Analysis of Evolution of Hispanic Wage Premium and its Reason

Ziying Cheng Zijun Meng Yanjing Qu Hanchen Wang Yuanjian Zhou

1. Question

This report mainly focuses on what has changed in income inequality between the Hispanic and the Non-hispanic, as well as the reason driving this change. To fully understand the wage gap between them, this report also focuses on whether the changes in the share of the wage bill paid to those with Hispanic origin accounted for by within industry or between industry changes.

2. Motivation

According to Wilkinson and Pickett(2009) [1], income inequality is highly associated with several social problems, such as violence and low educational performance. For the whole country, Turnovksy(2006) [2] also find that there is a mutual effect between income equality and national capital accumulation. As one of the largest ethnic groups in the United States, the population of hispanic group is still rising while their wage is significantly lower than non-hispanic groups: the median household income was 51450 dollars, which is about 10000 dollars less than white people, showing a significant income inequality(Census, 2019) [3]. We hope that this analysis will attract people's attention to this important issue and might be a reference for relevant policy.

3. Data

3.1 Basic Cleaning

We use 1971-2018 CPS(Current Population Survey) data [4] for this analysis, and restrict the samples to be between 17 and 65 years old. For subsequent process, we generate education indicators and create bins which representing combinations of demographic variables, such as education, origin and sex.

3.2 Deal with Top Coding

Due to some reason, CPS data will have top coding on extremely high income. If an individual has income over the threshold, his income will be topcoded as a certain number which is different in each year. As instructed, in order to represent the income distribution more precisely, we multiple 1.5 to each threshold numberto to adjust the top-level income.

3.3 Composition Adjustment

As we learn in the lectures, calculation of wage premium might be misleading because of change of composition. In order to control the weight of each bin, holding the share of income earned by each group fixed, we utilize composition adjustment method. First, we sum up incomes of each bin to get a total income for each origin group(hispanic and non-hispanic), and get the wage of each bin from data. Then we derive a weight of wage that will be fixed overtime, which is the sum of working hours in all year for this bin, divided by the total working hours of hispanic/non-hispanic group that this bin belongs to. The formulas is as follows:

$$w_g = \frac{\sum_t hours_{gt}}{\sum_{g't'} hours_{g't'}} \tag{1}$$

Then we multiply wage of each bin by this weight to get the total wage for each origin:

$$W_{Ht} = \sum_g w_g lnW_{gt} \tag{2}$$

Finally we construct hours of each origin by using:

$$ln(H_t) = ln(inc_t) - ln(W_{Ht}) \tag{3}$$

Under this method, we can have the weight unchanged and get the true hispanic wage premium.

4. Approach

4.1 Canonical Model

To investigate the relationship between wage premium and relative supply, we develop the canonical model. Initially, we used relative working hours, technological change to explain wage premium. However, technological change is not observable, therefore it goes into residual. Then our regression form of model becomes:

$$ln(\frac{W_{Ht}}{W_{NHt}}) = \beta ln(\frac{H_{Ht}}{H_{NHt}}) + \tau t + \epsilon_t \tag{4}$$

4.2 Instrument Variable

For causality, people may think there are confounding factors that affect both relative supply and wage premium, for example, education. Then instrumental variable is used to avoid endogeneity between variables. Instead of hours, we use population between hispanic and non-hispanic group. Ideally, population only affects working hours, not related to the error term in the regression. First we run regression with population and hours worked to get estimated population:

$$\frac{POP_{Ht}}{POP_{NHt}} = \beta_0 + \beta_1 t + \beta_2 ln(\frac{H_{Ht}}{H_{NHt}}) + \epsilon_t \tag{5}$$

Then we run this regression:

$$ln(\frac{W_{Ht}}{W_{NHt}}) = \beta_0 + \beta_1 t + \beta_2 \frac{POP_{Ht}}{POP_{NHt}} + \epsilon_t \tag{6}$$

4.3 Shift-Share Analysis

Unfortunately, with the dataset we have, we can not research the relationship between wage premium precisely. But at least we can take a quick-pass at whether a relative growth of hispanic-intensive industries, or a change in the relative importance of hispanic workers within industries accounts for the change of hispanic workers' wage bill share of total income. To analyze this, we do the so-called shift-share analysis: Decompose the share of total wage bill paid to hispanic in two

parts as follows:

$$\frac{WB_{Ht}}{WB_t} = \frac{1}{WB_t}\sum_i WB_{Hit} = \sum_i \frac{WB_{Hit}}{WB_{it}}\frac{WB_{it}}{WB_t} \tag{7}$$

where WB means wage bill, H denotes Hispanic, t denotes time and i denotes industry.

