PRICE CONSIDERATIONS IN THE MEASUREMENT OF INEQUALITY

by

Joseph Drost

Bachelor of Arts, University of New Brunswick, 2022

A Report Submitted in Partial Fulfilment of the Requirements for the Degree of

Master of Arts

in the Graduate Academic Unit of Economics

Supervisor: Mehmet Dalkir, PhD, Economics

Examining Board: Luc Thériault, PhD, Economics, Chair
Anthony Myatt, PhD, Economics
Murshed Chowdhury, PhD, Economics

This report is accepted by the
Dean of Graduate Studies

THE UNIVERSITY OF NEW BRUNSWICK

October 2023

ABSTRACT

This paper offers a unique lens for understanding inequality by examining the impact of subnational price level variations at the level of Core-Based Statistical Areas (CBSA). It demonstrates that regional price level differentials provide a pertinent perspective for comprehending inequality. The paper shows that the use of incomes that have not been adjusted by a regional price deflator tend to upwardly bias inequality measurements such as the Gini Coefficient and Atkinson Index. It reveals that adjusting for purchasing power can lead to changes in the Gini coefficient, with a more pronounced effect in states marked by significant price level disparities. The paper also highlights that the Regional Price Parity (RPP) adjustment alters the composition of the lowest income bracket, particularly impacting rural residents. The findings suggest that although standard Gini and Atkinson Indices are upward biased when using unadjusted data, the effect is small enough to be ignored in most cases.

DEDICATION

For Guin.

ACKNOWLEDGEMENT

I would like to first acknowledge my supervisor Dr. Mehmet Dalkir for the time, effort, and insight he offered throughout the writing of this thesis. I would also like to thank my readers; Dr. Murshed Chowdhury and Dr. Tony Myatt, as well as Dr. Luc Thériault who chaired my defense.

I would like to recognize the University of New Brunswick Department of Economics for the high level of professionalism and training I received. I want to especially recognize the hard work of Jodi O'Neill. I want to thank my peers in my program for their assistance, support, and friendship throughout the last year.

I want to recognize those who helped fund this research, notably Dr. Elizabeth Parr-Johnston and the UNB Graduate Research Arts Assistantship.

I want to thank Dr. Yuri Yevdokimov for encouraging me to apply to this program, Dr. Herb Emery for answering all my questions, and Dr. Don Wright for building me up.

I would like to thank my parents for their constant love and support.

Finally, I would like to thank my wife, Guin. For her support, time spent formatting, coffee and snacks, and mostly for making me be my best.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

Income inequality has become a major topic of interest to policy makers and researchers interested in the distribution of wealth over the last twenty years. The major catalyst for this increased interest was the perceived unfairness in the wake of the global financial crisis and ensuing significant class tensions which lead to Occupy protests across the globe. This created an environment for Thomas Piketty's seminal work *Capital: In the 21st Century* in 2014 which is claimed to be the highest selling book ever published by an academic publisher. Piketty's 2001 and 2003 papers listed in the bibliography did a lot to bring the topic into academic discussion, but *Capital* intelligently brought the issue to a much wider audience. While economic inequality has been well-researched in the years since *Capital,* but the relation between regional price levels and cumulative inequality has been under researched.

Following a review of the relevant literature, the first part of this MA report will provide theoretical justification for a claim that standard measurements of income inequality are structurally biased and tend to show higher levels of inequality than exist. This is because most studies use data which does not adjust for regional price differences. Using only nominal price levels and omitting a price adjustment can upward bias because high nominal earners tend to live in high price areas and vice versa. By leaving out relative price levels, the disparity between individuals in these two groups is overstated. To ensure we are truly comparing the same thing across all people an income adjustment should be done based on price levels. This work is not done to minimize the issue, but rather to give a better sense of the demographics of those near the bottom of the income distribution as this has many implications particularly in terms of place-based policies and their ability

to affect households at the bottom of the distribution. The second part of this MA Report will use 2021 data from the US Current Population Survey to adjust incomes relative to regional purchasing power parities. This will provide empirical evidence as to whether income inequalities are appreciably higher before a regional price deflator (RPP) is applied. It will also allow us to study how the bottom decile is affected by the adjustment.

Using American data from the Annual Social and Economic Supplement of the 2021 Current Population Survey this report shows that Gini coefficients that measure real income via the Regional Purchasing Parity (published by the Bureau of Economic Analysis) are lower than the Gini's shown by the nominal incomes. The effect was small but significant in most states where RPPs differ within the state. The effect was most appreciable in states with large differences in price levels within the state. Alaska and Wyoming provided no results because they contained no Core Based Statistical Areas (CBSAs). Therefore, Lorenz curves of the nominal and real incomes were identical. Rhode Island and the District of Columbia were also dropped because their entire population live within a single RPP which creates the identical problem.

Results for states that have large CBSAs which cross state boundaries should be taken with additional caution as it was not obvious what portion of the CBSA lives within each state. For larger CBSAs this makes it difficult to attain what portion of the population of the state lives outside a CBSA altogether and what percentage live within that given CBSA. We attempted to control for this by taking the percentage within the sample that live within each state, but because the sample of people who live in a given CBSA is so small, these results do not carry sufficient statistical weight. Future studies should use county population data to find what population of the CBSA lives in each state. States

where this is the biggest problem are New Hampshire due to the Boston CBSA, Delaware due to the various CBSAs which cross borders, and New Jersey. The New York City and Philadelphia CBSA's extend into multiple states also which cause more mild problems in the analysis of New York, Pennsylvania, and Connecticut. Fortunately, this does not affect findings for national statistics but states in which a CBSA that crosses a border makes up a significant portion of their population should be viewed with caution.

The findings show a small but significant decrease in Gini Coefficients when the Lorenz curves are drawn for real incomes rather than nominal incomes for most states. This phenomenon also exists for the national statistics. Due to relatively small samples within the individual states, it is hard to tell in some cases, but there seems to be an indication that adjusting for purchasing power decreases the percentage of rural dwellers who fall in the bottom decile of earners by about 8% from 30.37% of the bottom ten percent of earners to 27.57%. This occurs despite only 24.83% of the sample living outside of a CBSA.

2.0 LITERATURE REVIEW

Capital notes that the growth rate of capital has exceeded the growth rate of labour over most of the last century. This has led to extreme capital accumulation as capital returns are higher than labour growth. Over a long period even a small difference can cause extreme accumulation to investors rather than workers. Piketty's proposed solution is a small global tax on capital to bring its growth rate nearer to that of labour. Piketty has succeeded in creating serious discussions about ethical consideration around inequality, the ideal level of inequality, and the potential solutions. The "ideal" level of inequality would be one which balances maintaining incentives for people to work and create economic growth, with the social interest in maintaining a reasonably equal society. Society should not be so unequal that extreme class differences cause too much unrest or other such disutility, but should be unequal enough that people feel there is substantial reward to creating economic growth.

Since the publication of *Capital* it has been common for NGOs and researchers to publish inequality reports on some geography of interest to them. For example, Oxfam publishes the global inequality report on an annual basis.

