1 Literature

2 Ideas put forward by the literature

Explanations for the gender gap

- Gender-biased technological change (Black and Spitz-Oener, 2010).
- Differences in commuting patterns and transportation (Black et al., 2014; Liu and Su, 2020).
- Differences on job flexibility (Goldin, 2014).
- Structural change (Olivetti and Petrongolo, 2014, 2016)
- Changes in selection patterns (Mulligan and Rubinstein, 2008)

3 Data notes

3.1 Residualization procedure

Throughout this document I explore the relationship between average wages (gender wage gap) and CZ population density. I am also interested in how this relationship has changed over time.

To discard the possibility that the observed relationships arise because of simple changes in the population composition, I compute average wages net of individuals' characteristics.

Procedure

1. I regress log of the individuals' wages on individual characteristics separately by year, including a gender $\times$ CZ fixed effect. Models are estimated separately for each census year.

$$w_{igrt} = X_{irt}\gamma_t + \lambda^g_{rt}$$

2. $\hat{\lambda}_{grt}$ is my estimate of CZ wages, net of individual characteristics.
3. I compute the CZ-specific gender gap as:

$$gap_{rt} = \hat{\lambda}^m_{rt} - \hat{\lambda}^f_{rt}$$

Throughout the document, the control variables X_{irt} come into for different sets:

- **Basic controls:** age dummies, race dummies, and foreign born dummy.
- **Human capital:** basic controls + 4-level education dummies.
- **Industry controls:** human capital controls + $\approx$ 150 industry fixed effects.
- **Full controls:** industry controls + occupation fixed effects.

4 Basic findings

4.1 Densest CZ have experienced a largest decline in the gender wage gap

Consider the regression

$$gap_{r2020} - gap_{r1970} = \alpha + \beta \ln(density)_{r1970}$$

where $gap_r = w_r^{male} - w_r^{female}$, where w stands for average log-wage in the commuting zone.

Table 1 shows the estimates of β for different average wage measures. All the estimates are significant and negative $\implies$ densest CZ experienced a significantly larger reductions in the gender wage gap.

Several observations

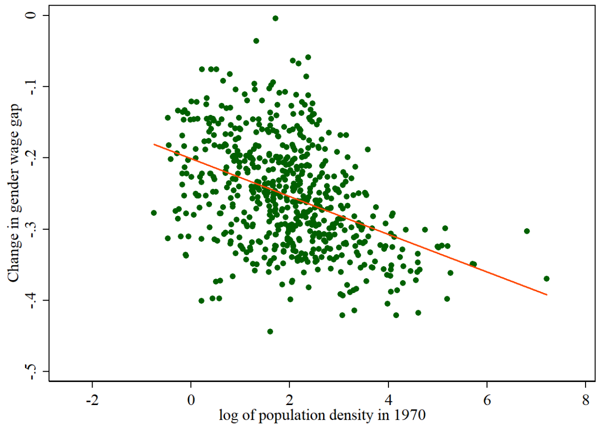
- **Density measure:** Duranton and Puga (2020) argue that total population is more linked to the "experienced" density in a labor market, than the "naive" measure of population density. Table 1 shows that the gradient is virtually unchanged when using: "naive" CZ population density, or total CZ population.
- **Magnitude of the coefficient:** The gradient is large under multiple benchmarks. Let's focus on interpreting the gradient for the raw wage gap on population density.
  - **Standardized effect** an increase of 1 s.d. in population density (1.18) $\implies$ a change of -.4 s.d. in the gender wage gap. A model with just just population density accounts for 17% of the variation 1970-2020 gender gap change.
  - **IC gap:** a CZ in the 25 percentile had an expected decline in the gap of 23 log points. In contrast, a CZ in the 75 percentile experienced a 27 log-point decline $\implies$ a decrease that is 17% larger.
  - **IC gap relative to the mean:** the 4 log-point gap is equivalent to 15.6% of the decline in the gender gap for *the average CZ.*

Table 1: Gender wage gap vs density

	Raw gap	**Net of basic controls**	**Net of human capital controls**
l_czone_density_70	-0.026***	-0.027***	-0.022***
	(0.002)	(0.002)	(0.002)
l_czone_pop_70	-0.025***	-0.025***	-0.023***
	(0.002)	(0.002)	(0.002)

Note: changes based on unweighted estimated elasticities. Sample restricted to full-time year-round workers. Table generated on 23 Nov 2020 at 16:25:24. Table generated with do file 2_analysis/code_files/write_regression_coefplots.do

Figure 1: Change in gender wage gap, 1970-2020



Note: figure restricts to CZ with more than 1 people per km^2. Figure generated on 23 Nov 2020 at 16:25:24. Figure generated using the dofile 2_analysis/code_files/write_regression_coefplots.do.

4.2 The relationship between density and gender gap went from positive to negative during 1970-2020

In figure 3 I show estimates of β_t in the regression:

$$gap_{rt} = \alpha_t + \beta_t \ln(density)_{rt}$$

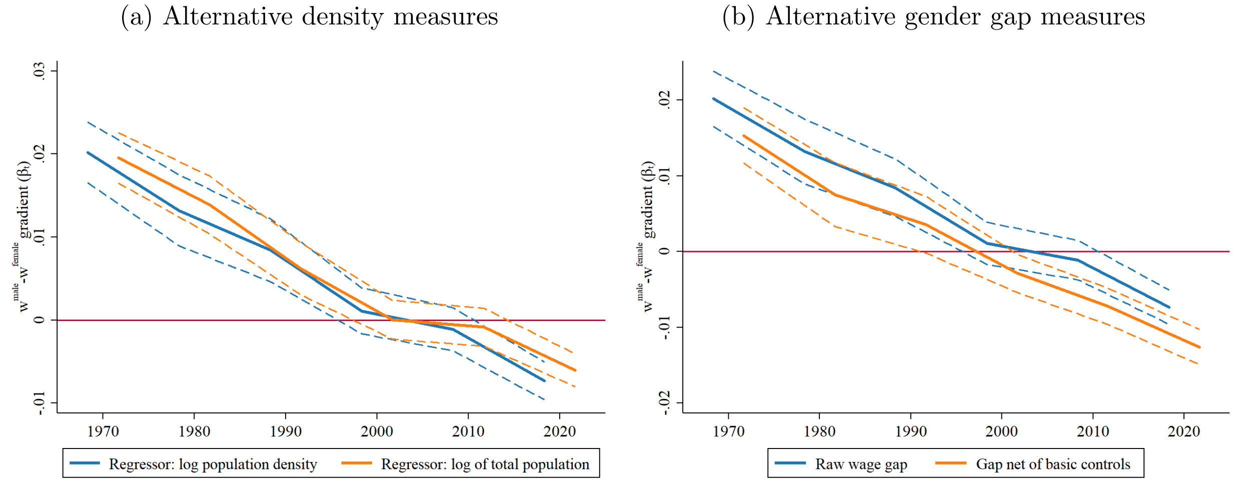
Several remarks

- The density gradient declined gradually over the period. It went from positive in 1970 to negative in the 2020.
- The gradient declines roughly in 2.7 log-points, which roughly corresponds to the coefficient found in in the previous section.
- The inversion of the gradient is robust to controlling for basic individual characteristics.

The cross-sectional gradient in perspective:

- **Implied IC gap**:
  - 1970: 2.8 log-points $\implies$ .4 s.d. increase in gender gap. This is also equivalent to an increase of 7% in the gender wage gap.
  - 2020: 1.16 log-points $\implies$ .3 s.d. increase in gender gap. This is also equivalent to an increase of 6% in the gender gap. If I account for differences in the age distribution of the population, then this difference becomes 2 log-points. Again, this is roughly equivalent to .2 s.d. in the adjusted wage gap.
- **In international perspective:**

Figure 2: Coefficient on population density β_t



Note: figure restricts to CZ with more than 1 people per km^2. Bars show 90% confidence intervals. Standard errors clustered at the CZ level. Figure generated on 23 Nov 2020 at 16:25:25. Figure generated using the dofile 2_analysis/code_files/write_regression_coefplots.do.