Using this formula, it is obvious that the change in the share can also be divided in two parts, the average share of working in certain industry multiplied by the change of share of hispanic wage in that industry, and the change in the share of specific industry multiplied by the average share of hispanic wage bill:

$$\Delta\frac{WB_{Ht}}{WB_t} = \sum_i \overline{\frac{WB_{Hit}}{WB_{it}}} \times \Delta\frac{WB_{it}}{WB_t} + \sum_i \Delta\frac{WB_{Hit}}{WB_{it}} \times \overline{\frac{WB_{it}}{WB_t}} \tag{8}$$

5. Results

5.1 Wage Premium and Relative Supply

First we look at what the data tell us about the evolution of wage premium and relative supply. As shown in figure 1, the hispanic wage premium has been decreasing generally since 1970 while the decreasing rate became slower in the past decade, probably owing to government policy setting minimum wage for the hispanic. In contrast to the wage premium, since 1970, the relative supply of hispanic group has increased, which can be easily explained by immigration or government policies. With more and more hispanic people are willing or allowed to enter the labor force and have a full-time job, the supply of hispanic labor constantly increases.

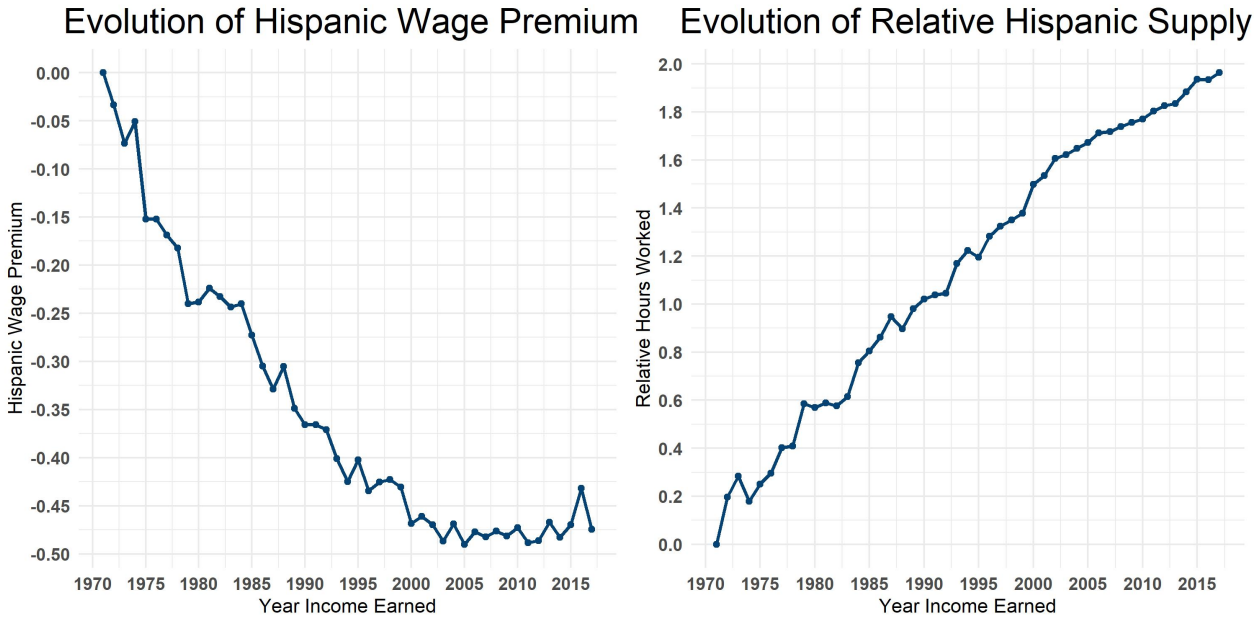

Figure 1: Evolution of Hispanic Wage Premium and Relative Supply

Then, we run the canonical model using OLS and 2SLS. Regression Results are shown in the following table:

Table 1: Regression Results

	Dependent variable:		
	Wage Premium	Relative Population	Wage Premium
	OLS	*FIRST STAGE*	*2SLS*
	(1)	(2)	(3)
Time Trend	0.016***	0.006***	0.042***
	(0.003)	(0.001)	(0.010)
Relative Supply	-0.597***	-0.053***	-1.215***
	(0.059)	(0.015)	(0.236)
Constant	-0.024*	-0.009***	0.084*
	(0.014)	(0.003)	(0.044)
Observations	47	47	47
R^2	0.956	0.982	0.848
Adjusted R^2	0.954	0.981	0.841
Note:			*p<0.1; **p<0.05; ***p<0.01