The level of geographical aggregation chosen to measure income inequality can have a significant influence on the results. For example, a large urban area may have significant inequality as unemployed people tend to relocate there in the hope of taking advantage of a thicker job market. However, at the level of the neighborhood there is significant sorting based on income resulting in lower inequality. People with high incomes tend to live in neighborhoods with others who have comparable incomes. This shows the importance of comparing reasonably similar geographical units. New York

City's Gini coefficient should not be compared to the nation as a whole. It is important to collect data for all geographical levels, but it is not especially instructive to compare across levels. Rather, states should be compared to other states and countries with other countries and so on. This remains the case after the price level adjustment.

This paper will look at inequality at the level of the state and of the nation. In this case it is acceptable to use different geographical units because it is not a comparison between units that is interesting, but rather a comparison before and after the price adjustment is made that is important. This paper compares the unit against itself rather than against other units. In this case the state is the smallest appropriate unit to analyze because it is a level within which there are observable price differentials. Except for the states mentioned earlier, every state has at least two RPP's and usually more. While there are only 51 states there are 260 unique CBSAs within the sample.

While much of the research on income inequality has taken place in France and the United States, there are researchers in Canada with an interest as well. Locke, May et. al. (2015) have published an inequality report for Newfoundland and Labrador as they advocate for certain changes in the taxation system in order to reduce these inequalities. And the Center for the Study of Living Standards included a section devoted to it in their report on New Brunswick Income Trends (2021).

One of the preeminent Canadian voices in this field is Michael Veall who has published several times on the top income shares in Canada. In 2003 Veall and Saez published "The Evolution of High Incomes in Canada, 1920-2000." This became the starting point for future revision and repetition of studying Canada's top income earners

using tax-filer data. Their data showed a consistent increase in top income shares from the 1970s to 2000. In 2003 they seemed alarmed that the top five percent of income earners had reached the prewar level. As of 2000 the top five percent had 29% of the income and the top one percent had 13.6%. Proponents of the use of income ratios prefer it to the Gini Coefficient because the Gini Coefficient is most sensitive to changes in the middle of the distribution. For researchers interested in the top of the distribution, this is less useful (Chu and Wang, 2021). In 2012 Veall extended this study to more recent data and analyzed reasons for the surge in top income shares in the mid to late 1980's. Veall suggests that much of this surge can be attributed to external factors such as corporate policies within the United States. They conject that as top incomes increased in the US, Canadian top incomes were forced to increase to compete for the same people. They show that Canadian top income shares could be modeled as a lagged response to American top income shares. Their second likely explanation relied on "skill biased technical change". The idea that improved communications technology allows a few high performing individuals to have an even larger impact and therefore be worth higher salaries.

While Veall has chosen to focus on income shares primarily in his research, other reports have focused on Gini coefficients to estimate inequality. In some cases, the Gini Coefficient is preferable to income ratios because it allows researchers the flexibility to get what can be considered a more full picture. Income ratios require the population to be separated into just two groups which may be useful when studying dynamics of poverty or of the very rich but misses out on a nuanced picture of the rest of the distribution. Gini coefficients give the area between the line of perfect equality and the Lorenz Curve (a

curve showing the share of income distributed at each percentile). Necessarily the Lorenz Curve runs from 0 to 1 as all the income in the economy is distributed by the 100th percentile. The Lorenz Curve approach creates difficulty in ranking inequality when the curves intersect and produce similar Gini Coefficients. Researchers have found ways of ranking Lorenz curves which intersect one (Chiu, 2007) or multiple times (Davies and Hoy, 1995; Aaberge, 2009) relying on the Pigou Dalton Principle which states that any transfer from richer to poorer should improve social welfare. The Pigou Dalton Principle is also known as aversion to downside inequality.

If two Lorenz Curves intersect each other, then a normative judgment must be made regarding whose income is more important. If the Lorenz Curves do not intersect it is always the higher of the two that is considered more equal because this indicates everyone except for the highest earner is better off which is no doubt more equal. However, in some cases Lorenz curves do intersect which can cause difficulty in ranking them. Addressing this, Fernando De Maio (2007) writes that "The Gini coefficient's main weakness as a measure of income distribution is that it is incapable of differentiating different kinds of inequalities. Lorenz curves may intersect, reflecting differing patterns of income distribution, but nevertheless resulting in very similar Gini coefficient values" (850). This has been addressed by other scholars by adding a social welfare function which can apply different weights along the income distribution. This is discussed further in section on the Atkinson Index. In this report, we will take the Gini Coefficient as a reasonably accurate ranking system for Lorenz curves. For robustness, the Atkinson Index is also used as an alternative measure.

This research report will take the Gini coefficient as a reasonably accurate measure of inequality despite its problems and will report Gini coefficients and provide bottom decile values of income as well. Aaberge's method could be used to rank the Lorenz curves that will be produced, but that will be beyond the scope of this project. Gini is also acceptable is it holds the desirable quality of adhering to the Pigou Dalton Principle as proven by Mehran (1976).This project will also be limited to adjusting the curves to reflect purchasing power inequality rather than nominal income inequality and seeing if this affects the spatial makeup of the lowest decile of earners. The expectation is that when incomes are controlled for price level, fewer individuals who make up the bottom decile will reside in low cost areas. Failing to make this adjustment should overrepresent low cost and rural dwellers in the bottom decile which would have significant policy implications for countries such as Canada which undertake significant regional economic support policies.

Some philosophers criticize the use of the Gini coefficient as well as the study of inequality entirely (Frankfurt, 1987). Their primary argument is based on the idea that if the lowest earners' incomes are increasing, then there should be no consideration to how they fare comparatively. However, if there is an assumption introduced that individuals tend to feel competitive regarding incomes and compare themselves relative to their peers, then it is necessary to study and understand inequality. The idea is that *relative income* matters. This is a very reasonable assumption as the competitive nature of the social system leaves most people comparing their situations to others. It naturally follows that this can be associated with a disutility when one does not stack up favourably. In this model individual welfares are a function of income and how far they are from the mean

income. Given the specification of this social welfare function, the socially optimal situation will be one where inequality is low, but not so low as to sacrifice too much income. In the socially optimal situation there is balance between incentive to earn, and equality.

Sheshinski demonstrates a social welfare function including relative income concerns and shows that the Gini coefficient can be substituted into it and follow natural theoretical intuitions. By writing the social welfare function instead as a function of individual incomes and the sum of all income differences, and reintroducing the Gini coefficient, the Gini's relevance is demonstrated. The example concludes with a social welfare function decreasing in the Gini Coefficient and increasing in average income. This shows mathematically that there is a social welfare equation which includes the Gini coefficient relying only on the additional assumption that there is disutility associated with having a lower income than another. The author leaves it open that other equations and necessary conditions should be found to expand the set of possible social welfare equations which include the Gini Coefficient. Shesinski's finding demonstrates that the further study of the Gini coefficient and its proper measurement continues to be worthwhile as it is likely to impact social welfare.

2.1 New Economic Geography

The new economic geography literature has provided strong theoretical reasons as to why the law of one price would not hold at the subnational level. Krugman (1971) suggests that differing price levels across regions can be traced back to increasing returns which cause an economy to roughly bisect (in numerical models) into a core periphery relationship. Any initial advantage of some region over another can make it more

profitable for manufacturing industries to locate in this area to reduce transportation costs associated with getting goods to market. This causes further concentration of people in the core region while the periphery becomes only populated by firms requiring large amounts of land unaffordable in the core where land rents are high due to the high demand for land. This creates a dynamic where land rents are high in the core, but the costs of manufactured goods are low due to the firm's proximity to the market. The periphery region has lower land rents but higher costs of manufactured goods. Food produced in the periphery are also theorized to have lower costs in the hinterlands, these are given the broad name "agricultural goods", but they need not necessarily be farmed goods.