- The US is the 6th country with the highest gender gap within the the OECD. Moving from a 10th percentile CZ to a 90th in 2020 $\Longrightarrow$ decline of 3 log-points in the gender gap. This is roughly equivalent to moving from the US to Germany $\approx$ erasing 50% of the gap between the US and the average for OECD countries see this data.
- **Tangential remark on gender sd:** there is as much variation in the gender gap within the US (2020), as there is in the across countries in the OECD.

- **In terms of the urban wage premium?**
  - **In 2020**: the urban wage premium for men was of 4.4 log points. For women was 5.2 log-points $\Longrightarrow$ 18% larger premium for women. Adjusting wages for age / race $\Longrightarrow$ premium for women was 25% larger than for men.
  - **Change across time:** since the peak in 1990, male's premium declined by 3.84 log-points 47% (41% when adjusting by age / race). By comparison, women's premium declined by 2.25 log points. This a decline of 30% (26% when adjusted by age / race) $\Longrightarrow$ this is a big difference in the evolution of men's and women's fortunes. Women's decline in the premium is just 58% of men's decline in the premium.

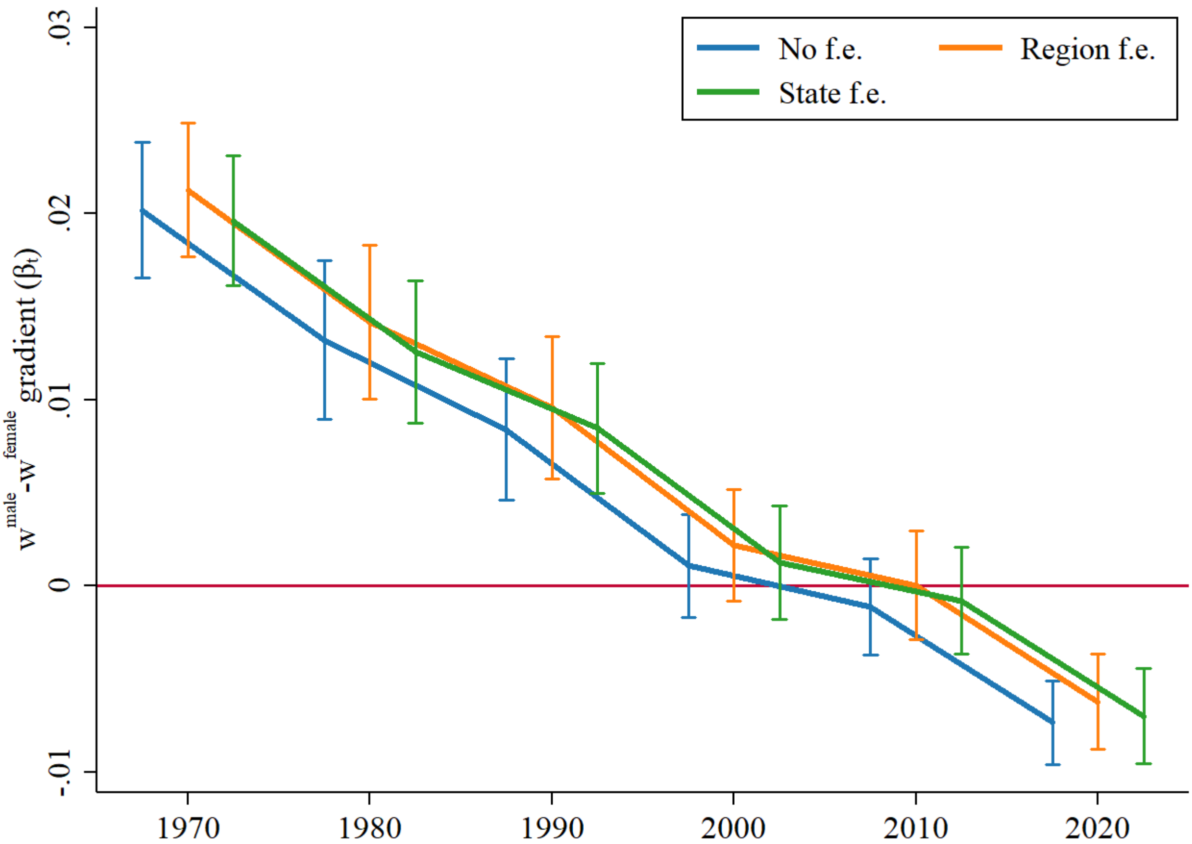
My conclusion here, is that the observed coefficients on density are big relative to the overall within-country variation in the gender gap, even though the IC seems modest.

4.3 Further checks

4.3.1 Is the density gradient just capturing cross-region variation?

No. The gradient and it drift is robust to controlling for region and state fixed effects!

Figure 3: Density gradient is robust to adding region and state fixed-effects



Note: figure restricts to CZ with more than 1 people per km^2. Bars show 90% confidence intervals. Standard errors clustered at the CZ level. Figure generated on 23 Nov 2020 at 16:25:26. Figure generated using the dofile 2_analysis/code_files/write_regression_coefplots.do.

4.3.2 The drift in the gradient is not driven just by small / rural CZ

One concern with the results in the previous sections is that they might be driven by very small CZ, which are relatively unpopulated and which have little bearing in the overall US labor market. To study whether this is a concern I also compute gradients for "big" CZ. I define big CZ as those which had a population density of at least 2.5 people per km^2 in 1950. These 175 CZs had a median population of 316k people and accounted for around 74% of the US population in all the years I am considering.

Figure 4 shows that this is not the case. There I compare the gradients for all CZ vs those I get when I limit estimation to big CZ only. Two features stand out:

1. The gradient *also arises for the* the largest CZ.

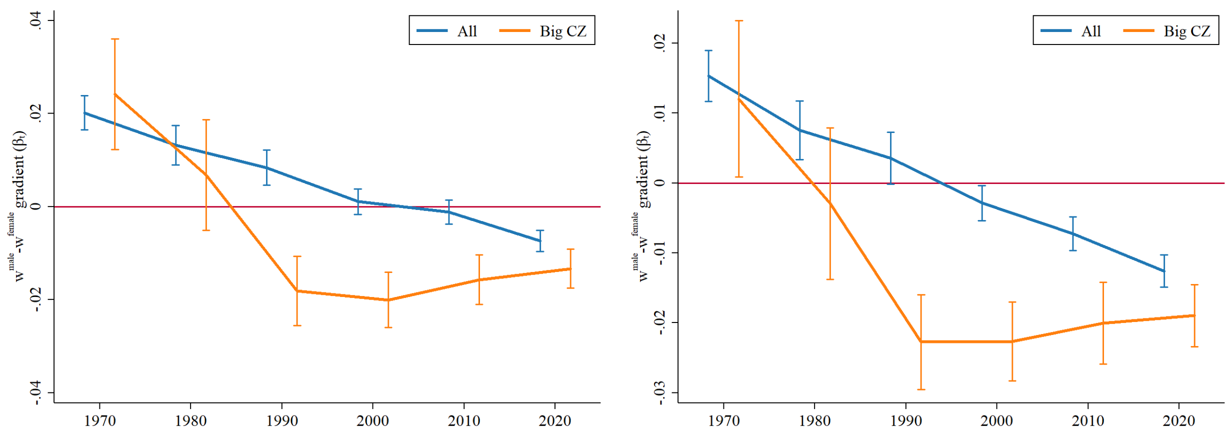
2. Big CZ also feature a decline in the gender-gap gradient. However, this decline is concentrated on the 1970-1990 period. From then on, the gradient remains negative and roughly constant. Panel (b) shows that the picture remains unchanged if I adjust wages for age / race.

