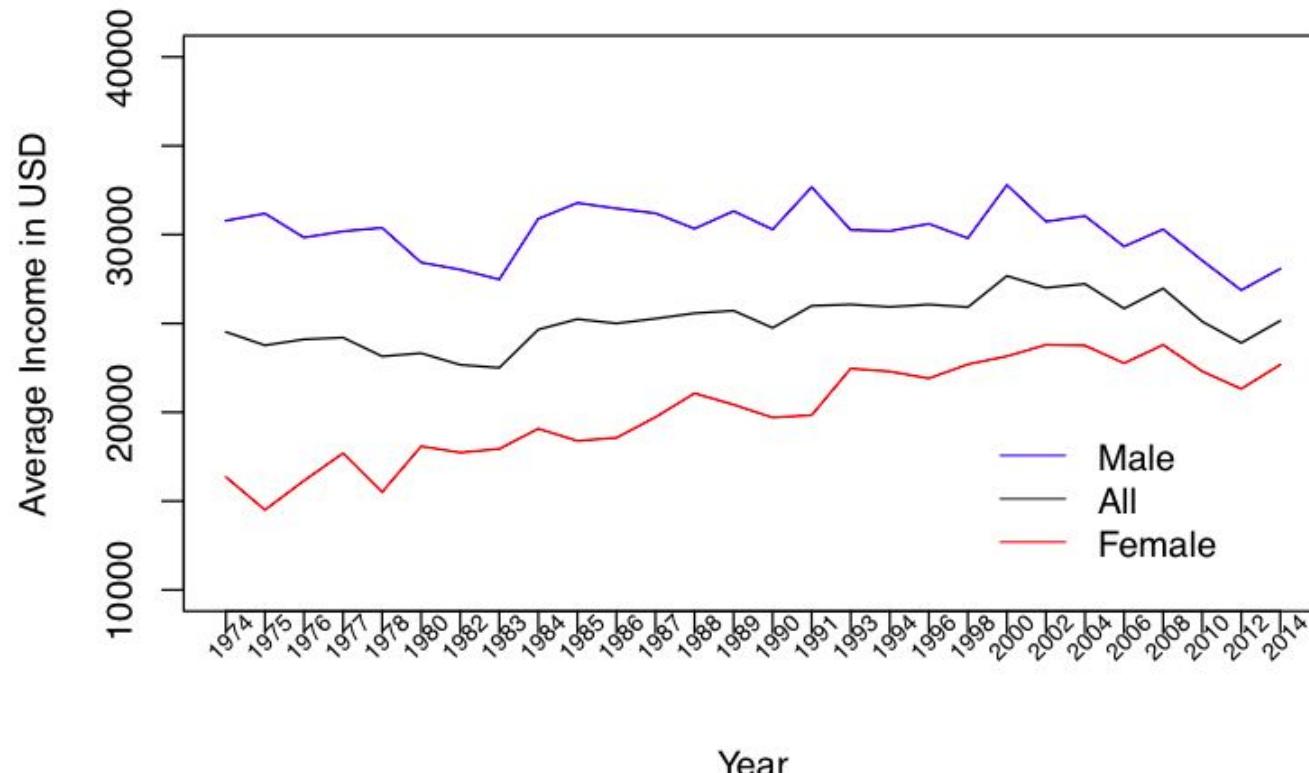


Describing the Data

the overall percent change inflation-adjusted income of women in the workforce for each year

- Of all the respondents of the GSS from 1973 to 2014, for each year we examined those respondents
 - had reported income greater than \$0
 - adjusted for inflation
 - as a USD amount
- Percent change was calculated in regards to income change from previous year

Average Income per Year (1975–2014)



Describing the Data

average job prestige of women-held occupations for each year

- developed by sociologists Robert W. Hodge, Paul S. Siegel, and Peter H. Rossi
- a standardized, subjective aggregate score of occupations based on relative social standing
 - the measures of prestige is a part of the concept of social economic status (SES). Job with high prestige are more likely to have a higher level of pay stability, better lateral career mobility, and established professional associations
 - Prestige is the measurement of the "desirability" of an occupation in terms of socioeconomic rewards
- prestige score ranges from 0 to 97, where 97 is the most prestigious
- ranking of occupations were updated in 1980 and again in 1989; suspended in 2010 for further updates