Many of the findings discussed by Krugman are derived from the Dixit Stiglitz (1977) model of monopolistic competition of tradeable and non-tradeable goods. Krugman concludes that the core periphery equilibrium suggests *real* wages end up equalizing with perfect labour mobility. In a practical application it is often true that a country like the United States imports most of its manufactured goods as well as a large portion of its agricultural goods. So, while location may have some small influence on these goods' prices within the country, it is unlikely to influence the RPP in a meaningful way. Rather, much of the difference can probably be traced to non-tradeable goods such as land rents, construction services, and things of that nature. Items where arbitrage is impossible, create a possibility for different price levels in different regions of the same country.

In Krugman's landmark 1998 paper "Space: The Final Frontier" the Dixit Stiglitz model is generalized for looser assumptions, more regions, and differently shaped geographies. This includes "racetrack" (circular city) models. Krugman found that

numerical simulations suggest that even with a very small perturbation to an original uniform distribution, one or multiple cores will tend to emerge. He finds again that due to transportation costs and economies of scale prices of manufactured goods will be lower in the core regions while the agricultural goods will have lower prices in the periphery. The relationship between price and location is clearly more complicated, but these publications offer strong evidence refuting the law of one price.

The grandest attempt to compile as many different constraints as possible on the location theory with transportation costs and increasing returns is found in *The Spatial Economy* by Fujita, Venables, and Krugman. Relying on computer simulations and numerical examples they find many models which give similar results to the above. In this book they identify "centrifugal forces" which work towards agglomeration and "centripetal forces" working against agglomeration. Centrifugal forces include but are not limited to thick markets, knowledge spillovers from dense social networks, and things like social interaction which have value to residents. Centrifugal forces include increasing land rents, congestion, and capital immobility which all have increasing disutility to residents as the population grows. This further supports a rough core-periphery relationship leading to price differentials but rough equal real wages. These agglomeration economies in part determine land rents and non-tradeable prices.

There also exist empirical findings of differing price levels within the United States. For example, the US Bureau of Labor Statistics publishes a regional price parity at the level of metropolitan areas and states which show that price levels are different within the United States. Aten (2017) provides commentary on the construction of the measurement by the Bureau of Labor Statistics and uses the data to construct real wage

levels for each region. Aten finds that goods prices are relatively higher and rents relatively lower in predominantly rural states supports the core-periphery argument of Krugman. This is likely due to transportation costs and land demand. If so, it closely fits the predictions of the new economic geography literature.

Diamond and Moretti (2021) take a similar idea, that prices and wages differ across geography, and attempt to measure a standard of living using a regional price parity at the level of commuting zones. They find evidence that gains in income for college educated workers are roughly equated by price changes with movement from low-cost to high-cost commuting zones. However, for high school dropouts moving to less affordable commuting zones is associated with a decrease in consumption. This suggests the real wage equality holds for educated workers but not for unskilled workers who on average incur a real income decrease when moving to urban or otherwise high-cost areas. This is only conditional support for the prediction that real wages are similar across geography as it holds just for educated workers. It is possible that part of the reason that it does not hold for uneducated workers is that they will tend to migrate to urban areas in an attempt to take advantage of thicker job markets. While searching for better work their low income brings down the average income of uneducated people in the area and prevents them from having equal real wages. This would be because of unjustified exuberance for the working possibilities available to them in larger cities.

In Canada, the Regional Price Parity approach would be the ideal way to construct an adjustment for wages. However, Statistics Canada provides insufficient data to replicate this for Canadian regions. Rather they publish CPI's for each province and many Census Metropolitan Areas, but their data is insufficient to replicate this for most

Canadian regions. Emery and Guo (2019) note that "...in New Brunswick… Price data used to construct the [Consumer Price] index are primarily collected in Saint John, which leads us to ask how well the CPI for New Brunswick represents the cost of living for residents outside of Saint John" (4). They estimate CPI bias for a large number of subgroups in the Canadian population including rural/urban. This is done using the Engel curve, which operates on the assumption that as real incomes increase the share of income spent on a given necessity will decrease. Emery and Guo construct Engel curves for food expenditure under the assumption of constant tastes. They suggest that if an overall price level is higher for some group, that their income share spent on food will be lower for the same nominal income. They construct these Engel curves and estimate the CPI bias for each group. This method can be applied to any number of regions, but Emery and Guo limit it to each province, and to a single rural/urban bias for all of Canada.

Emery and Guo employ a method introduced originally by Hamilton in 2001. Hamilton uses the theory that food budget share is negatively related to real income to create a method to estimate CPI bias in the US from 1974-1991. Hamilton suggests that this may be a stronger method than price aggregation as it automatically accounts for substitution effects and the introduction of new products. CPI bias is measured by regressing food budget share on the difference between unadjusted food and non-food price levels, the difference between nominal wage and the total unadjusted price level, and a number of other regressors and fixed effects for years. These can produce estimates of bias for CPI for any other variable you may choose to add.

While this project was originally meant to be undertaken using Canadian data, a sufficient data source could not be obtained on a satisfactory timeline. This is why we

resorted to American price data in the form of RPPs. RPPs are gathered using price data in a similar manner to how CPIs are calculated. While this comes with the issues inherent to the CPI (the lack of substitution effects and new technologies) it nonetheless shows a relative difference between urban and rural areas which can be used to attain real incomes. There is no obvious reason that these inherent limitations will over or undershoot the relative price levels of different regions as the issues are inherent to all price levels equally. Typical arguments about the CPI are expected to be of no concern to this study for that reason.

Geloso and Msaid (2018) use similar estimation methods to the ones used in this report but focus only on regional makeup of the bottom and top deciles of the income distribution before and after RPP adjustment. They use this income deflation to estimate how each state is represented in the top and bottom ten percent relative to their overall share of the national population. Doing this they find that urban states such as New York and California are overrepresented in the top 10%. Rural states where price levels are lower are overrepresented in the bottom decile of the income distribution. The present MA Report contains a few useful extensions that justify the undertaking of this project. First, there is nothing said about the level of inequality within states. While some data is admittedly poor, there are many states where meaningful estimates can be made. The estimation of representation for each state will not be recreated as the findings from Geloso and Msaid are recent enough to maintain relevance. Rather, the contribution of this paper will be to take this a step further and estimate the representation of rural dwellers in the top and bottom portions of the distribution for each state. In this case, a rural person will be anyone that does not reside within a CBSA. All people living outside a CBSA

reside in places of smaller than 100,000 people and an area in which less than 25% of the population travels to a central county for work. Ostensibly this means that the majority of this person's purchasing and earnings occur outside of a metropolitan area.