Takeway from this exercise:

The gender gap gradient is not driven just by small CZ. Big CZ feature a more striking pattern of gradient decline during 1970-1990. Smaller CZ do drive the gradual decline from 1990 onward.

Figure 4: Density gradient is robust to adding region and state fixed-effects

(a) Raw wages (b) Net of age/race



Note: figure restricts to CZ with more than 1 people per km^2. Bars show 90% confidence intervals. Big CZ are defined as those having a density of at least 2 people per square km in 1950. Standard errors clustered at the CZ level. Figure generated on 23 Nov 2020 at 16:25:27. Figure generated using the dofile 2_analysis/code_files/write_regression_coefplots.do.

5 Questions inspired by these facts?

- Why has the decline in the urban premium has been less severe for women over the period?
- What explains the inversion in the gender gap - density gradient?
- Do women benefit more from moving to a city today?

5.1 Incipient answers

5.1.1 Do women benefit more from cities?

Fixing ideas To fix ideas let's consider the following model for wages, where g indexes gender and r indexes region:

$$w^g_{iert} = \alpha_t + \delta_{et} + \gamma^g_{ert} + \varepsilon_{it}$$

For the sake of the argument, suppose e indexes education. Here δ_e represents the economy-wide return to having education e, while γ represents the region-specific return to education. Then, the average wage in a commuting zone is given by,

$$w^g_{rt} = \alpha_t + \sum_e s^g_{ert}\delta_{et} + \sum_e s_{ert}\gamma_{ert} + u_{rt}$$

To fix ideas, let us impose the strong assumption that $\gamma_{ert} = \omega^g_t \ln(density)_{rt}$ the above equation reduces to:

$$w^g_{rt} = \alpha_t + \sum_e s^g_{ert}\delta_e + \omega^g \ln(density)_{rt} + u_{rt}$$

By running the regression:

$$w^g_{rt} = \alpha + \beta^g_t \ln(density)_{rt}$$

we have that:

$$plim(\beta^g_t) = \omega^g_t + \frac{cov(\sum_e s^g_{ert}\delta_{et}, \ln(density)_{rt})}{var(\ln(density)_{rt})}$$

therefore β^g_t has two components

- ω^g_t = gender-specific urban wage premium.
- A term reflecting bias due to heterogeneous composition of the population across labor markets. For example, if high-wage groups are more likely to locate in cites $\implies \beta^g_t$ overestimates the gains from moving to a city.

I will impose another strong assumption. Let $\frac{cov(\sum_e s_{er}\delta_e, \ln(density))}{var(\ln(density)_r)} = \chi^g$.

Then, abusing a bit of notation, we have

$$\beta^g_t = \omega^g_t + \chi^g_t$$

This simple model suggests a procedure to decompose the density gradient into the selection / urban premium components.

1. β^g_t is obtained from a simple regression of average *raw wages* on population density.

2. Under the assumption of $\gamma_{er} = \omega^g \ln(density)_r$ an individual's wage is reduced to:

$$w^g_{iert} = \alpha_t + \delta_{et} + \gamma^g_{rt} + \varepsilon_i$$

which is just a regression of individual's wages on education fixed-effects, and a gender$\times$CZ fixed effects. In this case γ^g_{rt} gives an estimate of $\omega^g_t \ln(density)_{rt}$.

3. From there, it is straight forward to obtain an estimate of the urban wage premium, just estimate the regression:

$$\gamma^g_{rt} = \eta_t + \omega^g_t ln(density)_{rt}$$

4. The selection component is obtained out of the difference between the estimates in part 1 and 3.

Some remarks:

- In the discussion above I am assuming that the urban wage premium is the same for education groups $\gamma^g_{er} = \omega^g \ln(density)_{rt}$. If the premium were education-specific then,

$$w^g_{rt} = \alpha_t + \sum_e s^g_{ert}\delta_{et} + \sum_e s_{ert}\omega^g_{et} \ln(density)_{ert} + u_{rt}$$

which would introduce heterogeneity of density gradient across commuting zones. One can easily explore whether this is an issue by estimating ω^g by education group (this is feasible only when there is a limited number of education groups).

Results

Premium today: in the table below I show the decomposition described above under different sets of controls.

Interpretation notes:

- From left to right, each column adds a new set of regressors controls. Thus the interpretation of the selection component in, say the human capital column, is how much of the density gradient is accounted by the human capital controls, over an above the basic set of controls. As a result, the % explained column is computed relative to the adjusted premium in the basic column.
- A negative selection component $\Longrightarrow$ denser cities have higher shares of low wage groups.
- A positive component $\Longrightarrow$ denser cities have higher shares of low income groups.

What does stand out?

- Women's premium is always larger than men's.

- Selection component of basic (age, rage, foreign born dummies) is always negative (city's population is blacker and younger, which tend to be associated with a lower income). *Selection on basic demographics* is twice as important for women than for men.
- Selection on education, industry, occupation is very similar for both genders.
- **Overall message:** even after accounting for this rich set of characteristics, women's premium is positive and bigger than men. Cities appear to draw more women from low-wage groups $\implies$ accounting for different demographic composition actually *increases* women's relative density premium.

Premium over time The table below shows the decomposition of the density gradient for three selected years.
What does stand out?

- In 1970, city women were more likely to come from low education groups. This had reversed by 2020.
- 1970-1990: Accounting for industry of employment eliminates the gender gap in the urban premium. Even though both genders are employed in higher pay industries in cities, men are more likely to do so $\implies$ most of density premium can be explained by gender segregation across industries.
- Note however, that industry and occupation do not erase the gap in the urban premium across genders in 2020. Even after accounting for industry and employment, the women's premium is 56% larger than men's.
- The table also makes clear the pattern already noted by [cite autor here]. There's a precipitous decline in men's urban wage premium. If we concentrated in the last column, between 1990 and 2020 men's premium declines by 61%. What is remarkable, is that women's premium is relatively more resilient, declining by 40% over the same period.
- Without adjusting for industry and occupation men's decline in the premium was so strong that it went from being *above* women's, to being below it.

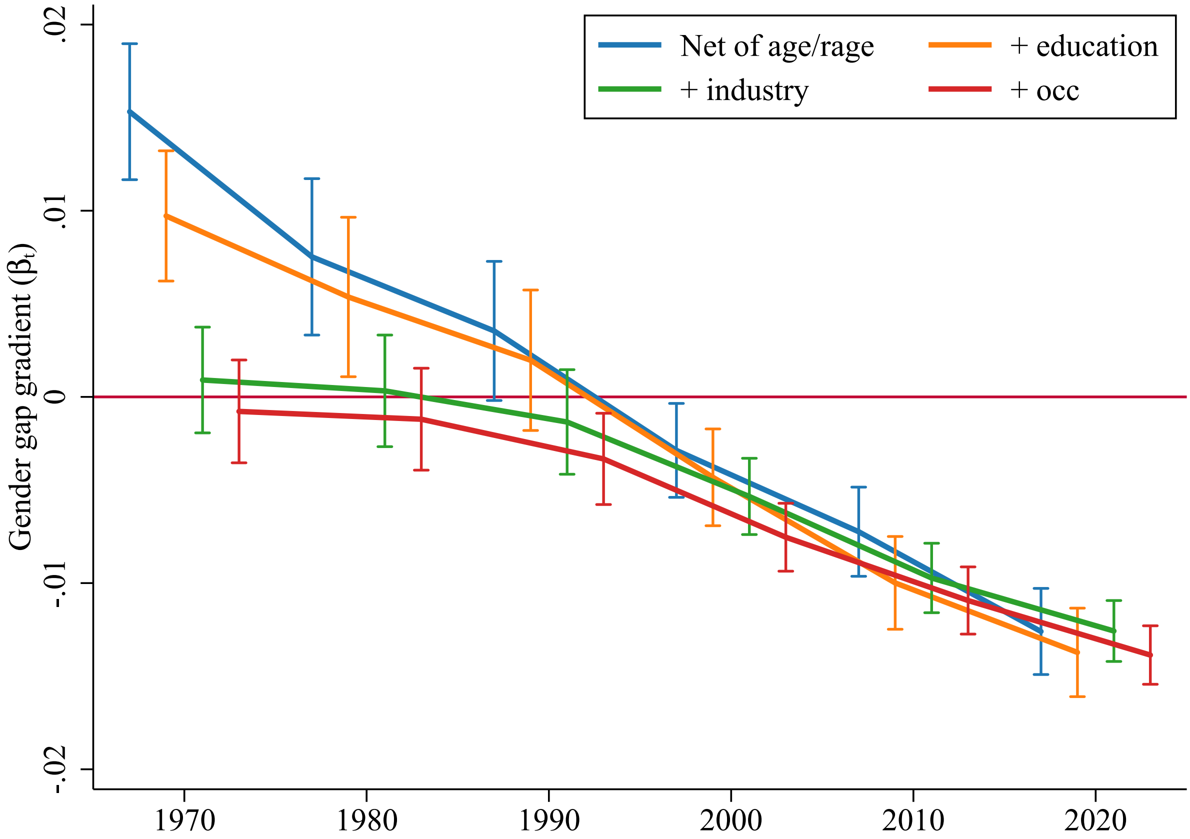
The attached coefficient plots basically paint the same picture. They also suggests that the reversal of fortunes happened during the 1990s. From then on, there is a gradual expansion of the gender gap in the premium.

