Undergraduate Honors Thesis: Examining local price levels and income distribution over time

Josh Archer *University of California, Berkeley* *

May 14, 2020

Abstract

While the relationship between price indexes and income inequality is studied on the state and national level, how do these measures relate at the city level? Due to many of the shortcomings with price indexes such as availability of goods, quality change, and non-homothetic preferences I exploit small variations in city level American Community Survey cost of living indexes to compare with a variety of income inequality measures. I find that measures of inequality and high incomes are positively related to local price indexes while low incomes have a negative relationship to price levels.

1 Introduction

Is there a relationship between the changes in income distribution and the changes in price indexes at the city level? There is a large literature in economics that attempts to identify

*I would like to thank Professor Emi Nakamura for her immense support as my undergraduate thesis advisor. I would also like to thank Professor Barry Eichengreen, Joan Martinez, Marcus Sander, Matt Tauzer for their help with a preliminary work and with the mechanics of writing a research paper. All mistakes and shortcomings are my own. Email: jearcher@berkeley.edu

this relationship in the state or country level, however not at the city level. It is well reported that folks in the US are moving out of state about half as much as they did a generation ago (HuffPost). If city-level price levels are related to city-level inequality, this begs a question for future research– why aren't people moving? In this paper, however, I attempt only to establish the relationship between the cost of living (price indexes) and income distribution (inequality). On the national level, the US experiences stable prices with approximately 2% inflation each year, yet many media outlets and researchers point to growing inequality in the US. In the literature at the state and national level, some report that the lowest income levels face higher price levels (Tipping 1970) while others report that income grows faster than price levels for low income households (Raymond 1985). Additionally, the relationship between inequality and price levels or price level growth (inflation) is somewhat up for debate. Some researchers find low inflation improves inequality at the country level (Bulıř 2001) or that inequality *causes* inflation (Beetsma and Van Der Ploeg 1996). In more recent research, some say the relationship is non-linear. For example, Balcilar et al find that inflation does increase inequality at the state level when inflation is above a certain threshold, but it improves inequality below this threshold (Balcilar et al. 2018). Given this research, I hope to investigate this relationship further and to add to the literature by investigation changes in the price levels of a variety of goods and changes to the income at the city level. I find the relationships to be much stronger when controlling for city-level fixed effects (compare the results from Table 2, Columns (1)-(4) to Table 1 Columns (5)-(6)).

A positive relationship between measures of inequality and price indexes has quite a few implications. For example, many consider home ownership a key channel to improve household economic incomes and house prices have steadily been increasing since the recovery from the Great Recession. I find that all incomes are increasing but higher incomes are increasing at a faster rate. Higher and lower income households face the very similar city level price indices, which leaves less available income for lower income households to purchase goods like a home. I do not find causality in this paper, however I note the possible implications for my results and that there is room for further research.

The paper is organized as follows: Section 2 gives background and context on price index construction in the United States, Section 3 outlines the data and the empirical strategy, Section 4 presents the results, section 5 is the discussion and conclusion, and section 6 is the appendix.

2 Background

While many researchers have shown that inequality is rising in the US, perhaps the most influential work comes from Emmanuel Saez and Thomas Piketty. In their 2003 paper, they describe the income inequality increases over the last few decades of the 20th century, particularly driven by an increase in the 'working rich' as opposed to the capital owners at the beginning of the 20th century (Piketty and Saez 2003). This rising inequality has persisted into and after the Great Recession, and some economists cite that consumption inequality follows a similar pattern (Attanasio, Hurst, and Pistaferri 2012). While many economists and media outlets agree that inequality has grown in the US since the 1970s, there is much disagreement on the relationship between inequality and inflation, including how to measure inflation. The typical measure of inflation is provided through the Consumer Price Index which is reported monthly by the Bureau of Labor Statistics (BLS). According to the CPI handbook of methods, one of the main CPI series is the CPI for all urban consumers, which is a single number released monthly to track aggregate relative price changes indexed to 1982-1984 (Bureau of Labor Statistics 2018). This informs many as to the levels of inflation year over year, however plenty of economists have pointed out the shortcomings of this measure. The BLS measure doesn't fully and correctly account for quality changes and new goods. There is a complex manner to introduce new goods in the price index from year to year where we must control for quality changes and some researchers suggest that these indices may be incorrect by pointing to specific items. William Nordhaus, for example, writes about quality changes in price indexes with regards to lightbulbs, with the introduction of high quality LED lights the price index may vastly overstate the true price increases (Nordhaus 1998)