$$\text{Prestige} = 12.5(\text{Rating} - 1)$$

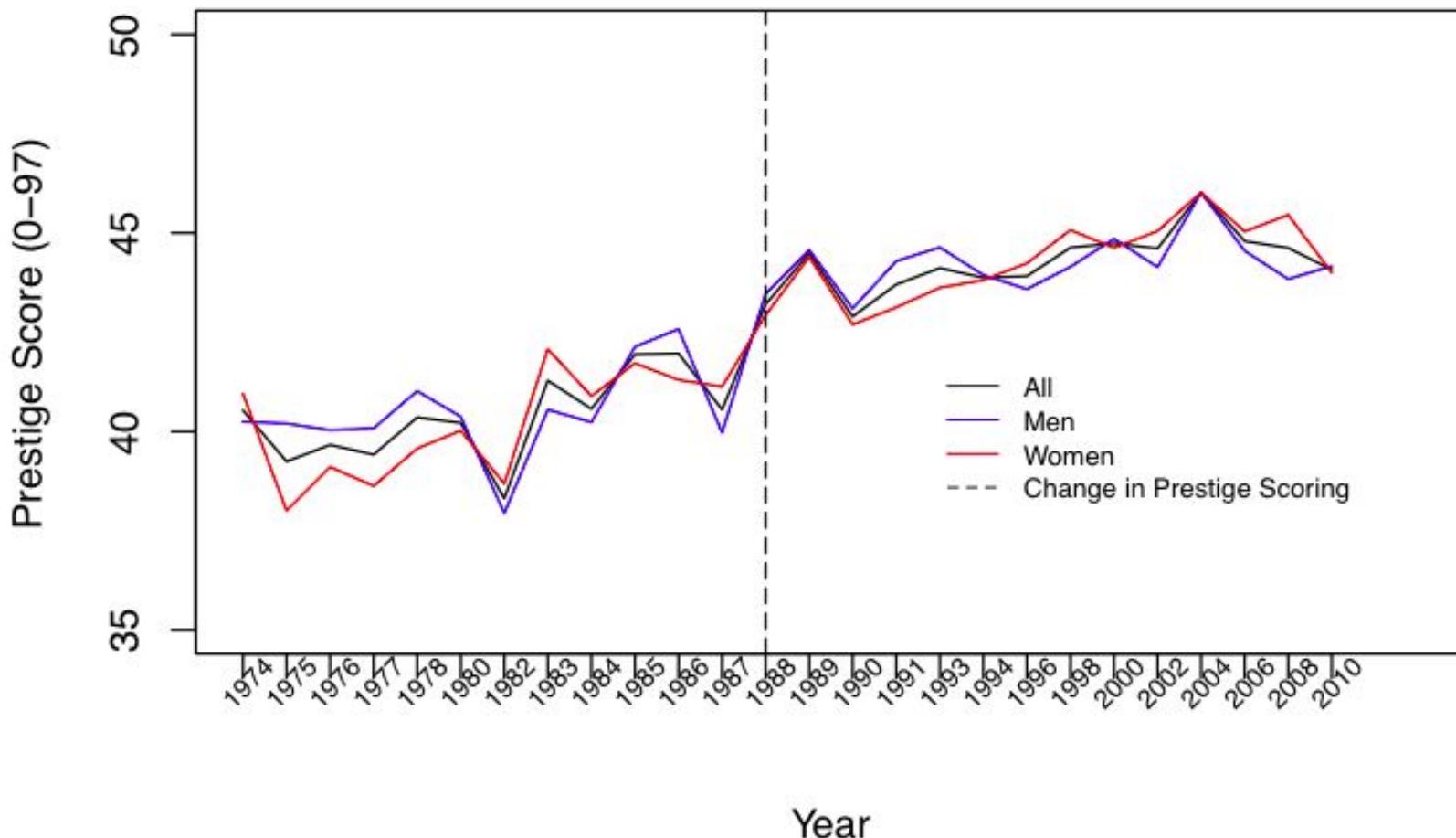
where the rating was provided by sample of individuals selected to rank different occupations

$$\frac{\sum(\text{prestige score of gender for year}, Y)}{\text{number of that gender in the workforce in year}, Y}$$

Describing the Data

average job prestige of women-held occupations for each year

Average Job Prestige of Respondents (1974–2010)



Describing the Data

the percent change in support of women in the labor force from all respondents from the previous year

Approve/Disapprove

“ Do you approve or disapprove of a married woman earning money in business or industry if she has a husband capable of supporting her? ”
(GSS 1972-1998)

$$\% \text{ support} = \frac{\Sigma \text{Approve}}{\Sigma \text{Approve} + \text{Disapprove}} * 100$$