3.0 THEORETICAL FRAMEWORK

This paper will use the theories of the new economic geographies and the measurement of Gini coefficients and Atkinson Indices before and after adjusting nominal wages for real purchasing power to show that there is a small but significant reduction in the level of inequality when real wages are used. The changes are small enough to maintain the legitimacy of the Gini and Atkinson measures while highlighting that in some cases states with significant price differences may have results that are upwardly biased. This will improve the understanding of the nature of inequality within subnational units and in a nation as a whole. It will also show the rural/urban makeup of those in the bottom decile of the income distribution. By deflating using RPPs, real incomes will increase in low price level places and decrease incomes in high price areas. It is expected that this will decrease the overall level of inequality as low-income areas should also be low price areas and vice versa. This will be empirically tested by estimating if there is a statistically significant change in the Gini coefficients when the deflator is added.

The new economic geography literature has conjected that high income areas also tend to be high price areas. The law of one price does not hold at a subnational level for at least three reasons; transportation costs, land rents, and the existence of non-tradeable goods. Land rents and non-tradeable goods are very similar, but it is nonetheless useful to think about them independently as rent and housing is more rival than other non-tradeable goods making it more sensitive to population effects than income effects as competition plays a bigger role for limited housing costs than goods for which quantity provided is more elastic.

Transportation costs mean that the total costs to the producer of providing a good near the source of its manufacturing is lower than to incur the transportation costs associated with shipping it to further places. Price level is positively linked with total costs of production, so it follows that the price of a good will be lower nearer to the factory and higher elsewhere. The firm makes the same profit selling for a price that is lower by exactly the transportation cost. This is not necessarily maximizing but nonetheless illustrates that there is incentive for a firm to prioritize sales to customers nearby. For this reason, we can expect that there will be a lower price level for manufactured goods in manufacturing regions and a higher price for manufactured goods away from the manufacturing region. Due to the interconnected nature of the modern global economy and the rate at which the United States imports manufactured goods, this effect is probably of minor significance. The majority of goods produced will still be imported internationally into any city within the United States as evidenced by its negative trade balance in manufactured goods. This means that goods prices of tradeable are still determined in global markets with probably only very slight variation between regions within the United States.

Housing is a significant portion of most Americans budget. Between 2017 and 2021 the Census Bureau estimated that 31% of Americans and half of all renters spent 30% or more of their income on housing. As noted in the new economic geography literature previously cited, as density increases in core regions land rents increase due to high demand. This will affect the overall price level as housing is a large budget item for most households. It would be natural that this will have a larger effect on the price level than the transportation costs as housing is a necessary good that takes up a large

portion of income and is not tradeable so there is a higher degree of competition for a small, relatively inelastic housing stock. To a certain extent there may also be a reverse causation effect wherein workers demand higher salaries to compensate for high land rents which creates an upward spiral of housing costs which are limited at the point where it would be more efficient for metropolitan firms to locate elsewhere with lower wages even if other costs increase somewhat. To that extent the centrifugal force of cheaper labour elsewhere becomes a limiting factor to wage spiral.

Non tradeable goods and housing may be the primary reason that price levels differ between areas. Housing, despite being the largest budget allocation for most people, is not the only non tradeable good. Other popular examples include things like haircuts, where no one would drive an hour for a haircut, and construction services, where it is very expensive and inefficient to transport labour and construction resources any significant distance. Because these goods are not exportable and must be acquired near a person's place of residence, this removes opportunity for arbitrage. Therefor these prices are not determined on global or national markets, but rather in a market that is only related to the distance people are willing to commute. In these markets which only trade locally, prices are determined only by the incomes and tastes of *local* consumers. For this reason, in places where nominal incomes are higher the firm can efficiently increase the price they charge for their good or service. This leads to a higher overall price level as the price level is some mix of the price of tradables and the price of non-tradables. This shows that in places where incomes are higher, perhaps in a city with a large banking and professional demographic, then the price level will also be higher. With only theory it is not easy to show whether income gains are fully offset by price

levels of non-tradeable and housing, but it becomes obvious that their income is overstated if not adjusted. Recall that Diamond and Moretti showed that gains in income by moving areas are fully lost by college educated people and more than fully lost by non-educated people due to overall price differentials.

4.0 DATA & METHODS

4.1 Income Data

Data for this project has been taken from the 2021 Annual Social and Economic Supplement to the Current Population Survey. This data is sourced from the Census Bureau and linked with RPP values taken from the Bureau of Economic Analysis. The entire dataset included 90,759 observations but the analysis was restricted to just 61,834 as all observations where total household nominal income was reported as less than $100 were dropped. This is justified because from all sources, including transfers, employment, and other forms of income, there is no conceivable way that so many people's annual income was less than $100. It seems that a great many people were uncomfortable answering questions about their annual income as in the breakdown of income into categories, 27, 909 observations have no response. Dropping all nominal incomes less than $100 loses about this same number of observations so is a reasonable cutoff for inclusion.

The primary variable of interest within this dataset is the household total income. This is provided as a continuous variable and, after unrealistic values were dropped, ranges from $100 to nearly $3 million in annual income. Because the CPS is a limited sample, it is likely that it misses out on many of the highest earners which are the source of much of the inequality. It is not necessary to correct for these super rich values for two reasons. First, it is unlikely that regional price differences are of concern to the top percentile as their consumption coefficient is on average quite small, so it is not reasonable to be used to deflate their entire income. Secondly, for the top percentile it is likely that a smaller percentage of their spending is done within their own CBSA. These

households earn several hundreds of thousands of dollars annually. A large portion of that is saved, and another large portion of that is spent outside of their city. This makes it unreasonable to deflate their entire income by the RPP. Alternatively, almost all the bottom decile's spending is done within their immediate area, so deflation by the RPP gives a good sense of their true purchasing power. The following Tables include the summary statistics for both the nominal and the real incomes.

Table 4.1.1: Income Summary Statistics

Household Nominal Income			Household Real Income		
Percentiles		Observations	Percentiles		Observations
1%	$3,019	61,834	1%	$3,185	61,685
5%	$11,372	Mean	5%	$11,690	Mean
10%	$17,335	$98,736	10%	$17,881	$99,602
25%	$35,000	Std. Dev.	25%	$35,768	Std. Dev.
50%	$69,708	116,382	50%	$70,688	116,195
75%	$124,451	Skewness	75%	$126,351	Skewness
90%	$201,227	6.11	90%	$202,372	6.52
95%	$272,323	Kurtosis	95%	$269,374	Kurtosis
99%	$508,820	78.83	99%	$498,788	93.22

The summary statistics of the two variables are very similar in their percentile ranges as well as their means and standard deviations. One important way in which they

differ is in their kurtosis. The distribution of the real income is shown to have a sharper peak in the middle. This is consistent with the idea behind this paper, which is that when we only compare nominal incomes, we are not really comparing the same thing across groups due to price effects. This higher kurtosis may indicate that we went from treating many distributions with different means as one, to truly having a single distribution. This represents an improvement in the data and provides support for making the income adjustment before doing analysis.