In addition to the new goods and quality change biases, these indices rest on the assumption of homothetic preferences, that is, the proportion of a households income spent on a particular product is constant throughout income changes. For example, suppose my income is $40K per year and I spend 30% of my income on food ($12,000). Under the assumption of homothetic preferences, if my income tripled to $120K per year, I would still spend 30% of my income on food ($36,000). This assumption is clearly invalid when we extend it to to all goods, that is, consumers' Engel curves are not horizontal. Related to this notion, one can expect high income households to purchase a compositionally different basket of goods than a low income household, yet the CPI applies uniformly. In addition to this, households of equal income are likely to have access to a very different set of goods in Berkeley, California than households in Uniontown, Ohio.

From the theoretical background, there is forthcoming research to find a theory for Engel Curves in order to back out peoples consumption habits which could be exploited to craft price indexes for households with non-homothetic preferences (Atkin et al. 2018). However, this theory has yet to be established and I will assume homothetic preferences for this analysis

To take this a step further, some researchers have pointed to the differences in goods availability in cities (Handbury and Weinstein 2014). If the basket of goods is not the same in each city, how can we evaluate all price indexes with just one number? For this reason, I attempt exploit the variation in price indexes at the city level to compare to income changes.

Overall I hope that by focusing on city level price indexes and income, my results will be less influenced by availability differences and quality differences. Consumer's within cities are likely to have more similar consumption purchases than consumers across the country.

3 Data and Methodology

3.1 Data Sources and Sample

My sample will include cross-sectional data on 38 representative cities from 2005 to 2017. The choice of cities is not entirely random, some cities had to be omitted due to merging issues between the ACS and IPUMS data. I only included data where the ACS and IPUMS data precisely agree in order to avoid erroneously comparing two different geographies. Additionally, the IPUMS data does not include the years 2001-2004, so I decided to use the years 2005-2017.

I use the American Community Survey (ACS) Cost of Living price indexes to form city level price indexes from. Most of the ACS price indexes are reported in each quarter, so I have aggregated the data to create annual cost of living indexes in order to compare this with annual income data. I take a sample of 38 cities which I can match up with income data. The list of cities in the sample can be found in the appendix (section 6.1)

I use the IPUMS data to create income statistics for each city in the sample. First I removed removed all data coded as Not In Universe (NIU)[1], then created a variety of income statistics for the IPUMS data. In order to protect the privacy of those in the IPUMS data, the income data is topcoded[2]. Thus the usual inequality statistics such as the Gini Coefficient are likely to biased.

1. The NIU data include people who are too young or too old to work.

2. If an individual's income is very high, it is coded as a some large number. Generally if an income is above the 99.5th percentile income for the state, it will be coded as the median of all incomes greater than the 99.5th percentlie income for the state. For more information, see https://usa.ipums.org/usa-action/variables/INCWAGE#codes_section

Instead I have used the ratio of percentile incomes as a measure of inequality. For example, $\frac{Y^{90}}{Y^{10}}_{c,t}$ is the 90th percentile income divided by the 10th percentile income for a particular city in a particular year. If this value grows over time, this means high incomes are growing faster than low incomes. I also use the 80th percentile income divided by the 20th percentile income as an additional measure ($\frac{Y^{80}}{Y^{20}}$).