Strongly Agree, Agree/Disagree, Strongly Disagree

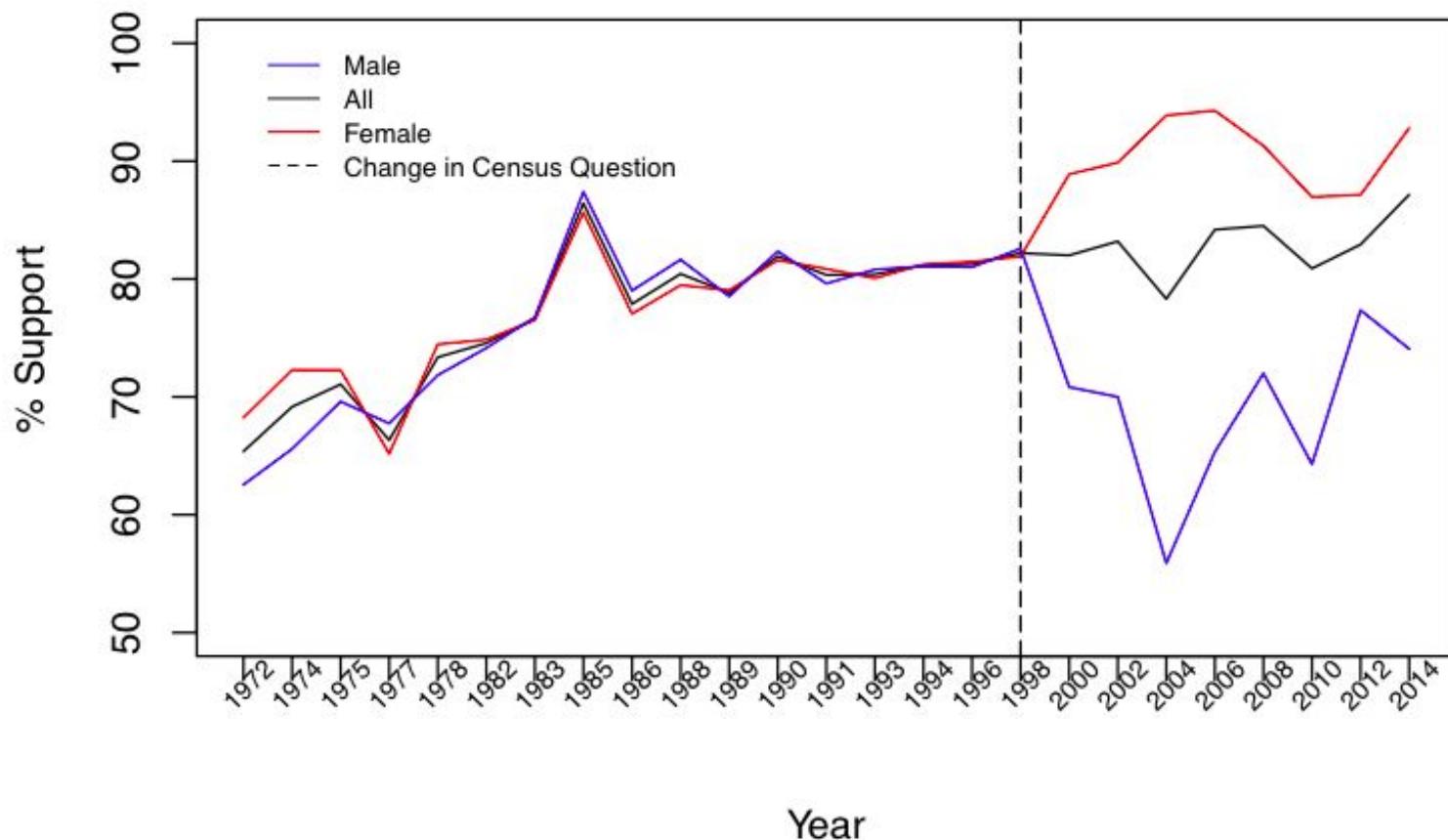
“ Because of past discrimination, employers should make special efforts to hire and promote qualified women. ”
(GSS 2000-2014)

$$\% \text{ support} = \frac{\Sigma (\text{StronglyAgree} | \text{Agree})}{\Sigma (\text{StronglyAgree} | \text{Agree}) + (\text{StronglyDisagree} | \text{Disagree})} * 100$$

Describing the Data

the percent change in support of women in the labor force from all respondents from the previous year

Percent Support of Women as Economic Agents (1972–2014)



Describing the Data

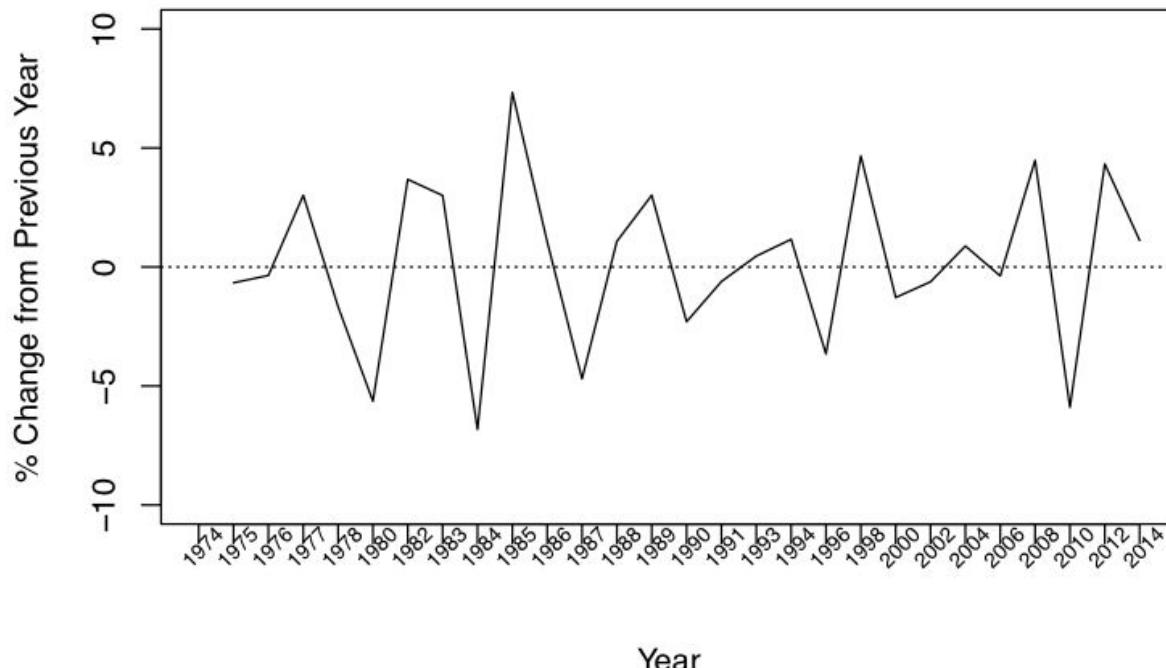
the percent change of job satisfaction for respondents in the workforce