Some interpretation caveats

- In the discussion above I am interpreting $\hat{\omega}_t^m - \hat{\omega}_t^f$ as evidence of differneces genders in the urban wage premium. This interpretation might not be correct. For example, suppose that urban wage premium differs by education level, but not across genders. That is:

$$\gamma_{ert}^m = \gamma_{ert}^f = \omega_{et} \ln(density)_{er}$$

Figure 5: Coefficient on population density β_t controlling for worker characteristics



Note: figure restricts to CZ with more than 1 people per km^2. The regressions are done on data aggregated at the CZ level. Bars show 90% robust confidence intervals. Standard errors clustered at the CZ level. Figure generated on 23 Nov 2020 at 16:25:28. Figure generated using the dofile 2_analysis/code_files/write_regression_coefplots.do.

note that in this case:

$$w^m_{rt} - w^f_{rt} = \alpha^m_t - \alpha^f_t + \sum_e \delta_e (s^m_{ert} - s^f_{ert}) + \sum_e \omega_{et}(s^m_{ert} - s^f_{ert}) \ln(density)_{rt} + u_{rt}$$

Under this scenario the gradient on density is given by:

$$\sum_e \omega_{et}(s^m_{ert} - s^f_{ert})$$

The expression above will likely be negative if:

– Women concentrate more heavily in high urban wage premium groups.

If we now focus on the drift down of the premium, men's advantage will decline if:

– The premium for men-heavy groups is declining over time.

– Or women are increasingly concentrating in high urban premium groups.

Note that this model also suggests' a quick way to discard this possibility. Note that the within-group gender wage wage is given by:

$$w^m_{ert} - w^f_{ert} = \alpha^m_t - \alpha^f_t + u_{rt}$$

thus there shouldn't be a within-group gender-gap gradient on density. Given the increase in women's educational attainment since the 1970 education is a natural target to apply this test.

The discussion above takes the proposed model too seriously, but it gives an useful intuition. If the drift in the gender-gap premium is driven by differences in group premia + changes in the the composition of the labor force across genders, then these issues should be less prevalent when I compute the regressions by group.

Urban wage premium over time

	Raw	Basic	Human capital	Industry	Industry and occupation
1970					
Men					
Adjusted premium	0.067	0.075	0.077	0.058	0.052
Selection component		-0.008	-0.003	0.020	0.006
		-0.120	-0.035	0.255	0.107
Women					
Adjusted premium	0.047	0.060	0.068	0.057	0.052
Selection component		-0.013	-0.008	0.011	0.005
		-0.277	-0.138	0.161	0.081
Wage gap					
Adjusted premium	0.020	0.015	0.010	0.001	-0.001
Differences in selection		0.005	0.006	0.009	0.002
% explained		0.243	0.366	0.907	1.778
1990					
Men					
Adjusted premium	0.083	0.088	0.084	0.072	0.064
Selection component		-0.005	0.015	0.003	0.007
		-0.063	0.165	0.035	0.093
Women					
Adjusted premium	0.074	0.085	0.082	0.073	0.067
Selection component		-0.011	0.003	0.008	0.006
		-0.144	0.038	0.103	0.083
Wage gap					
Adjusted premium	0.009	0.004	0.002	-0.001	-0.003
Differences in selection		0.005	0.002	0.003	0.002
% explained		0.576	0.472	1.684	-1.615
2020					
Men					
Adjusted premium	0.044	0.049	0.035	0.032	0.025
Selection component		-0.005	0.015	0.003	0.007
% explained		-0.118	0.296	0.083	0.210
Women					
Adjusted premium	0.052	0.062	0.049	0.045	0.039
Selection component		-0.010	0.014	0.004	0.006
% explained		-0.202	0.218	0.082	0.124
Wage gap					
Adjusted premium	-0.007	-0.013	-0.014	-0.013	-0.014
Differences in selection		0.005	0.001	-0.001	0.001
% explained		-0.703	-0.087	0.080	-0.095

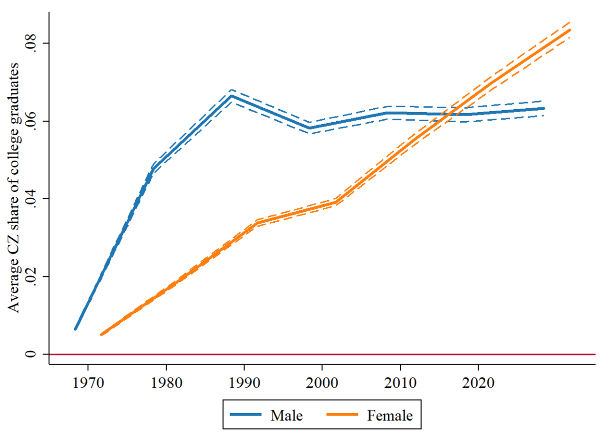
5.2 Are the results above driven different wage premium across educational groups?

Short answer: no. But the drift in the gender-gap gradient is driven by workers without a bachelor degree.

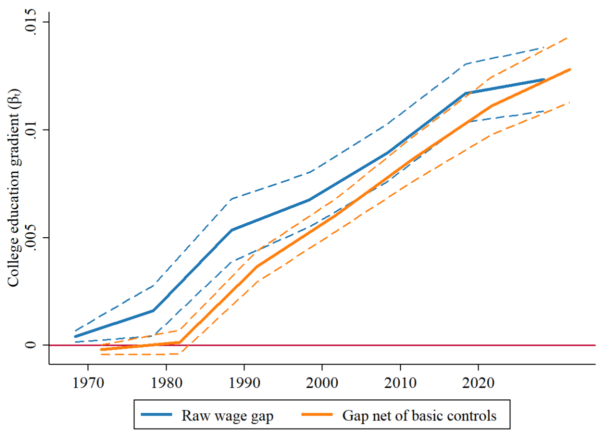
Panel a in figure 6 shows that women educational attainment have increased dramatically relative to men. Panel b shows that it has increased more in the densest labor markets. If the urban wage premium of college graduates is larger than non-college graduates, then this trends could drive the patterns described in previous sections.

Figure 6: Educational attainment by gender

(a) Average share college graduates



(b) College share-density gradient



Note: figure restricts to CZ with more than 1 people per km^2. Bars show 90% confidence intervals. Standard errors clustered at the CZ level. Figure generated on 23 Nov 2020 at 11:41:41. Figure generated using the dofile 2_analysis/code_files/write_regression_coefplots.do.