In total, I use the 90th percentile income, the 20th percentile income, the ratio of the 80th percentile of income divided by the 20th percentile of income, and the ratio of the 90th percentile of income divided by the 10th percentile of income. These will serve as a variety of income and inequality statistics to compare to the price indexes. Overall, with missing values omitted, the sample of all cities over all years is 336 observations.

3.2 Empirical Strategy

The goal of my empirical analysis is to compare various income and inequality measures with price levels at the city level over time. I have run the following regressions on the sample of 38 cities from 2005-2017

$$p_{c,t} = \beta_0 + \psi_t + \theta_c + \beta_1 \log(Y^{20}_{c,t}) + u_{c,t} \tag{1}$$

$$p_{c,t} = \beta_0 + \psi_t + \theta_c + \beta_1 \log(Y^{90}_{c,t}) + u_{c,t} \tag{2}$$

$$p_{c,t} = \beta_0 + \psi_t + \theta_c + \beta_1 \left(\frac{Y^{80}}{Y^{20}}\right)_{c,t} + u_{c,t} \tag{3}$$

$$p_{c,t} = \beta_0 + \psi_t + \theta_c + \beta_1 \left(\frac{Y^{90}}{Y^{10}}\right)_{c,t} + u_{c,t} \tag{4}$$

The subscript c denotes cities and t denotes time measured in years. $p_{c,t}$ is the city level price index (from the ACS data). θ_c is a set of city fixed effects and ψ_t is the set of time fixed effects. $Y^{20}_{c,t}$ and $Y^{90}_{c,t}$ are the 20$^{\text{th}}$ and 90$^{\text{th}}$ percentile of income, respectively. $\left(\frac{Y^{80}}{Y^{20}}\right)_{c,t}$ and $\left(\frac{Y^{90}}{Y^{10}}\right)_{c,t}$ are the ratios of the 80$^{\text{th}}$ over the 20$^{\text{th}}$ percentile incomes and the ratio of the 90$^{\text{th}}$ over the 10$^{\text{th}}$ percentile incomes, respectively.

The coefficient of interest in each regression is β_1. For regressions 1, 2, the coefficient estimate will reveal how related the increases to income of the 20th or 90th percentiles are with the rises in price levels. That is, as incomes for the 20th or 90th percentile rise, how much does the price level rise? Regressions 3 and 4, therefore, may include the most accurate measures of inequality: the ratio of top incomes with lower incomes. A positive and significant $\hat{\beta}_1$ estimate for regressions 3 and 4 imply that price level increases are associated with high incomes rising *faster* than lower incomes.

4 Interpreting Results

4.1 Summary of Results

Tables 1 and 2 outline the regression results with and without city and/or time fixed effects. Table 1 columns (1) - (4) omit the city and time fixed effects. Columns (5) - (6) include both city and time fixed effects. My hope is that I can exploit the city and time variation to test if there is relationship between income distribution and price level changes at the local level.

Table 2 runs the same regressions with only city fixed effects, (1) - (4), or only time fixed effects, (5) - (8).

Table 1:

	Dependent variable: Price Index, $p_{c,t}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\log(Y^{20})$	10.452^{***}				-3.323^{*}			
	(2.475)				(1.922)			
$\log(Y^{90})$		35.118^{***}				19.620^{***}		
		(2.998)				(4.760)		
$\frac{Y^{80}}{Y^{20}}$			1.746^{***}				0.888^{***}	
			(0.432)				(0.240)	
$\frac{Y^{90}}{Y^{20}}$				0.590^{***}				0.175^{***}
				(0.106)				(0.053)
Constant	11.161	-289.776^{***}	94.419^{***}	92.410^{***}	138.179^{***}	-106.496^{**}	99.879^{***}	103.674^{***}
	(22.470)	(33.790)	(2.972)	(2.562)	(16.284)	(52.592)	(3.195)	(2.525)
City Fixed Effects	No	No	No	No	Yes	Yes	Yes	Yes
Year Fixed Effects	No	No	No	No	Yes	Yes	Yes	Yes
Observations	336	336	336	336	336	336	336	336
R^2	0.051	0.291	0.047	0.085	0.940	0.942	0.942	0.941
Adjusted R^2	0.048	0.289	0.044	0.082	0.930	0.933	0.932	0.931

Note: *p<0.1; **p<0.05; ***p<0.01

Regressions (1) - (4) are simple regressions without city or year fixed effects. Regressions (5) - (8) include fixed effects for year and for city as a full set of dummies for each.