Satisfied, Moderately Satisfied/Moderately Dissatisfied, Dissatisfied

“ On the whole, how satisfied are you with the work you do? ”
(GSS 1970-1980)

$$\% \text{ satisfied} = \frac{\sum (\text{Very Satisfied} | \text{Moderately Satisfied})}{\sum (\text{Very Satisfied} | \text{Moderately Satisfied}) + (\text{Very Dissatisfied} | \text{Moderately Dissatisfied})}$$

Percent Change in Female Job Satisfaction (1975–2014)



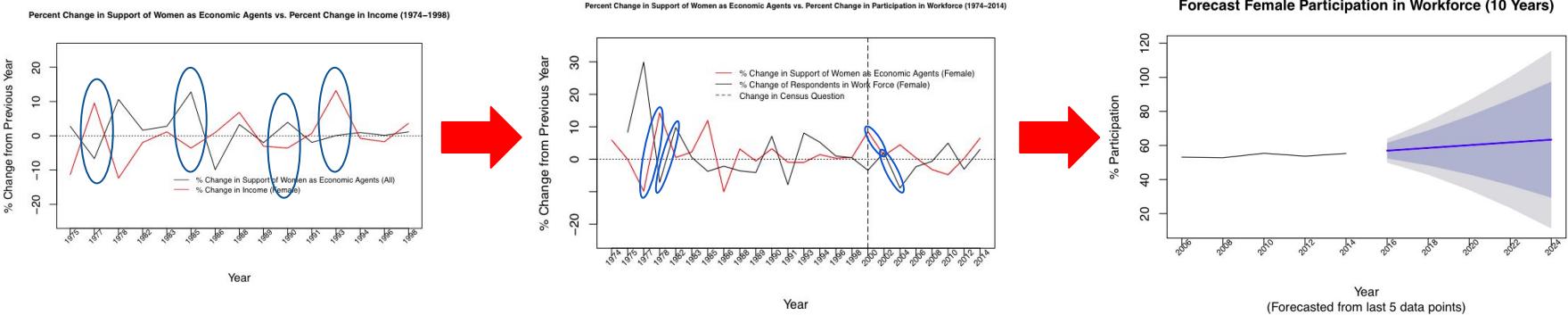
Assumptions Embedded

what did we assume within the model, and why?

- Missing years' reported GSS answers will be omitted since the few years missing towards the end are sparse, and the years missing in the beginning cannot be accounted for with no known data points prior to 1970.
- We assume that the questions -- whether women should work and whether affirmative action should be taken to hire and promote women at work -- about women in the workforce are comparable since they both address the idea of sentiments about approval of women as economic agents.
- The prestige metric developed by Hodge, Siegel, and Rossi for the GSS is altered in 1989 to reflect contemporary perceptions on occupations and new jobs. The 1960s and 1989 metrics are comparable as they were developed with the same goals in mind. However, in 2010 the prestige metric was suspended for re-evaluation.
- We grouped ordinally-ranked variables into buckets.
 - e.g. Satisfied and Moderately Satisfied -> Satisfied
- Outliers for income and prestige were removed so as not to influence the metrics for those variables.
- We converted non-numerical data into percentages to be able to compare across variables
 - We did percentage of Approval as a ratio of those who approved + disapproved, not the general population since most answers were non-answers.

Our Model of Analysis

in the analyses, what questions did we pose regarding the variables and what methods did we use to address them?



1. Is there any relationship between any of the variables?
 - Correlational analysis

2. Does one variable necessarily cause changes in another? Is there a causal relationship between variables over multiple, lagged time frames?
 - Granger causation

3. And looking ahead, what will the variables look like?
 - ARIMA for forecasting

Correlational Outcome

	% change in W in workforce	% change in W income	% change in support of W as economic agents	% change in W job satisfaction	% change in W job prestige
% change in W income	not significant		<i>p < .05 overall and W group</i>	not significant	not significant
% change in support of W as economic agents	<i>p < .05 for W support only</i>			not significant	<i>p < .10</i>
% change in W job satisfaction	not significant				
% change in W job prestige	<i>p < .05</i>				

Significant Correlational Outcomes

Variables (X,Y)	Correlation Coefficient	p-value (in order of decreasing significance)
Workforce Part., Support for Women	R = -0.4951612	p = 0.01629
Workforce Part., Job Prestige (W)	R = -0.4458283	p = 0.02551
Support for Women, Income (W)	R = 0.429088	p = 0.04629
Support for Women, Job Prestige (W)	R = 0.3988611	p = 0.07328