In panel a of figure 7, I show estimates of β_{et} for the following regression model:

$$w_{rt}^m - w_{rt}^f = \alpha_{et} + \beta_{et} \ln(density)_{rt}$$

If all the variation is driven by changes in education-specific urban wage premium, then we should expect β_{et} to be zero in all years. This is not what the data shows. The gender-gap gradient is positive and mostly constant for college graduates. The decline in the gradient arises for non-college graduates only. Panel b shows that adjusting the gap for age and race does not change this finding (panel b).

Panels c and d show the density premiums by gender. There I plot the density slope in regressions of the form:

$$w_{rt}^g = \alpha_{et}^g + \beta_{et}^g \ln(density)_{rt}$$

- Women with a bachelor degree benefit *less* from cities than men. Moreover, panel (b) shows that there is no decline in the urban wage premium for college workers. This is consistent with the findings by [autor].

- The decline for non college workers happens in both genders. But again, women are less affected by this development $\Longrightarrow$ *the reversal of fortunes is happening for college without a college education.*

Although education is an endogenous outcome, it is remarkable that the urban wage premium for college graduates is so constant for both genders over the whole period.

- Women's labor force participation and educational attainment has increased massively over the years. Moreover, their concentration in urban areas has also increased. If anything, a simple supply-demand framework would have suggested that the gender-gap premium should have decreased.

- This begs the question of whether the urban wage premium is smaller for college women. If so, what prevents them from reaping the same benefits as men from cities?

5.3 1970-1990: a tale of women's accession to high pay industries

Figure 11 suggests that industry of employment accounts for both the (i) cross-sectional gender-gap gradient, and (ii) its decline during 1990-2020. Here I zoom in on the relationship between industrial structure, population density, and employment shares across genders.

5.3.1 Which are the high-pay industries?

6 Descriptive analysis

6.1 Basic facts and trends

- Most CZ are relative small. 40% of CZ account for 85% of the of the total population in almost all years.

- CZ at the very top have lost population share. "Mid"-tier cities are the ones with the fastest population growth.

- Overall, the US has experienced manufacturing decline. But decline has been much faster in the top 30% CZ (see here).

6.2 The gender gap

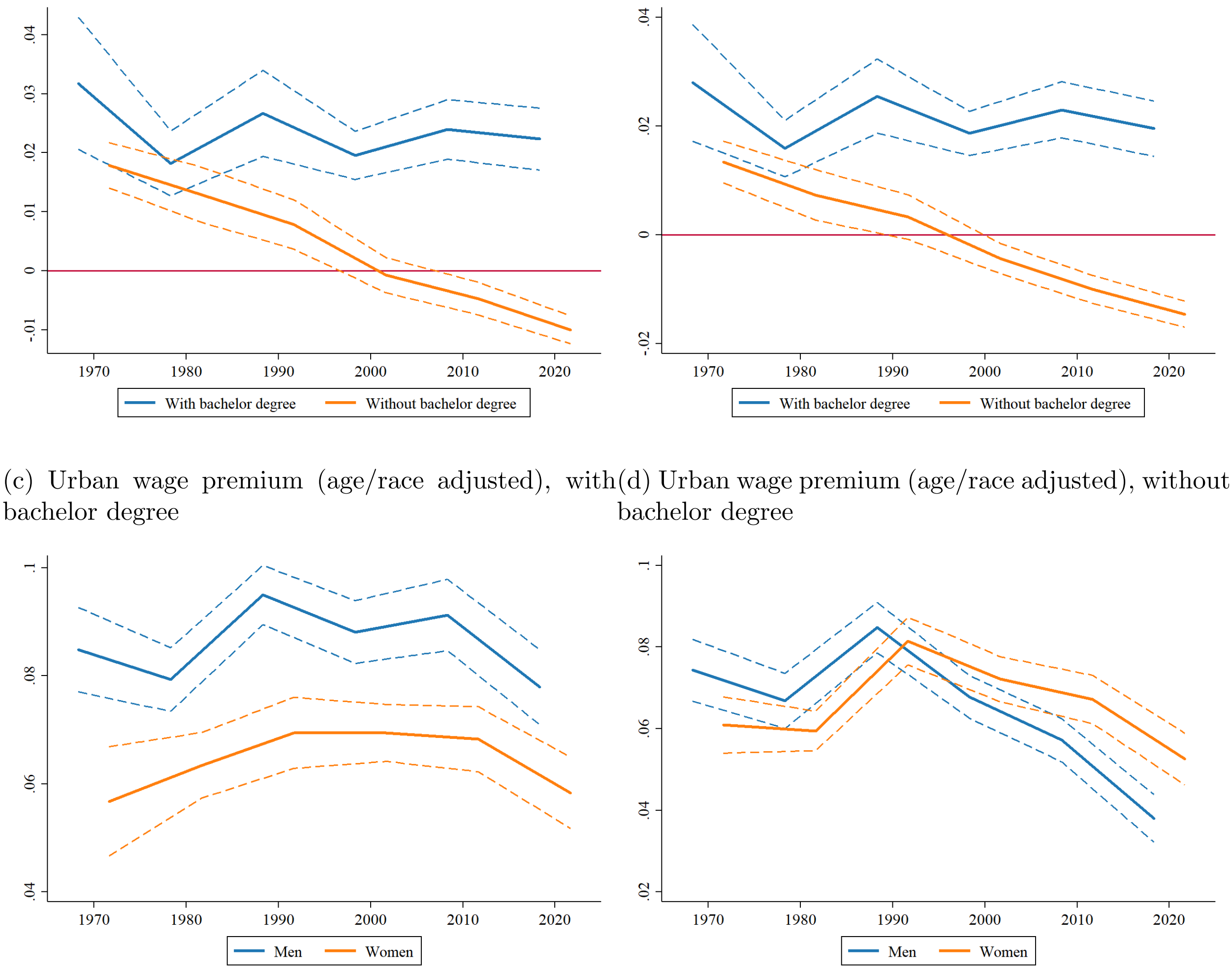
- Wage gap is decreasing everywhere. But decline is faster in densest CZ.

- There is a clear inversion of the gender-gap density gradient.

- This inversion is also present if I focus attention to the top 248 CZ. [see this graph].

-

Figure 7: The density gradient by education level



(a) Gender gap gradient

(b) Gender gap gradient, age and race adjusted

(c) Urban wage premium (age/race adjusted), with bachelor degree

(d) Urban wage premium (age/race adjusted), without bachelor degree

Note: figure restricts to CZ with more than 1 people per km^2. Dashed lines represent 90% confidence intervals. Figure generated on 23 Nov 2020 at 11:41:43. Figure generated using the dofile 2_analysis/code_files/write_regression_coefplots.do.