4.2 RPP Data

RPP data is provided for CBSA's and for states. The RPP is an index where 100.00 would be the average overall price level across the entire United States. Then 105.00 would be 5% higher prices than average and 95.00 would be 5% lower prices. The RPP is constructed using interview data much the same way CPI data is collected. While there are RPPs for metropolitan dwellers and for dwellers of the entire state, this leaves out the price level of residents living outside of metropolitan areas. To recover this data, it was assumed that the state's RPP can be represented by the following formula where "sj" represents state j, "rj" represents the rural component of that state, and "mi" metropolitan area i.

$$RPP_{sj} = RPP_{rj} * \frac{Pop_{rj}}{Pop_{sj}} + \sum RPP_{mi} * \frac{Pop_{mi}}{Pop_{sj}}$$

That is, it was assumed that the states RPP was the weighted average of all its metropolitan areas and its rural component. To recover the rural component the formula is rearranged to the following.

$$RPP_{rj} = RPP_{sj}\frac{Pop_{sj}}{Pop_{rj}} - \sum RPP_{mi} * \frac{Pop_{mi}}{Pop_{sj}}$$

In most cases where every CBSA was fully contained within a single state this worked directly to provide an estimate of the price level in the rural component of each state where one exists. The limitation of this is that American statistics share the flaw commented on by Emery and Guo in the Canadian CPI statistics. That is, they interview primarily people within metropolitan areas and provide no weighting, so the urban price levels are overrepresented. Because of this, the RPP recovered for the rural component will be systematically too high and the change in Gini coefficients due to price levels will be lower than the actual difference. It would also mean the proportion of rural dwellers that make up the bottom decile of the real income distribution would still be too high. In the results it is shown that while the price adjustment does cause the proportion of the bottom decile who live in rural areas to become nearer to the overall proportion of rural dwellers in the population, rural dwellers still tend to be overrepresented at the bottom of the distribution.

Aten (2016) notes that the RPP calculation use, in part, the BLS CPI price survey which "was not designed for place-to-place comparison and does not fully represent smaller geographical units (1)." Because the smaller geographical units and rural areas tend to have lower RPPs, this means that the effects in the results section are probably under reported.

4.3 Methods

Following the completion of these data cleaning and recovery techniques the estimation was straightforward. Using the Stata command "fastgini" in the package of the same name, Gini coefficients were estimated for each state and as a national total. The national total in particular may not necessarily align with other published Gini's due to the fact that the state proportions in the sample were not the same as the true population shares. Rather, smaller states will be overrepresented in this sample. The effect of this is ambiguous as there is not necessarily a link between a state having a smaller population and having lower incomes. However, it would be more likely that the small states have lower incomes as well as they tend to be primarily rural. Think of states such as Wyoming, Alaska, and Montana. This will in no way prevent from testing if Gini is sensitive to the inclusion of RPPs.

The command "fastgini" was preferred over others as it returned Gini coefficients as well as "jack knife" standard errors which can be used to perform a T-Test on the Gini estimates. Jack knifing is the process of calculating the gini coefficients of all subsets of (n-1) observations. Using these n Gini coefficients, a standard error is estimated in the regular way. The primary advantage that jackknifing has over bootstrapping is that it is reproducible. Jackknife results are fully determined while bootstrapping results are random due to the random subsamples drawn from the population. However, it has been shown by Efron (1982) that jackknife estimates of variance and standard error are upward biased. This means that had we opted to use bootstrapping instead we may have not only found significant results in more states. This

problem does not affect the Gini outcomes, only the standard errors so it is not a cause for major concern.

Outliers in this study were found to have a large effect due to the fact that the state samples are much smaller than the national sample. This means that outliers could dramatically influence the standard errors of the estimates, a problem that would be inherent to bootstrap or jackknife standard errors. These outliers caused issues in smaller states. For example, the largest listed income in the data (nearly $3 million) resides in Montana. Montana also has the second largest standard error when the outliers are included. With incomes over 1 or 2 million excluded, it ranks quite average. It was decided to leave the high values in.

5.0 ESTIMATION & RESULTS

5.1 Gini Coefficients:

Results mainly supported the hypothesis that Gini coefficients would systematically decrease if adjusted for real purchasing power. The below table shows Gini coefficients for nominal and adjusted real income levels as well as their jackknifed standard errors and the T Value for the test of the two Gini coefficients being equal. Nominal Gini coefficients ranged from .4113 (Wyoming) to .5368 (Washington D.C.). However, Washington DC presents a unique case as it is not truly a comparable geographical unit to the states. Cities tend to have higher inequality than states. The highest Gini found for a state was .5165 (Michigan). The range of real Gini coefficients was .4111 (Hawaii) to .5368 (DC) or .5159 (Michigan)

Table 5.1.1 Gini Coefficients of Nominal and Real Incomes

State	Gini of Nominal Incomes	SE of Nominal Gini	Gini of Real Incomes	SE of Real Gini	Difference in Gini	State Variation in RPP
AL	0.4921	0.0166	0.4905	0.0134	-0.0016	6.67
AR	0.4905	0.0158	0.4876	0.0157	-0.0029	6.46
AZ	0.4674	0.0130	0.4701	.0131	0.0027	9.64
CA	0.4715	0.0051	0.4679	0.0050	-0.0036	21.25
CO	0.4617	0.0142	0.4568	0.0139	-0.0049	14.49
CT	0.4803	0.0134	0.4782	0.0164	-0.0021	5.80
DE	0.4257	0.0135	0.4253	0.0138	-0.0004	5.25
FL	0.4757	0.0076	0.4749	0.0076	-0.0008	15.97
GA	0.4942	0.0128	0.4898	0.0126	-0.0044	11.71
HI	0.4118	0.0100	0.4111	.01016	-0.0007	6.28
IA	0.4357	0.0201	0.4349	0.0206	-0.0008	6.88
ID	0.4207	0.0113	0.4190	0.0112	-0.0017	4.46
IL	0.4894	0.0105	0.4865	0.0106	-0.0029	18.63

IN	0.4747	0.0173	0.4719	0.0171	-0.0028	18.63
KS	0.4376	0.0120	0.4353	0.0118	-0.0023	5.17
KY	0.4718	0.0127	0.4670	0.0126	-0.0048	8.48
LA	0.4913	0.0133	0.4906	0.0135	-0.0007	8.01
MA	0.4918	0.0118	0.4887	0.0117	-0.0031	11.90
MD	0.4789	0.0133	0.4767	0.0133	-0.0022	17.38
ME	0.4452	0.0188	0.4407	0.0184	-0.0045	8.32
MI	0.5165	0.0158	0.5159	0.0160	-0.0006	10.59
MN	0.48	0.0156	0.4720	0.0152	-0.0080	13.57
MO	0.4696	0.0209	0.4668	0.0203	-0.0028	7.82
MS	0.5017	0.0165	0.5003	0.0161	-0.0014	5.67
NC	0.4703	0.0086	0.4666	0.0085	-0.0037	8.33
ND	0.4449	0.0122	0.4447	0.0121	-0.0002	2.19
NE	0.4542	0.0164	0.4527	0.0166	-0.0015	3.48
NH	0.4229	0.0119	0.4221	0.0118	-0.0008	13.18
NJ	0.4756	0.0123	0.4736	0.0121	-0.0020	17.74
NM	0.4554	0.0111	0.4531	0.0109	-0.0023	7.25
NY	0.4884	0.0070	0.4862	0.0069 4	-0.0022	21.81
OH	0.4665	0.0104	0.4613	0.0100	-0.0052	7.62
OK	0.4673	0.0122	0.4671	0.0122	-0.0002	6.03
OR	0.4501	0.0131	0.4475	0.0130	-0.0026	6.32
PA	0.487	0.0122	0.4852	0.0123	-0.0018	12.61
SC	0.4772	0.0144	0.4738	0.0142	-0.0034	18.83
SD	0.4249	0.0164	0.4241	0.0166	-0.0008	4.22
TN	0.4729	0.0113	0.4708	0.0113	-0.0021	9.65
TX	0.4869	0.0113	0.4841	0.0086	-0.0028	16.09
UT	0.4151	0.0138	0.4142	0.0138	-0.0009	6.21
VA	0.4687	0.0109	0.4639	0.0110	-0.0048	19.15
VT	0.4973	0.0275	0.4961	0.0278	-0.0012	6.13
WA	0.4345	0.0088	0.4282	0.0087	-0.0063	15.86
WI	0.4422	0.1402	0.4392	0.0138	-0.0030	16.35