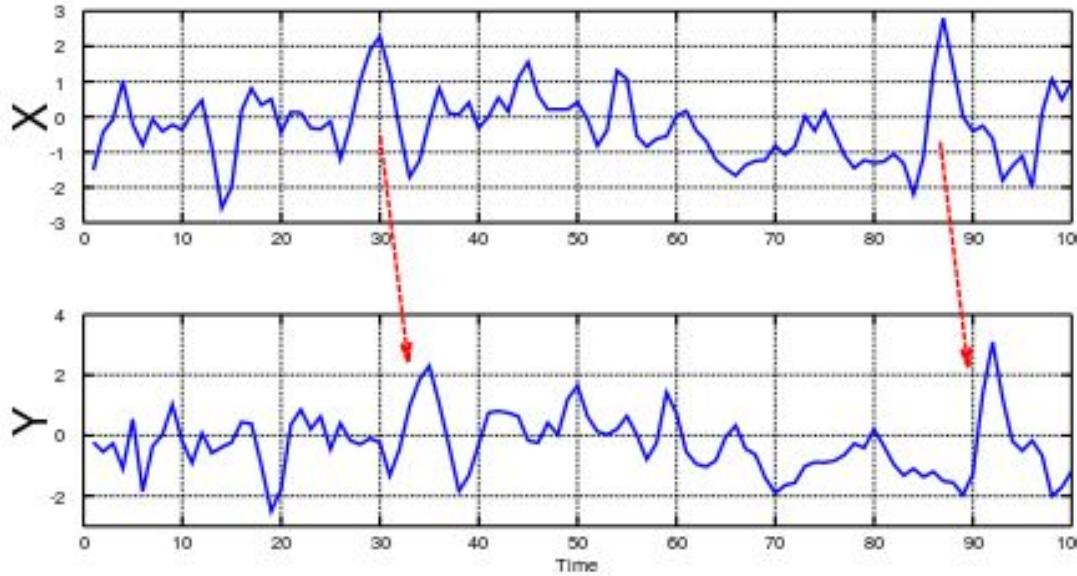
(W) indicates subset of data consisting of women in the workforce

Support for Women = Overall Public Support for Women as Economic Agents

Workforce Part. = Participation in Workforce (W)

Granger Causality

explaining the method



A statistical test to determine whether one time series forecasts another time series:

$$Y_t = \alpha + \delta t + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p}$$

$$+ \beta_1 X_{t-1} + \dots + \beta_q X_{t-q} + e_t.$$

where X and Y are two time series, p and q are lag variables, and beta is the measure of causation of X on Y

Granger Causation Outcome

	% change in W in workforce	% change in W income	% change in support of W as economic agents	% change in W job satisfaction	% change in W job prestige
% change in W income	<i>working % -> income</i>		<i>income -> overall support, income -> W support</i>	not significant	not significant
% change in support of W as economic agents	<i>overall support -> working %, W support -> working %</i>			<i>job sat -> overall support, job sat -> W support</i>	<i>overall support -> prestige, W support -> prestige</i>
% change in W job satisfaction	not significant				
% change in W job prestige	not significant				

Significant Granger Causations

Granger Causality	F-value	p-value (in order of decreasing significance)
Workforce Part. (W) → Income (W)	F = 11.718	p = 0.002555
Income (W) → Support for Women	F = 9.2287	p = 0.006497
Support for Women → Job Prestige (W)	F = 6.9516	p = 0.01582
Job Satisfaction (W) → Support for Women	F = 4.4501	p = 0.0477
Support for Women → Workforce Part. (W)	F = 3.2309	p = 0.08738

→ indicates Granger causality

ex. Time Series X → (Granger causes) Time Series Y

(W) indicates subset of data consisting of women in the workforce

Support for Women = Overall Public Support for Women as Economic Agents

Workforce Part. = Participation in Workforce (W)

ARIMA Forecasting

	2016	2018	2020	2022	2024
Participation in Workforce (W)	56.9%	58.6%	60.2%	61.8%	63.4%
Participation in Workforce (M)	68.1%	70.4%	72.8%	75.1%	77.4%
Percent Change in Average Income (W)	17.2%	28.0%	38.8%	49.6%	60.4%
Percent Change in Average Income (M)	14.8%	25.2%	35.5%	45.9%	56.2%

percent values have been rounded to 1 decimal place

Evaluating Our Model

p-value thresholds	Correlation (X,Y)	Causation X → Y
p < .01		<ul style="list-style-type: none"> • Workforce Part. (W) → Income (W) • Income (W) → Support for Women
p < .05	<ul style="list-style-type: none"> • Workforce Part., Support for Women • Workforce Part., Job Prestige (W) • Support for Women, Income (W) 	<ul style="list-style-type: none"> • Support for Women → Job Prestige (W) • Job Satisfaction (W) → Support for Women
p < .10	<ul style="list-style-type: none"> • Support for Women, Job Prestige (W) 	<ul style="list-style-type: none"> • Support for Women → Workforce Part. (W)

→ indicates Granger causality

ex. Time Series X → (Granger causes) Time Series Y

(W) indicates subset of data consisting of women in the workforce

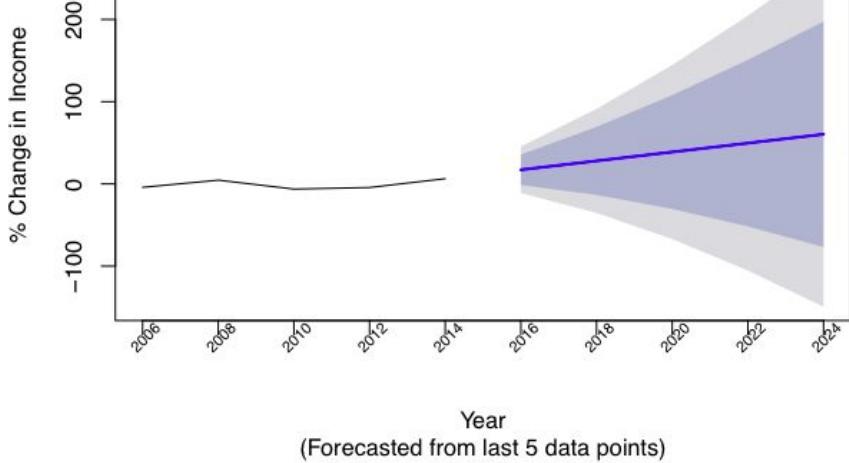
Support for Women = Overall Public Support for Women as Economic Agents