Table 2:

	Dependent variable: Price Index, $p_{c,t}$							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\log Y^{20}$	0.187				10.660***			
	(1.525)				(2.602)			
$\log Y^{90}$		5.675**				36.982***		
		(2.444)				(3.195)		
$\frac{Y^{80}}{Y^{20}}$			0.373*				1.883***	
			(0.206)				(0.467)	
$\frac{Y^{90}}{Y^{10}}$				0.094**				0.630***
				(0.047)				(0.114)
Constant	107.967***	45.734*	104.740***	105.541***	7.269	−308.074***	93.856***	92.334***
	(12.906)	(27.506)	(2.985)	(2.424)	(23.847)	(35.703)	(3.745)	(3.413)
City Fixed Effects	Yes	Yes	Yes	Yes	No	No	No	No
Year Fixed Effects	No	No	No	No	Yes	Yes	Yes	Yes
Observations	336	336	336	336	336	336	336	336
R^2	0.936	0.938	0.937	0.937	0.065	0.305	0.063	0.101
Adjusted R^2	0.929	0.930	0.930	0.930	0.027	0.277	0.025	0.065

Note: *p<0.1; **p<0.05; ***p<0.01

Regressions (1) - (4) include only city fixed effects while regressions (5) - (6) include only year fixed effects.

4.2 Nominal percentile income growth in sample

Before examining the relationships between price indexes and income at the local level, I decided to examine the 20th (Y^{20}) and 90th (Y^{90}) percentile income growth in the sample. I went back to the sample and ran the following regressions:

$$\log Y_i^{20} = \beta_0 + \theta_i + \beta_1 T_i + u_i \tag{5}$$

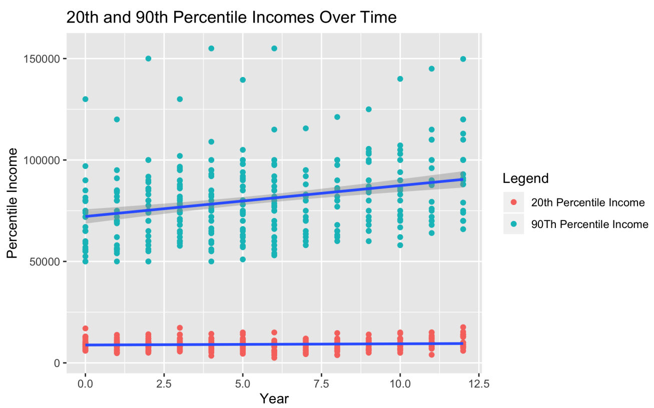
$$\log Y_i^{90} = \beta_0 + \theta_i + \beta_1 T_i + u_i \tag{6}$$

Where the dependent variables use the same notation as previous regressions in this paper, i is each observation (13 year $\times$ 38 cities = 336 observations), θ_i is a set of City fixed effects, where $\theta_i = \theta_c \ \forall \ i \in c$, and T_i is the numeric year index from 0-12. To build T_i, I subtracted 2005 (the initial year) from the numeric year variable of each observation in order to examine the effect of time passing on percentile income. I did not include City fixed effects in the first two regressions of Table 3, while the last two regressions include the city fixed effects in order to control for local variation. Figure one simply plots the percentile income data and the regression lines.

As shown in Table 3, without including city fixed effects, 20th percentile wage growth is insignificantly different from zero while 90th percentile income grows at about 2% per year. However, in these regression the R^2 value is relatively small– less than 10%. I then included city fixed effects which yielded much stronger results. 20th percentile incomes increase by about 1% per year while 90th percentile incomes rise at about 2% per year, on average($p < 0.01$).