6.3 LFP

- Employment to population ratio has always had an U pattern for men. The U pattern has exacerbated over the years. Decline in middle places has been faster. This is in line with the Ely Lecture (see here)
- For women, densest places offered more employment at the start. Over time, they have become more like men. U pattern seems to be intriguing (see here)
- What about the absolute values:
  - For men, the story if one of decline in LFP. With faster decline in the middle of the distribution. See here

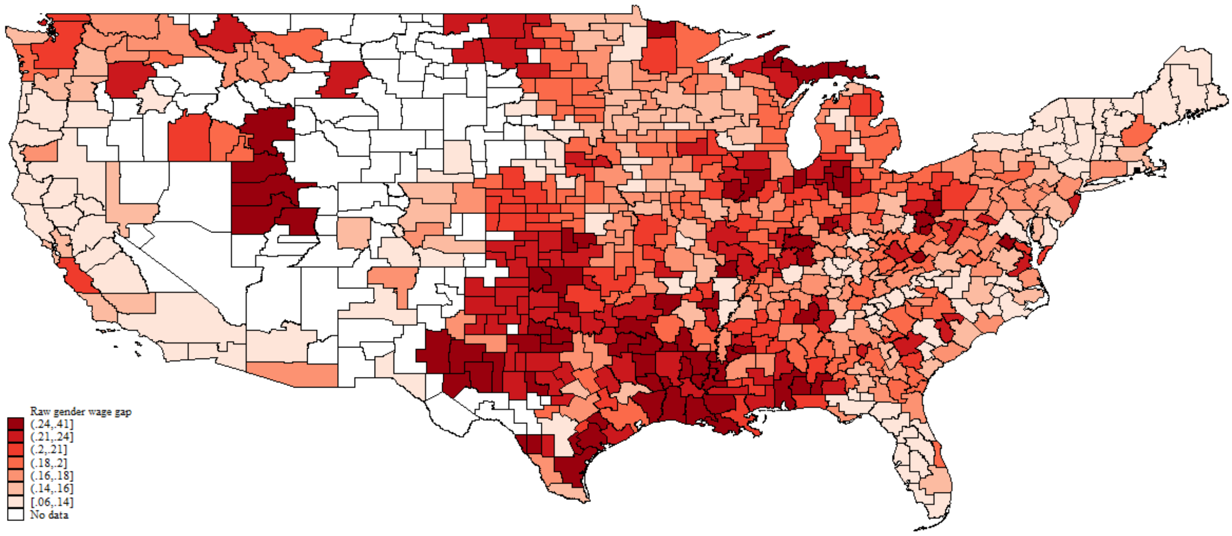
- For women, the story is one of faster progress at the tails of the distribution. here

7 Main findings

Fact 1: there are substantial differences in the level of the gender gap across CZ Figure 8 illustrates variation of the gender gap across US CZ in 2020. The figure restricts to CZ with population densities above 1 person per square kilometer in 1950. In 2020, men had an unconditional average wage 19 (se 5) log-points larger than women's. The map however, shows that there are wide variations from this average across CZ. Men's wage advantage is below 14 log-points in the Northeast and most of the West Coast, while it is above 24 log-points in the parts of the South West. The standard deviation in gender-wage gap across the 625 CZ shaded in the graph is of 5 log-points, which represents 26% of the average national gap.

[make an argument here that these differences are economically significant]

Figure 8: The gender gap in the US in 2020



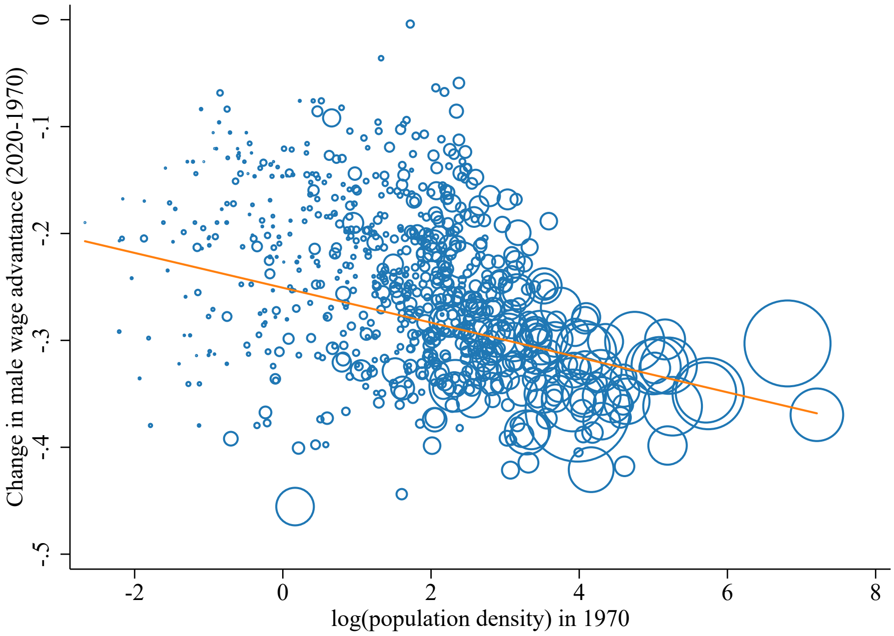
Note: darker colors denote higher relative wages for men. Figure restricts to czones with population densities above 1 person per km^2 and full-time year-round workers.

Fact 2: there are substantial differences in the evolution of the gender gap across CZ

- Take two CZ as an example and show the evolution of the gender gap in these two places.
- Then show statistics on the change of the gender gap across places.

Fact 3: the gender gap has decreased the most in the densest CZ [add residualization at the individual level]

Figure 9: Change in male wage advantage in US CZ



Fact 4: the relationship between population density and the gender gap has inverted over the period [write regression I am writing here]

[graph of cross-sectional slope goes here]

8 Robustness of facts 3 and 4

8.1 Composition of the sample

- Results are robust to including all male and female workers.
- Results are also results for controlling for basic demographics

8.2 Weighting of regressions

Both facts are robust to alternative weighting mechanisms. Weighting only changes the timing of the decline in the population density gradient.

- This happens because decline of the gap happens first in denser places and then it decelerates.
- Places at medium levels of density speed the decline from 1990 on. This would explain the difference between between weighted and unweighted estimates.

8.3 Alternative measures of density

- Results are robust to using population as a measure of density.

9 Possible explanations

9.1 Causes for the gender gap in the literature

9.2 Increased sorting

- Regressions above do not control for observable characteristics.
- If women with stronger ability sort themselves:
  - Increasingly in denser cities.
  - This sorting is stronger than men.
- This would generate faster decrease of the gender gap in denser places.

Suppose wages are determine as follows:

$$y_{igr} = X_{igr}\gamma + \varepsilon_{igr}$$

taking averages by gender at the CZ level we have:

$$\bar{y}_{gr} = \bar{X}_{gr}\gamma + \bar{\varepsilon}_{gr}$$

therefore:

$$\bar{y}_{mr} - \bar{y}_{fr} = (\bar{X}_{mr} - \bar{X}_{fr})\gamma + \bar{\varepsilon}_{mr} - \bar{\varepsilon}_{fr}$$

so by running the regression:

$$\bar{y}_{mr} - \bar{y}_{fr} = \beta \log(density)_r + \bar{\varepsilon}_{mr} - \bar{\varepsilon}_{fr}$$

β would be reflecting the correlation between CZ population density and the average gap between male and female characteristics. This omitted variable problem is easily resolved by running the regression:

$$\bar{y}_{mr} - \bar{y}_{fr} = (\bar{X}_{mr} - \bar{X}_{fr})\gamma + \beta \log(density)_r + \bar{\varepsilon}_{mr} - \bar{\varepsilon}_{fr} \tag{1}$$

Things to have in mind

- These regressions impose the same return to observable characteristics for men and women in all CZ. Differential returns across CZ will go into the residual.

Procedure

Aggregate level data I run the regression

$$\bar{y}_{mr} - \bar{y}_{fr} = \alpha_t + (\bar{X}_{mr} - \bar{X}_{fr})\gamma_t + \beta_t \log(density)_r + u_{rt}$$

where I allow the return to observable characteristics to vary by year. The main interest is looking at the resulting evolution of β_t.