To summarize Figure 2, 32 of the 46 relevant states have significant findings, as does the national coefficient. In these 32 states the adjustment for prices decreases the level of inequality in state as calculated by the Gini coefficient. This supports the theory that places with higher incomes also have higher prices, showing that Gini coefficients are systematically too high. However, this effect is very small for most states and does not seriously affect its national rank or really the understanding of the severity of the issue. However, for a select few states, the problem may be misunderstood in a way worth. For this reason, Figure 3 includes the states rank when calculated nominally and its rank when calculated with real wages. Figure 3 also shows the change in Gini estimates between the two calculations.

There are also a number of states with questionable results due to the CBSA crossover issue discussed earlier. New Jersey does not have a large CBSA of its own but falls into CBSAs for both Philadelphia and New York City. 15 New Jersey respondents also did not list an MSA, however this is impossible. This caused the program to create for them an unrealistic RPP, these observations were dropped for consistency. This is better than giving them the state average because then we can get no effect of their location within the state, which is the whole point of these operations.

The largest change in Gini coefficients due to the adjustment occurred to Minnesota while the smallest of all states with multiple RPPs was Oklahoma. The ordinal ranking of inequality also remained pretty similar before and after the adjustment, however there were some notable movers. Minnesota moved several

positions due to its relatively large RPP adjustment. Minnesota also showed large effects because close to one half of its population live in a single CBSA while the remainder live relatively rurally, leading to significant variation within the price variable for this state. Important to supporting the theory is the fact that inequality in California moved 5 positions down when adjusting for RPPs. This is likely because California has some of the most pronounced high RPPs in the entire country. Seven of the eleven highest RPPs in the country belong to California cities including the 2 highest. These seven CBSAs represent about 16.7 million people. California RPPs ranges from as low as 96.7 (El Centro) to as high as 119.8 (San Francisco-Oakland-Berkeley).

Intuitively, results were smaller and often insignificant in states where there was very little variation in RPP across the state. Montana, for example, only varied 1.12 points across its RPP zones. South Dakota only varied 6.2 points and was insignificant, and Idaho only varied 5.1 points and was insignificant. In states such as these the adjustment is not relevant because most people in the state tend to experience similar price levels. This can be pinpointed as the primary reason that several states did not produce significant results. It is not necessarily that the effect does not exist there, but rather that is too small to be perceived on a sample of this size. In states with moderate to large variation, in almost all cases there were significant results. Arizona remains an anomaly where inequality increased following the adjustment. An explanation for this is not readily available.

Table 5.1.2: State Inequality Rankings and Change due to Adjustment

State	Nominal Rank	Real Rank	Rank Change	Gini Change
DC	1	1	0	0
MI	2	2	0	-0.0006
MS	3	3	0	-0.0014
VT	4	4	0	-0.0012
MA	5	7	2	-0.0031
AL	6	6	0	-0.0016
LA	7	8	1	-0.0007
GA	8	5	-3	-0.0044
AR	9	9	0	-0.0029
IL	10	10	0	-0.0029
NY	11	11	0	-0.0022
PA	12	12	0	-0.0018
TX	13	13	0	-0.0028
CT	14	14	0	-0.0021
NJ	15	20	5	-0.002
MD	16	16	0	-0.0022
MN	17	15	-2	-0.0028
FL	18	19	1	-0.0008
WV	19	17	-2	0.0000
SC	20	18	-2	-0.0034
IN	21	21	0	-0.0028
KY	22	23	1	-0.0048
MO	23	27	4	-0.0048
MT	24	25	1	0.0000
TN	25	22	-3	-0.0021
AZ	26	29	3	0.0027
VA	27	28	1	-0.0048
OK	28	30	2	-0.0002
CA	29	24	-5	-0.0036
NC	30	26	-4	-0.0037
OH	31	31	0	-0.0052
CO	32	32	0	-0.0049
NV	33	33	0	0.0000
NM	34	34	0	-0.0023
NE	35	35	0	-0.0015
RI	36	37	1	0.0000
OR	37	36	-1	-0.0026

AK	38	40	2	0.0000
ME	39	38	-1	-0.0045
ND	40	39	-1	-0.0002
WI	41	41	0	-0.0030
KS	42	42	0	-0.0023
IA	43	43	0	-0.0008
WA	44	44	0	-0.0063
DE	45	45	0	-0.0004
SD	46	46	0	-0.0008
NH	47	47	0	-0.0008
ID	48	48	0	-0.0017
UT	49	49	0	-0.0009
WY	50	51	1	0.0000
HI	51	50	-1	-0.0007

Figure 5.1.1: RPP Variation and Change

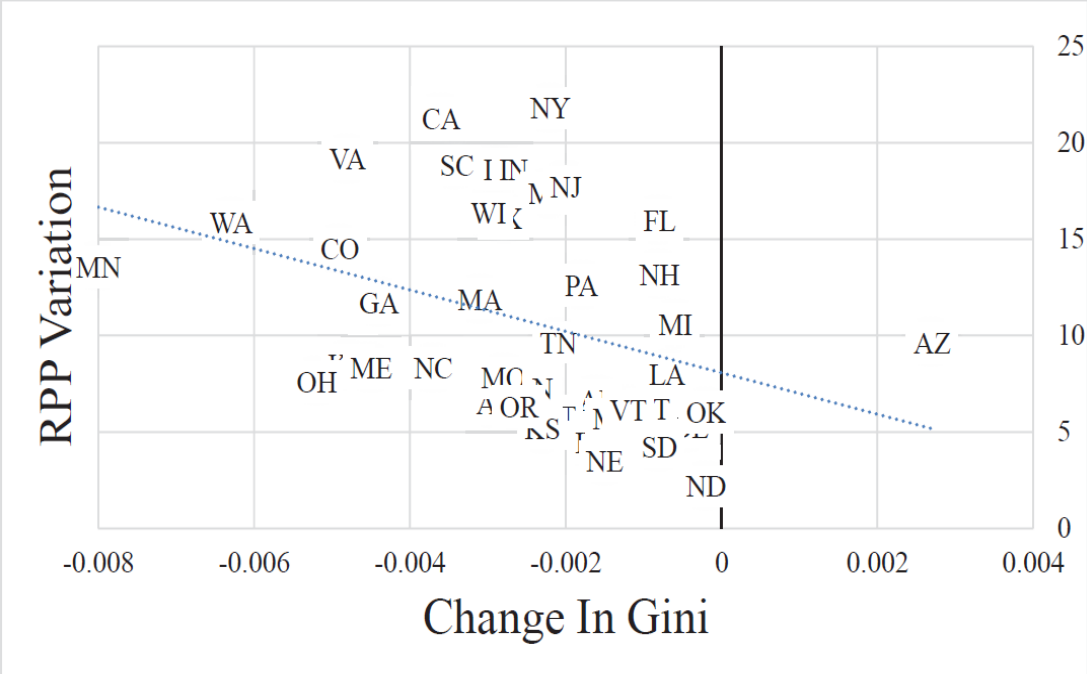

Table 5.1.3: Regression Table for RPP Variation and Change in Gini Coefficient

Regression Statistics	
Multiple R	0.3695
R Square	0.1365
Adjusted R Square	0.1155
Standard Error	0.0018
Observations	43