Table 3:

	Dependent variable:			
	$\log(Y^{20})$	$\log(Y^{90})$	$\log(Y^{20})$	$\log(Y^{90})$
	(1)	(2)	(3)	(4)
Year	0.005	0.019***	0.009***	0.023***
	(0.005)	(0.003)	(0.003)	(0.001)
Constant	9.047***	11.166***	8.338***	11.046***
	(0.031)	(0.021)	(0.056)	(0.019)
City Fixed Effects	No	No	Yes	Yes
Observations	336	336	336	336
R^2	0.003	0.096	0.811	0.957
Adjusted R^2	0.0002	0.093	0.789	0.952
Note:	*p<0.1; **p<0.05; ***p<0.01			
	City fixed effects included in regressions (3) and (4)			



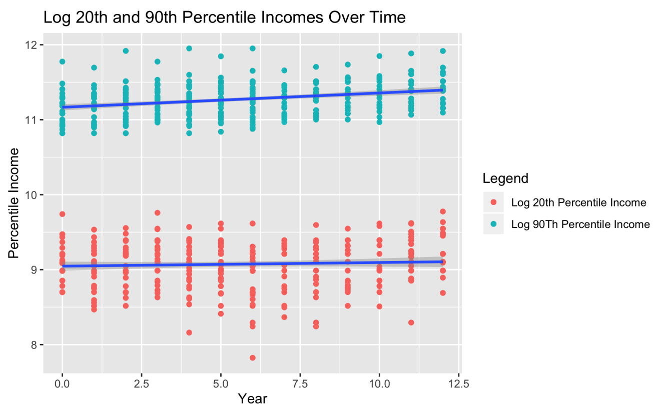


Figure 1: Scatter plots percentile incomes over the years of the sample. The blue lines represent thge regression equations, with a 2 standard error band around them.

4.3 Interpreting Table 1

I will mostly examine regressions (5)-(8) on Table 1. Log 20th percentile incomes (Y^{20}) inversely related with price levels with $p < 0.1$. As log 20th percentile incomes rise by 10%, price levels drop by about .3323 points. The remaining covariates, however, are positively and significantly related to price levels ($p < .01$). As log 90th percentile incomes (Y^{90}) rise by 10% there is a 3.5 point increase in price levels. This shows that price levels and high incomes have a much stronger relation than price levels do to low incomes. In this sample, improvements to high incomes drive prices up, whereas changes to lower incomes do not.

The remaining covariates ($\frac{Y^{80}}{Y^{20}}$ and $\frac{Y^{90}}{Y^{10}}$) serve to track the inequality in the cities in the sample. A higher ratio of incomes means there is more inequality. For example, if $\frac{Y^{80}}{Y^{20}} = 20$, then the 80th percentile income is 20 times the 20th percentile income for that observation.
Regression (7) on Table 1 shows a coefficient on $\frac{Y^{80}}{Y^{20}}$ of 0.888. Thus as the ratio of incomes increases by 1, the price index increases bt 0.89 points. This shows that more unequal cities have higher price indexes. Regression (8) supports this idea– for every increase in $\frac{Y^{90}}{Y^{10}}$, there is a 0.175 point increase in the price index.

5 Discussion and Conclusion

Over the last thirteen years, price levels have been on the rise. Nationally we have experienced low and stable price level increases (with national inflation statistics below 2%) and increases to nominal wages of workers. However, there has also been quite a bit of discussion on rising inequality in the United States. This paper adds to the literature regarding price levels and income inequality by inspecting this relationship at the city level. While there are many shortcomings in the data, including topcoding of IPUMS income data, I was able to utilize ratios of high income and low income percentiles as a measure of income inequality. I find that all of my measures of income and income inequality are positively related to local price indexes. However, high incomes rise much faster than lower incomes. Using the ratio of percentile incomes as a proxy for inequality, I find that as inequality at the local level is increasing over time, price levels are also increasing over time. Since all of this data is in nominal terms, this implies that *real* measures of inequality (adjusted for changes to price levels) is increasing at an even faster rate.