In table 1, OLS model shows that the relative supply has a statistically significant negative impact on the wage premium; The first stage of 2LS has R-squared 0.982, which means that population ratio is a good instrument; The 2SLS also shows a negative effect on the wage premium, and the significance is still high. Noteworthy, the time trend also has a high level of significance in both OLS and 2SLS models. Therefore to pin down the actual effect of relative supply, we will represent the actual-fitted plots for both original and detrended data.

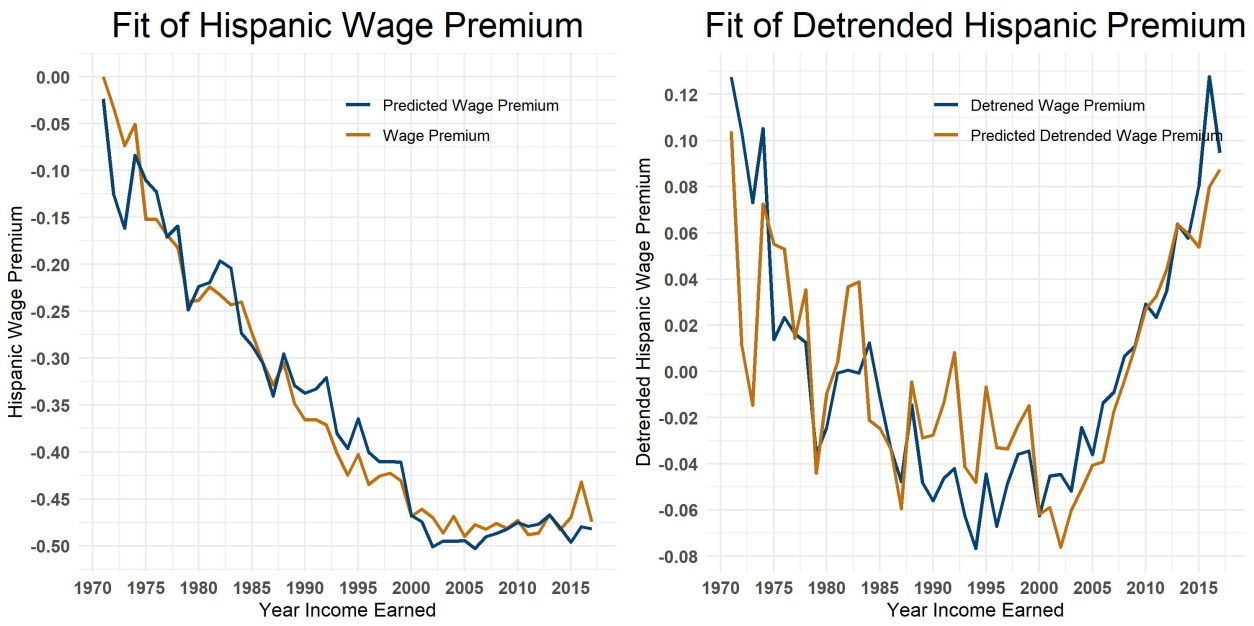

Figure 2: Fit of Original and Detrend Hispanic Wage Premium

As shown in figure 2, the fitted line and the actual line is quite close, which demonstrates that the model explains the wage premium well. Removing the time trend from wage premium and hispanic relative supply, we do the OLS regression again. The two lines still come together at most time, demonstrating a strong interpreting power of hispanic relative labor supply towards hispanic wage premium. However, instead of constantly decreasing in shape, this graph has frequent slope change, which means that the hispanic wage premium is under fluctuation caused by revolution of relative supply.

Finally, Figure 3 displays the relationship between the detrended hispanic relative supply and detrended hispanic wage premium. The detrended hispanic relative supply always moves towards the opposite direction of wage premium, implying a strongly negative effect between them, as shown in the OLS model.