Workforce Part. = Participation in Workforce (W)

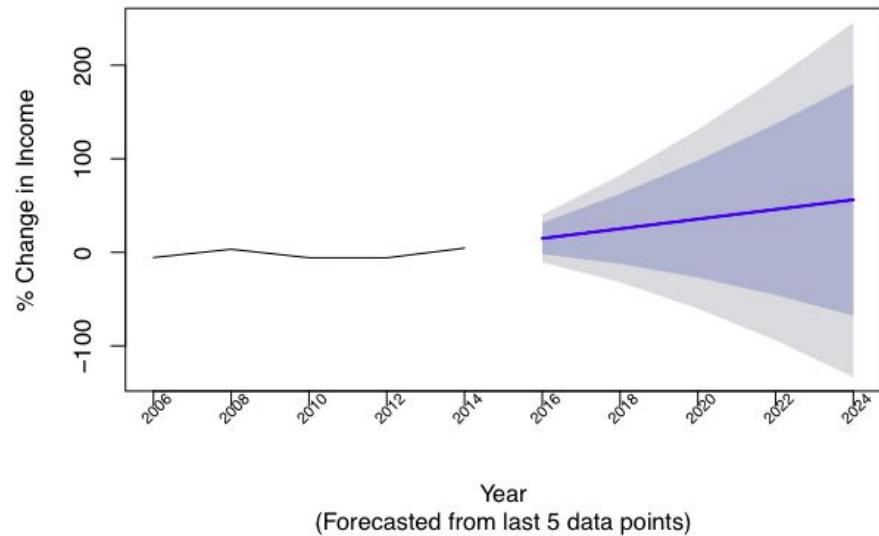
Forecasting

1) *the overall percent change inflation-adjusted income*

Forecast Percent Change in Female Income (10 Years)



Forecast Percent Change in Male Income (10 Years)



> `income.forecast`

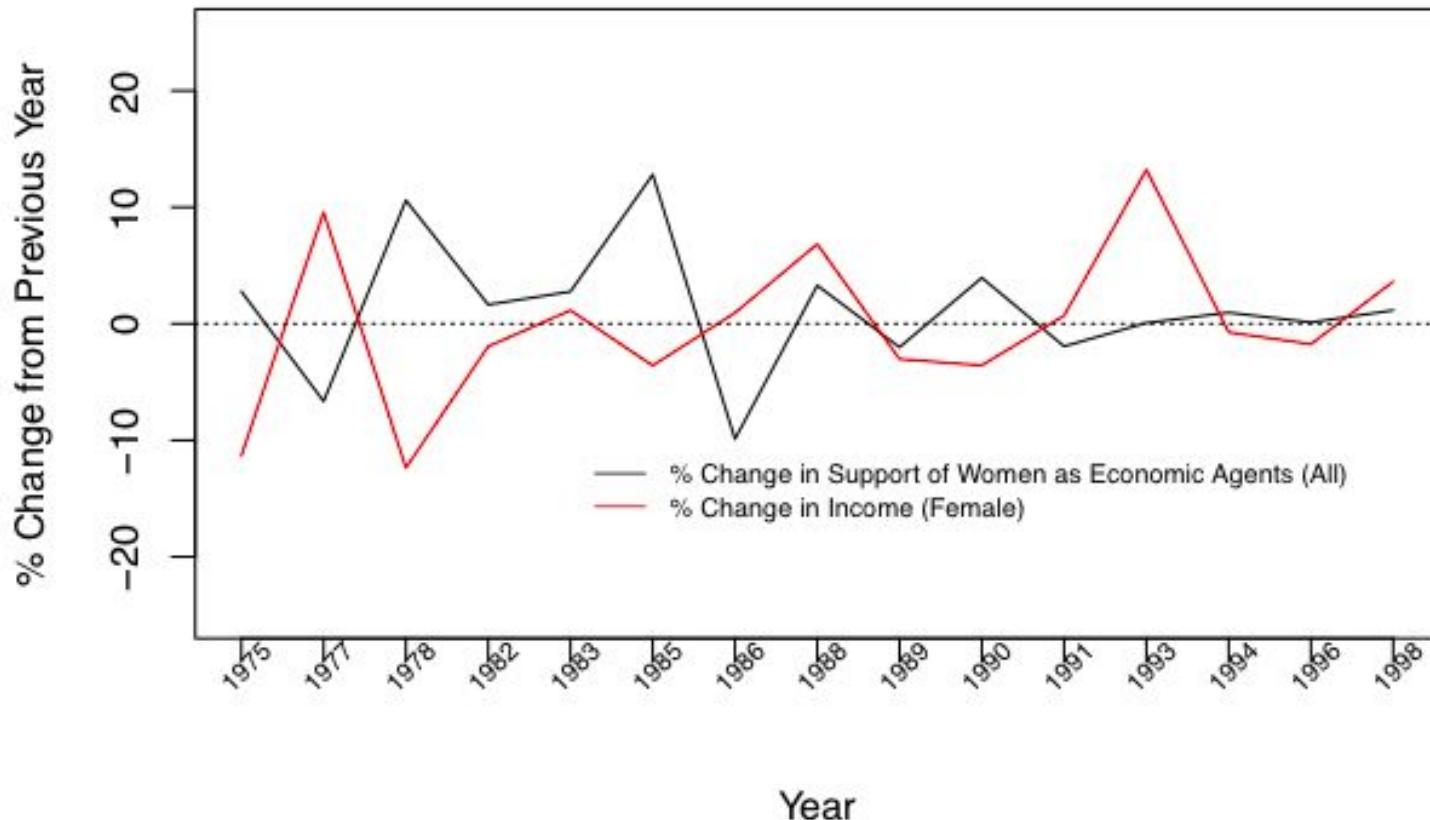
	2016	2018	2020	2022	2024
Percent Change in Average Income, Female	17.17627	27.97357	38.77087	49.56817	60.36546
Percent Change in Average Income, Male	14.83477	25.17862	35.52248	45.86633	56.21018