While I have not established causality, the results imply that there is in fact an empirical relationship between price indexes and inequality. However, if the trend continues, as the inequality amongst households at the city levels grows, so will the price levels. This can have many implications for future research and policy on inequality. As households make decisions to pursue educational attainment or homeownership, the cost of living reduces the available funds to lower income households to spend on homes and education. Additionally, as home prices and education prices rise, the higher earners at the city level will be able to afford much more due to increasing inequality and increasing price index, which could effectively push lower earners out of the housing and education market. Since homeownership and educational attainment are two significant ways to improve a household's economic outcomes, this could lead to further increases in inequality.

References

Atkin, David, Ben Faber, Thibault Fally, and Marco Gonzalez-Navarro. 2018. *A New Engel on the Gains from Trade.* Technical report. Working paper.

Attanasio, Orazio, Erik Hurst, and Luigi Pistaferri. 2012. *The evolution of income, consumption, and leisure inequality in the US, 1980-2010.* Technical report. National Bureau of Economic Research.

Balcilar, Mehmet, Shinhye Chang, Rangan Gupta, and Stephen M Miller. 2018. "The relationship between the inflation rate and inequality across US states: a semiparametric approach." *Quality & Quantity* 52 (5): 2413–2425.

Beetsma, Roel MWJ, and Frederick Van Der Ploeg. 1996. "Does inequality cause inflation?: The political economy of inflation, taxation and government debt." *Public Choice* 87 (1-2): 143–162.

Bulıř, Aleš. 2001. "Income inequality: does inflation matter?" *IMF Staff papers* 48 (1): 139–159.

Bureau of Labor Statistics. 2018. *The Consumer Price Index.* https://www.bls.gov/opub/hom/pdf/cpi-20180214.pdf. Accessed: 2019-11-10.

Handbury, Jessie, and David E Weinstein. 2014. "Goods prices and availability in cities." *The Review of Economic Studies* 82 (1): 258–296.

HuffPost. *Why Aren't Americans Moving Anymore.* https://www.huffpost.com/entry/why-arent-americans-moving-anymore_b_586fc32be4b0a5e600a78a23. Accessed: 2019-11-30.

Jaravel, Xavier. 2018. “The unequal gains from product innovations: Evidence from the us retail sector.” *The Quarterly Journal of Economics* 134 (2): 715–783.

Nordhaus, William D. 1998. “Quality change in price indexes.” *Journal of Economic Perspectives* 12 (1): 59–68.

Piketty, Thomas, and Emmanuel Saez. 2003. “Income inequality in the United States, 1913–1998.” *The Quarterly journal of economics* 118 (1): 1–41.

Raymond, J RING. 1985. “Variability of Inflation and Income Across Income Classes.” *Social Science Quarterly* 66 (1): 203.

Tipping, David Guy. 1970. “Price changes and income distribution.” *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 19 (1): 1–17.

6 Appendix

6.1 Cities in Sample

Charlotte NC
Chicago IL
Cincinnati OH
Cleveland OH
Denver CO
Des Moines IA
Detroit MI
El Paso TX
Fort Collins CO
Grand Rapids MI
Greensboro NC
Hartford CT
Houston TX
Jacksonville FL
Lansing MI
Las Vegas NV
Manchester NH
Memphis TN
Minneapolis MN
New Orleans LA
Austin TX
Oklahoma City OK
Albany NY
Bakersfield CA
Baltimore MD
Philadelphia PA
Phoenix AZ
Portland OR
Richmond VA
Salt Lake City UT
Seattle WA
South Bend IN
Spokane WA
Springfield MO
Stamford CT
Boise ID
Boston MA
Buffalo NY

6.2 R code for replicability

R Markdown