Individual level data This just allows for a more flexible variation on the returns of age birth place. Here the estimation is done in two steps:

1. Estimate the regression:

$$y_{igr} = X_{igr}\gamma + \lambda_{gr} + \varepsilon_{igr} \tag{2}$$

2. Compute CZ-adjusted wage gap:

$$\tau_r = \lambda_{mr} - \lambda_{fr}$$

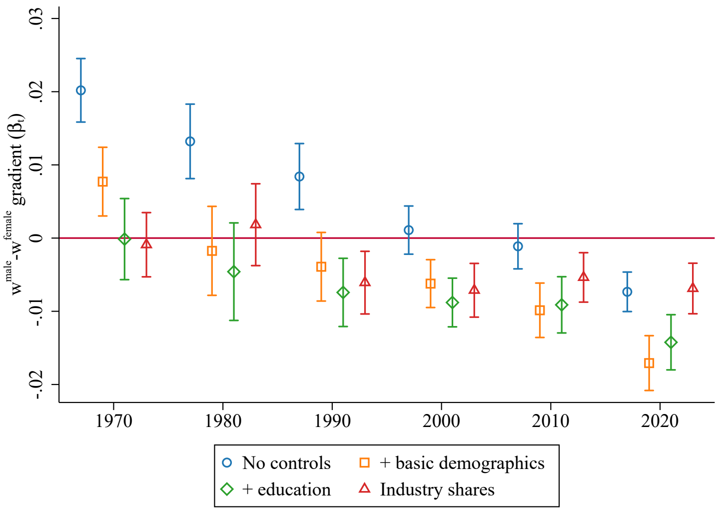
3. Run the regression:

$$\tau_r = \alpha_t + \beta_t \log(density)_r$$

I prefer this method as it exploits the individual level data in a richer way.

Results Overall individual level characteristics have limited value in accounting for the cross-sectional gradient and time-variation. Industry-level dummies are much more successful in accounting for the 1970-1990 period.

Figure 10: Coefficient on population density β_t controlling for worker characteristics



Note: figure restricts to CZ with more than 1 people per km^2. The regressions are done on data aggregated at the CZ level. Bars show 95% robust confidence intervals. Standard errors clustered at the CZ level.

9.3 Changes in CZ industrial structure

Figure 4 suggests that changes in the CZ industrial structure can go a long way in accounting for the cross-sectional variation. Here I do several exercises to explore this possibility.

9.3.1 Some national level facts

Women are initially concentrated in low-pay industries See here => regions specialized on these high-pay industries will show a higher gender gap.

Getting a workable definition of a high-wage 70s industry A high wage industry is one that has a high worker-adjusted average pay in 1970. To be more precise, using individual level data in 1970 I run:

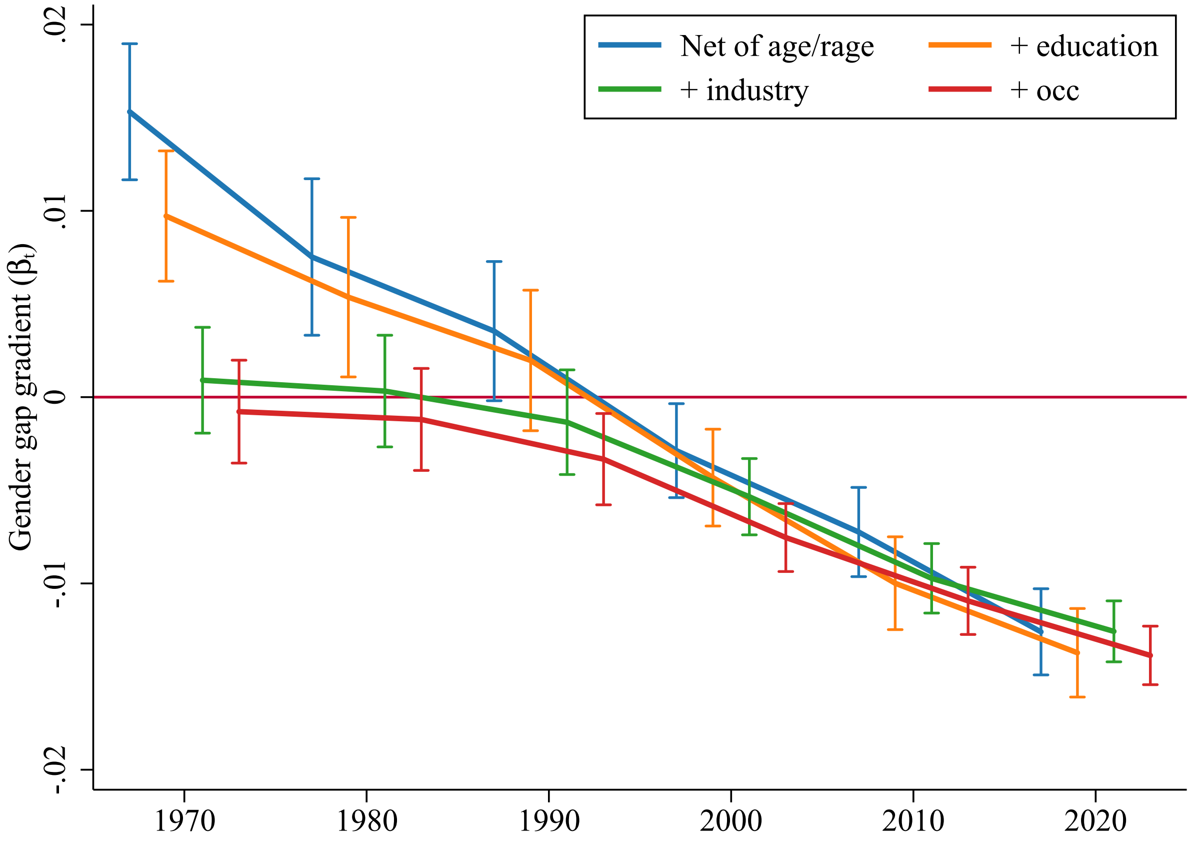
$$y_i = X_i\beta + \lambda_s \tag{3}$$

where s denotes the industry. I define an industry as having high-pay if they are in the top quartile of the λ_s distribution. When computing the quartiles, industries are weighted by employment share so that in 1970 each quartile accounts for 25% of the national level.

High pay industries were disproportionately concentrated in denser places in 1970 (graph)

High pay industries are in decline at the national level (graph) decline at the national level starts in 1990

Figure 11: Coefficient on population density β_t controlling for worker characteristics



Note: figure restricts to CZ with more than 1 people per km^2. The regressions are done on data aggregated at the CZ level. Bars show 90% robust confidence intervals. Standard errors clustered at the CZ level. Figure generated on 23 Nov 2020 at 16:25:28. Figure generated using the dofile 2_analysis/code_files/write_regression_coefplots.do.

High pay industries belong mainly to manufacturing (log file) the rest are mostly oil or utilities.

Employment share in highly paid industries accounts for most cross-sectional gradient on density during 1970-90 (graph) it also accounts for the time variation from 1970-90s. There's still something going on from 90 to 20

What can be happening:

- High wage industries are in decline in denser places... then the decline decline in male advantage comes from employment reallocation.
- It can be that at the start of the period, women are getting better access these industries. *I think for the 90's this seems to be the case.*

These industries continue to be highly paid industries See here

Denser CZ are more specialized in 70s high pay industries See here

70s high-pay industries decline disproportionately more in denser places

Women here

10 When would the national level fixed-effects account for the cross-sectional variation?

Suppose the wage profile of individuals is determined as follows:

$$w_{ir} = \lambda_e \tag{4}$$

where λ_e is a national-level fixed effects. Then,

$$\bar{w}_r^g = \sum_e \lambda_e s_r^g \tag{5}$$

then the gender gap in a given commuting zone is given by,

$$\bar{w}_r^m - \bar{w}_r^f = \sum_e^g \lambda_e^g (s_r^m - s_r^f) \tag{6}$$

suppose that:

$$s_r^g = \alpha_e^g + \beta_e^g \log(density)_r \tag{7}$$

then gender gap equation becomes,

$$\bar{w}_r^m - \bar{w}_r^f = \sum_e \lambda_e (\alpha_e^m - \alpha_e^f) + \sum_e \lambda_e (\beta_e^m - \beta_e^f) \log(density)_r \tag{8}$$

It follows that $\sum_e \beta_e^g = 1$. This follows from the identity below holding for all CZ:

$$1 = \sum_e s_{er}^g = \sum_e \alpha_e^g + (\sum_e \beta_e^g) \log(density)_r \tag{9}$$

So a negative coefficient in density requires:

$$\sum_e \lambda_e (\beta_e^m - \beta_e^f) < 0 \tag{10}$$

Which roughly requires that have a higher gradient on density in employment in "high-pay" groups relative to men.