There is a statistically significant negative correlation between the changes in women's Income and overall support of Women as Economic Agents

correlation = -0.4836227

p-value = 0.06778

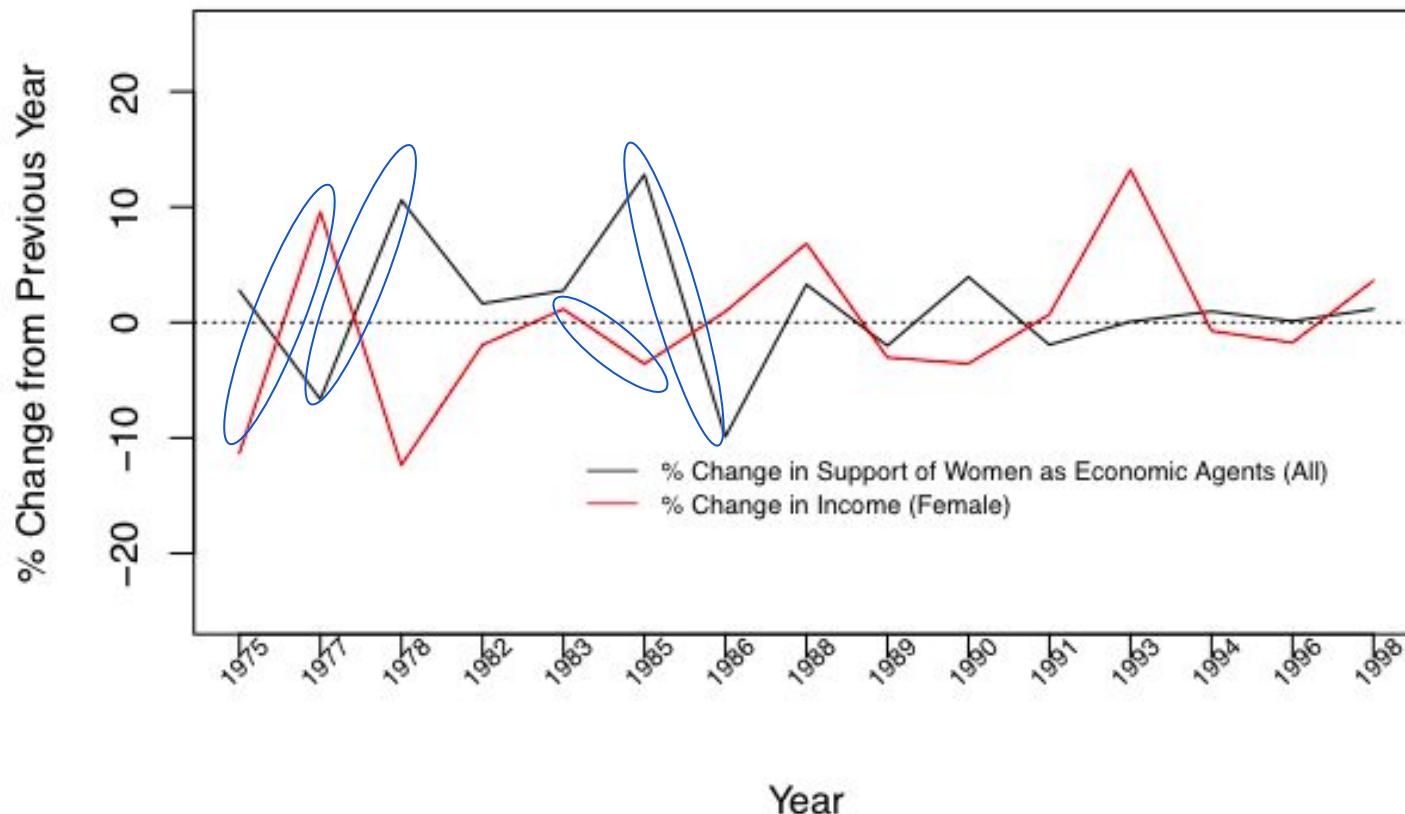
Percent Change in Support of Women as Economic Agents vs. Percent Change in Income (1974–1998)