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	0.0000	2.0795E-05	6.4820	0.0148
Residual	41	0.0001	3.2081E-06		
Total	42	0.0002			

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	-0.0010	0.0006	-1.6771	0.1011
RPP Variation	-0.0001	0.0001	-2.5460	0.0148

Table 5.1.4: RPP Ranges

	Maximum	**Minimum**	**Difference**
USA	119.83	80.31	39.52
AK	104.44	104.44	0.00
AL	92.09	85.42	6.67
AR	93.30	86.85	6.46
AZ	99.18	89.54	9.64
CA	119.83	98.58	21.25
CO	109.16	94.67	14.49
CT	105.80	100.00	5.80
DC	111.34	111.34	0.00
DE	99.21	93.96	5.25
FL	109.92	93.96	15.97

GA	99.12	87.40	11.71
HI	114.74	108.46	6.28
IA	93.64	86.79	6.85
ID	93.85	89.39	4.46
IL	105.42	86.79	18.63
IN	105.42	87.03	18.39
KS	93.66	88.49	5.17
KY	94.42	85.87	8.55
LA	94.42	86.41	8.01
MA	109.69	97.79	11.90
MD	111.34	93.96	17.38
ME	101.74	96.42	5.32
MI	99.75	89.16	10.59
MN	103.60	90.02	13.58
MO	96.25	88.42	7.82
MS	90.62	84.95	5.67
MT	92.50	91.38	1.12
NC	97.58	89.25	8.33
ND	92.58	90.38	2.19
NE	93.64	90.15	3.48
NH	109.69	96.51	13.18
NJ	114.58	96.84	17.74
NM	93.39	86.14	7.25
NV	95.54	95.54	0.00
NY	114.58	92.77	21.81
OH	95.02	87.41	7.62
OK	92.76	86.73	6.03
OR	105.37	99.05	6.32
PA	99.21	86.56	12.65
RI	101.20	101.20	0.00
SC	99.15	80.31	18.83
SD	93.04	88.82	4.22
TN	95.53	85.88	9.65
TX	103.85	87.76	16.10
UT	96.23	8.02	88.21
VA	111.34	92.19	19.15
VT	102.64	96.51	6.13
WA	114.57	98.71	15.86
WI	105.42	89.07	16.35
WV	95.55	88.54	7.01

WY	91.42	91.42	0.00

5.2 Atkinson Index

In 1972, Atkinson provided a new measure of inequality which avoided the ambiguity associated with ranking Lorenz curves. He argued that the Gini Coefficient was ineffective in cases in which Lorenz curves intersected. To avoid this problem, Atkinson specified a form for the utility curves. The form he found that most closely fit the expected characteristics of individual utility and social benefit was the iso-elastic function. Atkinson draws from the concept of the "risk premium" to claim that there is an equally distributed level of income which if everyone earned would provide the same social benefit as the existing current distribution. Using this Atkinson creates a measurement defined by

$$I = 1 - \frac{y_{EDE}}{\bar{y}}$$

Where y_{EDE} is the equally distributed equivalent and $\bar{y}$ is the mean income. Of course, this equally distributed equivalent cannot be directly observed. For that we must choose some form of the utility curve. In general, Atkinson's inequality measurement (denoted *I)* can be written as follows.

$$I = 1 - \left[\sum_i \left(\frac{y_i}{\bar{y}}\right)^{1-\epsilon} f(y_i)\right]^{\frac{1}{1-\epsilon}}$$

Where ϵ is the societal aversion to inequality. For robustness it is necessary to try a few different values of ϵ. In the calculations used for this paper, ϵ of 1.5 and 2.0 have been used. These are commonly used in the literature so this follows standard convention. When the iso-elastic form is assumed, *I* takes the following form.

$$I = 1 - \frac{1}{\bar{y}}\left(\frac{1}{n}\sum_{i=1}^{n} y_i^{1-\epsilon}\right)^{\frac{1}{1-\epsilon}}$$

This form was used to calculate the Atkinson inequality for the United States before and after the income adjustment. *I* gives the inverse of the proportion of the mean income which is the equally distributed equivalent. For example if some sample had an *I* of .2, this would mean that the same utility would only require that everyone have 80% of the mean income. For any *I*, there is an EDE equal to (1 – *I)* proportion of the current mean income.

Table 5.2.1 Atkinson Index by State and Inequality Aversion Level

	Nominal *I*, 1.5	**Real *I*, 1.5**	***I* Difference, 1.5**	**Nominal *I*, 2**	**Real *I*, 2**	**I Difference, 2.0**	**RPP Difference**
AL	0.6261	0.6240	-0.0021	0.8568	0.8554	-0.0014	6.67
AR	0.5187	0.5158	-0.0029	0.6670	0.6649	-0.0021	6.46
AZ	0.4681	0.4730	0.0049	0.6109	0.6179	0.0070	9.64
CA	0.5334	0.5304	-0.0030	0.7409	0.7424	0.0015	21.25
CO	0.5340	0.5272	-0.0068	0.7347	0.7250	-0.0097	14.49
CT	0.5121	0.5097	-0.0024	0.6390	0.6370	-0.0021	5.80
DE	0.4210	0.4209	-0.0001	0.5424	0.5428	0.0004	5.25
FL	0.5425	0.5422	-0.0002	0.7694	0.7693	-0.0001	15.97
GA	0.5785	0.5725	-0.0060	0.8056	0.8026	-0.0031	11.71
HI	0.4624	0.4608	-0.0016	0.6922	0.6919	-0.0003	6.28
IA	0.4584	0.4578	-0.0006	0.6415	0.6392	-0.0023	6.88
ID	0.4433	0.4412	-0.0021	0.6351	0.6342	-0.0009	4.46
IL	0.5412	0.5372	-0.0040	0.7054	0.7016	-0.0038	18.63