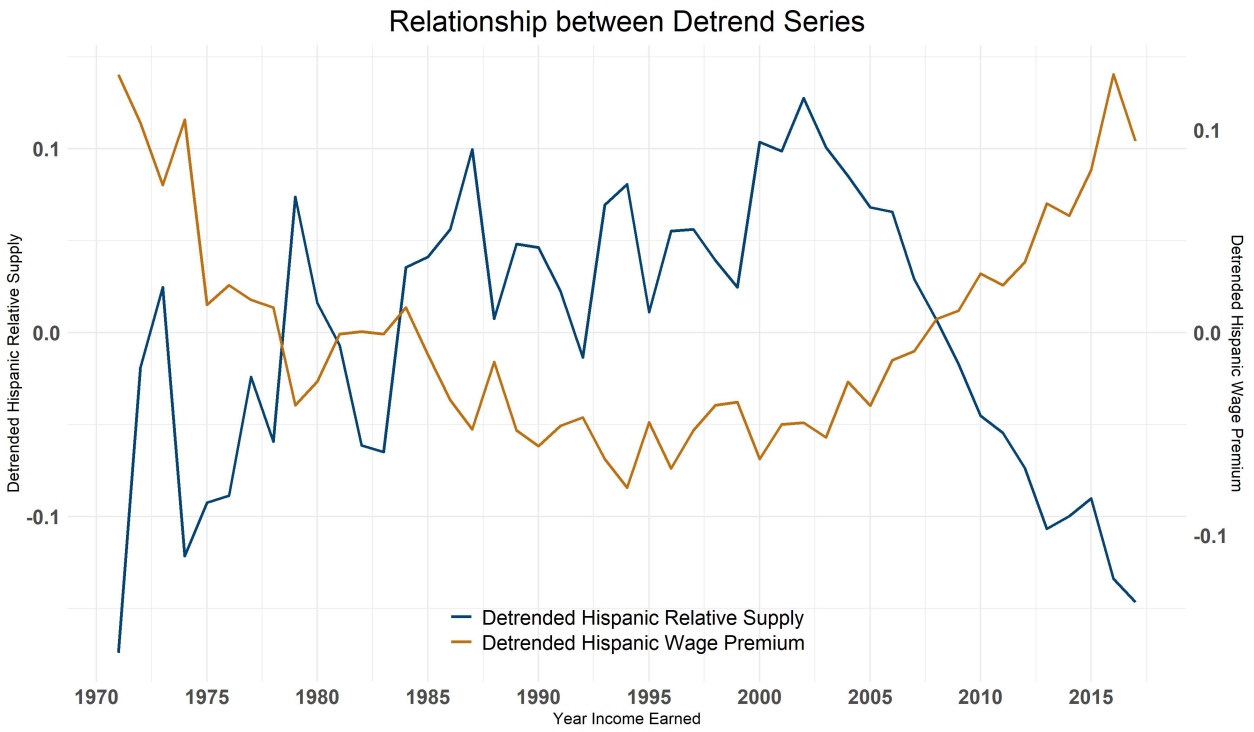

Figure 3: Relationship between Detrended Series

5.2 IV Robustness

In this section, we will plot the same four plots above using IV estimated relative supply to replace supply.