There is statistical significance when modeling the changes in women's Income as a causal variable of overall support of Women as Economic Agents

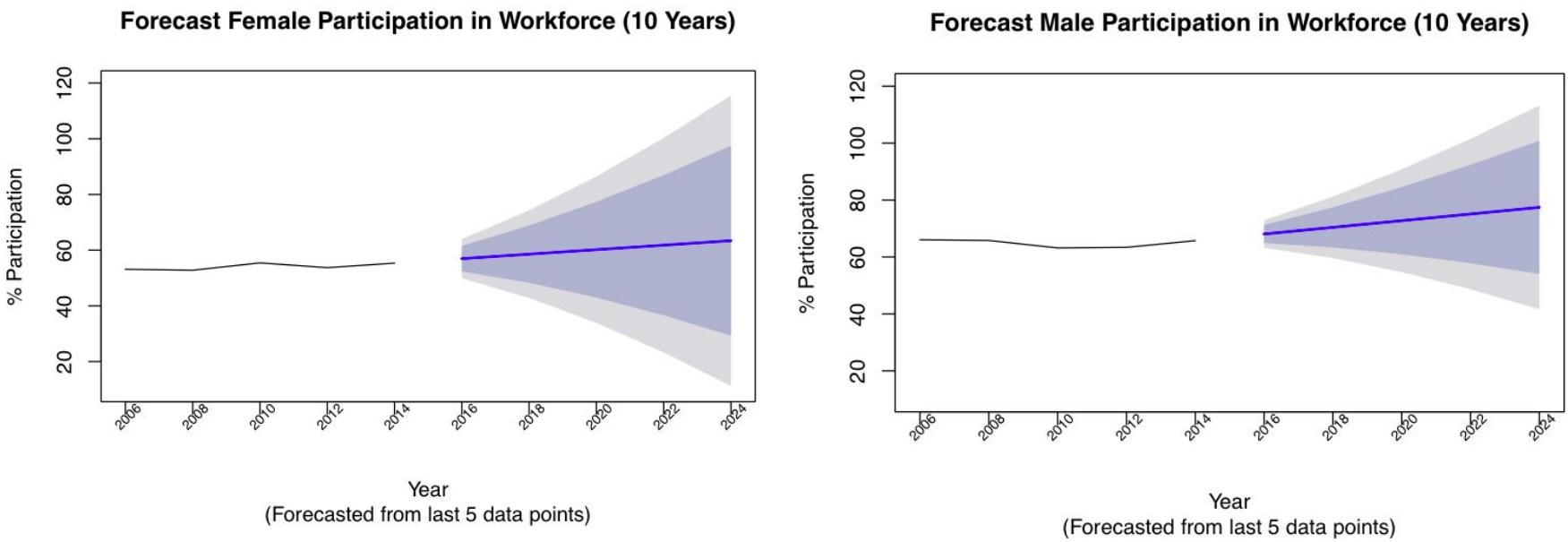
GRANGER CAUSALITY
F-Value = 9.2287
 $p = 0.006497$

Percent Change in Support of Women as Economic Agents vs. Percent Change in Income (1974–1998)



Forecasting

average job participation



> participation.forecast

	2016	2018	2020	2022	2024
Percent Participation in Workforce, Female	56.9467	58.56055	60.17440	61.78825	63.40210
Percent Participation in Workforce, Male	68.0738	70.41578	72.75777	75.09975	77.44173

Evaluation

how did we evaluate the results for each method?

- Pearson's Product-Moment Correlation
 - P-Values
 - Significance Testing
- Granger Causality
 - ADF Test for stationarity
 - F-Values
 - P-Values
- ARIMA
 - AIC



Unanticipated Challenges

what were the obstacles and how were they resolved?



- Originally we intended to focus on women in the field of Computer Science, however we found that subsetting the data in that way produced incredibly small sample sizes
 - the average sample size was 6.6 per year; we had to expand our scope to be all women in the survey that had an income > 0 to be considered included in the workforce
- Using the question change in support for women in the workforce, we found that the two questions, although addressing the same idea, were not that comparable
 - our solution was to analyze the time series for the years that asked the first question since there were more years of data

Confounders

what are some additional things to consider?



- Using the data from the GSS was not as comprehensive as we'd hoped
 - not all questions/variables we used were asked every year
 - the years for the time series is not consistently spaced out
- Causation metric for income and percent change of women participation were both significant
 - but income had a more significant p-value than participation (0.002 vs 0.04)
- A lot of respondents did not respond to the affirmative action question
- Forecasted only on recent data points -- only gave forecast based on that
 - mirrors the latest trends only

Confounders

what are some additional things to consider?



- We did not include all possible census questions that may be related to women's circumstances in the workforce (i.e. race, age, education levels)
 - this may affect our Granger Causality analysis since it assumes all relevant variables are included to make a conclusion about causal relationships
 - Granger Causality also does not account for instantaneous or non-linear causality
 - we only looked at women as one category; we did not segment this population

What is completed now, and what is left to do?

- refine Granger visualizations for all causal relationships
- consider forecasting with greater range of years
- incorporate feedback on existing analyses, visualizations