Now suppose that the wage profile is given by:

$$w_{ir} = \lambda_e^g \tag{11}$$

then,

$$\bar{w}_r^m - \bar{w}_r^f = \sum_e \lambda_e^m s_{er}^m - \lambda_e^f s_{er}^f \tag{12}$$

$$= \sum_e (\lambda_e^m - \lambda_e^f) s_{er}^m + \lambda_e^f (s_{er}^m - s_{er}^f) \tag{13}$$

thus it would have to be that the employment structure of the densest CZ is concentrated in those groups that experience the largest gender gap decline at the national level.

A On weighting

Here the basic question I want to answer is, can I have a good answer as to why I am not weighting.

Suppose wages are determined according to the model:

$$w_{ir}^g = \beta X_{ir}^g + \varepsilon_{ir}^g \tag{14}$$

where $\varepsilon_{ir}^g = \gamma_r^g + u_{ir}$ where γ_r^g and u_{ir} are independent.

$$\begin{aligned} \bar{w}_r^m - \bar{w}_r^f &= \beta(X_r^m - X_r^f) + \bar{\varepsilon}_r^m - \bar{\varepsilon}_r^f \\ &= \beta(X_r^m - X_r^f) + v_r \end{aligned}$$

note that if we assume that $var(\gamma_r^g) = \sigma_\gamma^2$ and $var(u_{ir}) = \sigma_u^2$.

$$\begin{aligned} var(v_r) &= var(\bar{\varepsilon}_r^m) + var(\bar{\varepsilon}_r^f) - cov(\bar{\varepsilon}_r^m, \bar{\varepsilon}_r^f) \\ &= 2\sigma_\gamma^2 + \sigma_u^2 \left(\tfrac{1}{N_m} + \tfrac{1}{N_f}\right) \end{aligned}$$

so, in the end I can test whether heteroskedasticity is a problem by running the regression (15) by OLS, extract the residuals and then run the regression:

$$\hat{u}_r = \alpha + \beta \left(\frac{1}{N_m} + \frac{1}{N_f}\right)$$

The results from this exercise give little justification for weighting the regressions. See this log-file.

B On the interpretation of the coefficients

The main findings come from regressions of the form:

$$\ln\left(w^{male} - w^{female}\right) = \alpha_t + \beta_t \ln(pop\_density)_{rt} + \epsilon_{rt}$$

without giving any causal interpretation to the coefficients, what is the interpretation of β_t?

B.1 Mathematical interpretation

First note that:

$$\beta_t = \frac{\partial \ln \left(w^{male} / w^{female} \right)}{\partial \ln pop\_density_{rt}}$$

thus β_t can be interpreted as the elasticity of the male wage advantage with respect to population density. Table 2 shows the estimated elasticities.

Table 2: Elasticities of male wage advantage to population density

Regression specification	**1970**	**1980**	**1990**	**2000**	**2010**	**2020**
Unweighted OLS	0.046***	0.030***	0.019***	0.003	-0.003	-0.017***
	(0.005)	(0.006)	(0.005)	(0.004)	(0.004)	(0.003)
Weighted by population	0.034**	0.007	-0.019*	-0.026**	-0.024***	-0.026***
	(0.013)	(0.013)	(0.008)	(0.008)	(0.006)	(0.005)
Observations	625	625	625	625	625	625

Note: Robust standard errors in parenthesis. Sample restricts to full-time year-round workers.. Table generated on 14 Aug 2020 at 15:39:24.

B.2 Economic interpretation

My basic results show that this elasticity:

- Has continually declined since the 1970s.
- It went from positive in 1970, to negative by 2020.

These elasticities are big and economically significant. Tables and 4 shows two ways of putting these numbers into perspective. Table 3 shows that moving form a CZ with a population density in the 25th percentile, to one in the 75th percentile translates into an increase of 4 p.p. in the male wage advantage in 1970. However, this same movement is associated with a decrease of 1.5 p.p. in the male advantage in 2020. These movements are equivalent to:

- A change of ∼10% relative to the average male advantage in each respective year.
- If we translate these figures to dollar amounts using as reference the annual wage income of the average full-time female worker, moving from the 25th to the 75th percentile in population density translates into \$1.1 k relative gain for men in 1970, but a \$0.6k relative loss in 2020.

Additionally, table 4 shows the estimated elasticities when I transform both the wage gaps and the CZ population density. If we focus on the unweighted OLS estimates, an increase of 1 sd in the log of population density is associated with 0.3 sd in the male wage advantage in 1970, but with a 0.2 sd deviations decrease in 2020.

Table 3: Male advantange changes implied by estimated elasticities

p.p. change in male advantage	**1970**	**1980**	**1990**	**2000**	**2010**	**2020**
Average male advantage	0.44	0.41	0.33	0.26	0.20	0.19
p75-p25	0.04	0.03	0.02	0.00	-0.00	-0.01
Relative male gain ($ USD)	1,116	793	534	73	-80	-559
p85-p15	0.08	0.05	0.03	0.00	-0.00	-0.02
Relative male gain ($ USD)	1,898	1,313	877	122	-134	-910
p90-p10	0.09	0.06	0.04	0.00	-0.00	-0.03
Relative male gain ($ USD)	2,379	1,644	1,088	153	-167	-1,159

Note: changes based on unweighted estimated elasticities in table 2. Sample restricted to full-time year-round workers. I compute the dollar figures using the wage of the average full-time year-round woman in my sample, assuming she worked 40 hrs a week during 40 weeks. All figures are in 2018 dollars. Table generated on 14 Aug 2020 at 18:35:47. Table generated with do file 2_analysis/code_files/create_IC_table.do

Table 4: β_t on standardized data

Regression specification	**1970**	**1980**	**1990**	**2000**	**2010**	**2020**
Unweighted OLS	0.330***	0.205***	0.176***	0.029	-0.030	-0.192***
	(0.036)	(0.040)	(0.048)	(0.044)	(0.042)	(0.036)
Weighted by population	0.244**	0.047	-0.173*	-0.301**	-0.283***	-0.300***
	(0.089)	(0.089)	(0.076)	(0.091)	(0.074)	(0.057)
Observations	625	625	625	625	625	625

Note: Robust standard errors in parenthesis. Sample restricts to full-time year-round workers.. Table generated on 14 Aug 2020 at 16:48:21.

References

Black, D. A., Kolesnikova, N., and Taylor, L. J. (2014). Why do so few women work in New York (and so many in Minneapolis)α Labor supply of married women across US cities. *Journal of Urban Economics*, 79:59–71.

Black, S. E. and Spitz-Oener, A. (2010). Explaining women's success: Technological change and the skill content of women's work. *Review of Economics and Statistics*, 92(1):187–194.

Duranton, G. and Puga, D. (2020). The economics of urban density.

Goldin, C. (2014). A grand gender convergence: Its last chapter. *American Economic Review*, 104(4):1091–1119.

Liu, S. and Su, Y. (2020). The Geography of Jobs and the Gender Wage Gap. Technical report.

Mulligan, C. B. and Rubinstein, Y. (2008). Selection, investment, and women's relative wages over time.

Olivetti, C. and Petrongolo, B. (2014). Gender gaps across countries and skills: Demand, supply and the industry structure. *Review of Economic Dynamics*, 17(4):842–859.

Olivetti, C. and Petrongolo, B. (2016). The Evolution of Gender Gaps in Industrialized Countries. *Annual Review of Economics*, 8(1):405–434.