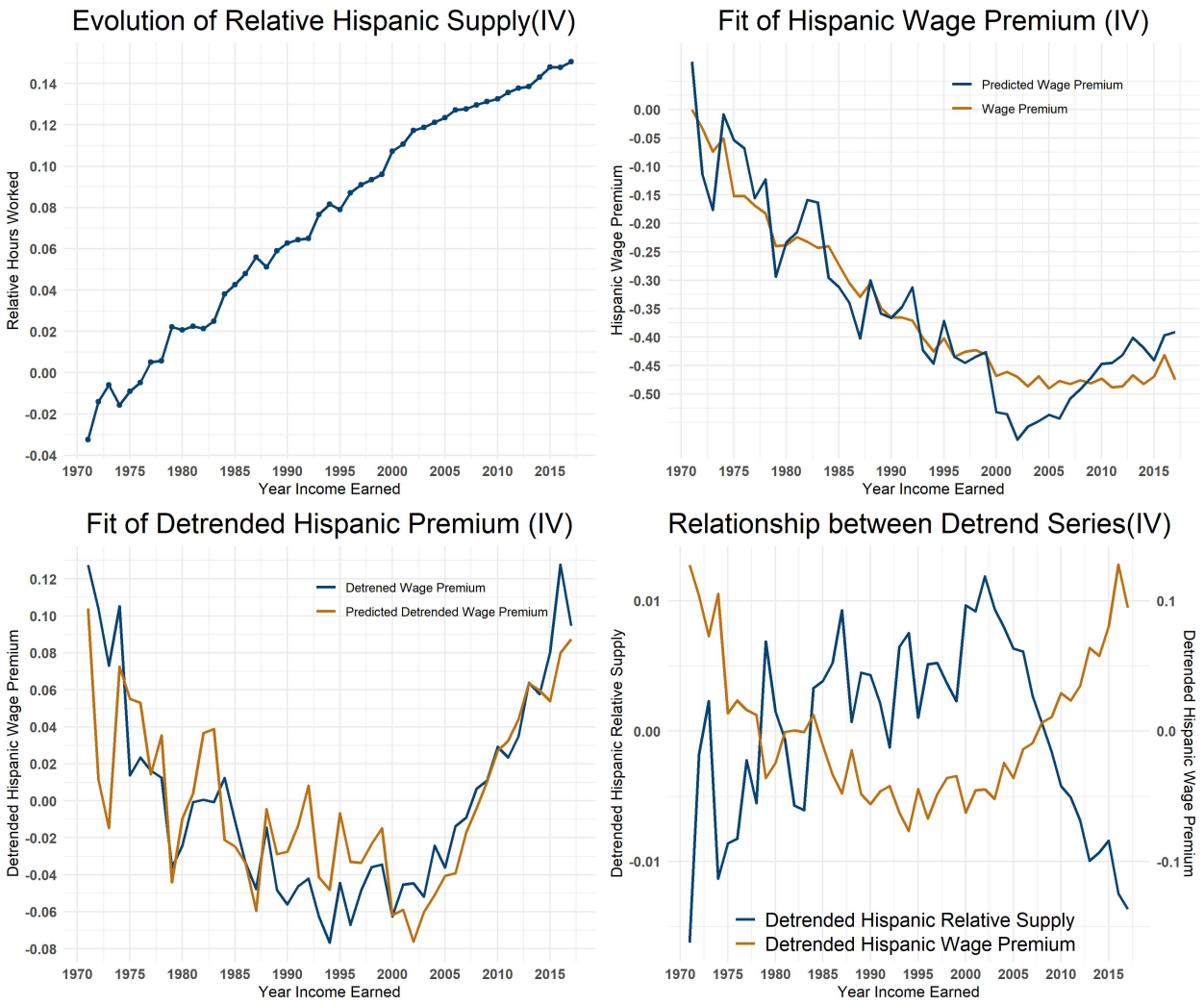

Figure 4: Relevant Plots using IV

Noteworthy, the two lines in the upright graph go apart between 2000 and 2005, which is different from Figure 2. We might try to fix our eye on this time range to figure out what has happened in the following study.

5.3 Shift-Share Analysis

Then we take a quick look at the Shift-Share analysis result using the formula in 4.2.

Table 2:

between group change	within group change	total change
-0.6%	10.2%	9.7%

According to Table 2, the share of wage bill paid to hispanic increased by 9.7% from 1970 to 2017. To be more specific, the share of wage bill paid in hispanic-intensive industries descended by 0.6% while the share of income of hispanic workers within industry increased by 10.2%. In other words, from 1971 to 2017, the relative importance of hispanic-intensive industries has shrinked, while the average income gap within industries shortened.

However, interpreting this decomposition as a causal statement about the relative importance of changes in relative industry demand versus relative demand for hispanic workers is dangerous, partly because it does not take changes of working hours into consideration. What's more, these results seem to be in contradict with the actual revolution of hispanic wage premium.

6. Conclusion

In conclusion, our analysis shows that the wage premium of hispanic workers has been decreasing and illustrates a negative relationship between relative supply and the wage premium. According to the shift share analysis, at least from an accounting perspective, relative importance of hispanic-intensive industries has become smaller while the income gap between hispanic and non-hispanic workers has shrinked.

While there are some conclusions suggested in this report, there is yet to be much more future work. For instance, what has happened during 2000-2005 making the IV model fit not so well ? Can we investigate further to find the casual relationship between relative demand and hispanic wage premium?

References

[1] Richard G. Wilkinson and Kate E. Pickett. Income inequality and social dysfunction. *Annual Review of Sociology*, 35:493–511, 2009.

[2] Cecilia García-Peñalosa and Stephen J. Turnovsky. Growth and income inequality: A canonical model. *Economic Theory*, 28(1):25–49, 2006.

[3] Income and poverty in the united states: 2018. `https://www.census.gov/library/publications/2019/demo/p60-266.html`.

[4] Renae Rodgers Steven Ruggles Sarah Flood, Miriam King and J. Robert Warren. Integrated public use microdata series, current population survey: Version 6.0 [dataset], 2018.