IN	0.4985	0.4976	-0.0009	0.6480	0.6515	0.0035	18.63
KS	0.4488	0.4459	-0.0029	0.5827	0.5796	-0.0030	5.17
KY	0.5468	0.5420	-0.0048	0.7992	0.7968	-0.0024	8.48
LA	0.5910	0.5896	-0.0014	0.8328	0.8309	-0.0019	8.01
MA	0.5625	0.5584	-0.0041	0.7661	0.7616	-0.0045	11.90
MD	0.5239	0.5210	-0.0029	0.6586	0.6559	-0.0027	17.38
ME	0.4277	0.4213	-0.0064	0.5364	0.5292	-0.0072	8.32
MI	0.5693	0.5686	-0.0007	0.7537	0.7540	0.0003	10.59
MN	0.5124	0.5006	-0.0117	0.6729	0.6596	-0.0133	13.58
MO	0.5448	0.5437	-0.0010	0.8036	0.8059	0.0023	7.82
MS	0.5757	0.5739	-0.0019	0.7891	0.7872	-0.0018	5.67
NC	0.5368	0.5329	-0.0039	0.7472	0.7462	-0.0011	8.33
ND	0.4873	0.4871	-0.0002	0.6775	0.6772	-0.0003	2.19
NE	0.4892	0.4879	-0.0013	0.6584	0.6570	-0.0014	3.48
NH	0.4785	0.4750	-0.0035	0.6955	0.6888	-0.0068	13.18
NJ	0.5290	0.5284	-0.0007	0.7253	0.7254	0.0001	17.74
NM	0.4746	0.4715	-0.0031	0.6320	0.6291	-0.0029	7.25
NY	0.5641	0.5634	-0.0007	0.7701	0.7700	-0.0001	21.81
OK	0.5221	0.5211	-0.0010	0.7207	0.7207	0.0000	6.03
OR	0.4628	0.4597	-0.0031	0.6119	0.6085	-0.0034	6.32
PA	0.5336	0.5318	-0.0017	0.7125	0.7118	-0.0006	12.61
SC	0.5500	0.5449	-0.0052	0.7549	0.7520	-0.0029	18.83
SD	0.4667	0.4656	-0.0011	0.6430	0.6419	-0.0011	4.22
TN	0.5026	0.4998	-0.0028	0.7002	0.6985	-0.0017	9.65
TX	0.5444	0.5407	-0.0036	0.7373	0.7355	-0.0017	16.09
UT	0.4245	0.4238	-0.0007	0.5983	0.5990	0.0007	6.21
VA	0.5428	0.5371	-0.0057	0.7672	0.7620	-0.0052	19.15
VT	0.5375	0.5360	-0.0015	0.7041	0.7041	0.0000	6.13
WA	0.4732	0.4664	-0.0069	0.6436	0.6373	-0.0063	15.86
WI	0.4525	0.4500	-0.0025	0.6093	0.6092	0.0000	16.35

This table demonstrates very little difference between *I* before and after the income adjustment. Differences due to the income adjustment display a similar pattern as with the Gini Index for and Epsilon of 1.5 but shrink quite a lot for an epsilon of 2. This further supports the hypothesis that price considerations can lead to a modest but significant

change in income inequality statistics. Largest changes were again found in states with significant urban rural divides such as Minnesota and Washington while most states hovered between a 0 and 1% change due to the price considerations with and epsilon of 1.5.

Figure 5.2.1: RPP Variation and Change

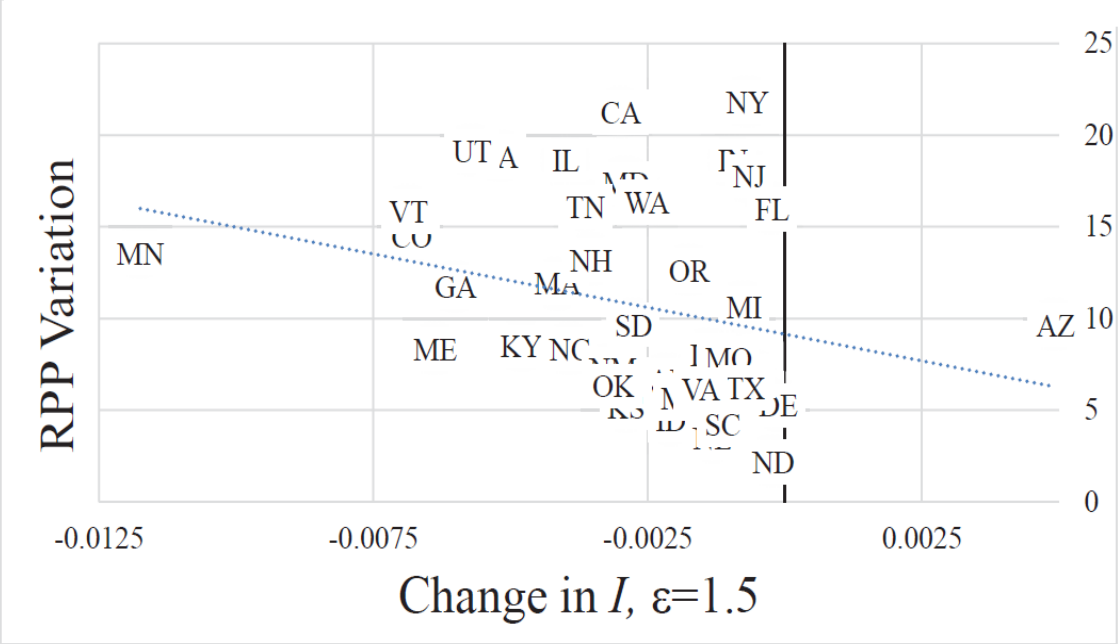

Table 5.2.2: Regression table for RPP variation and Change in Atkinson Index

Regression Statistics	
Multiple R	0.2762
R Square	0.0763
Adjusted R Square	0.0532
Standard Error	5.3449
Observations	43

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	94.3369	94.3368	3.302	0.0767
Residual	41	1142.6992	28.5675	2	

Total	42	1237.036031		

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	9.2556	1.1861	7.8036	0.0000
Atkinson Index	- 575.9259	316.9290	1.8172	0.0767

5.3 Bottom Decile Demographics

The bottom decile represents America's poorest people. Geloso and Msaid found that nationally the makeup of this bottom 10% was quite sensitive to RPP adjustment. They found that before the adjustment many households in the United States' bottom decile lived in rural states. After the adjustment, they found that the proportion of dwellers in more urban states such as New York or California was much higher. The following analysis takes this a step further by adjusting incomes based not off state RPPs but off the RPP of their specific location within these states. This should improve the accuracy of these findings and tell us something about the bottom declines of state distributions rather than being limited only to national statistics.

First it is best to get a sense for how the adjustment affects the overall national statistics. To get a sense of how the adjustment affects the demographic first it was calculated what percentage of the bottom ten percent of the distribution live in RPPs less than 100. In total, within the sample 65.3% of respondents live in such areas. While the mean is near 100, the frequency distribution seems to be somewhat leftward skewed. If RPPs did not matter, we would expect that the proportion of people in the bottom decile who live in low-cost areas would be similar to this number. However, the data shows that the proportion of the bottom decile living in low-cost areas is about 70.5%. After the

adjustment low-cost area dwellers make up 65.4% of the bottom decile. This is a nearly identical proportion in the total population. This is very strong evidence that real incomes are not very different between regions within the US.

The preceding analysis was harder to replicate for each state because RPPs are not standardized to 100 within the states. It would be possible, however, use the overall state RPP and do a similar operation based off of this number. While this should be done in the future, this paper has not extended into that. Rather, it has been favored to observe how the RPP affects the rural population within the bottom decile. This is because previous research has under researched how rural/urban dynamics may play into inequality. Table 5 has instead found the rural population of each state and the percentage of the bottom decile that is rural both before and after the RPP adjustment. Rural components tend to have lower RPPs than most places within the state, so this still shows how the adjustment affects things overall. The limitation, however, is that in states where the rural component was calculated crudely, the percentages may be significantly off.

Table 5.3.1: Rural Components of Bottom Decile (N>1000)

	Proportion Rural	Nominal Bottom 10% Proportion	Real Bottom 10% Proportion	Change	Number
USA	0.248	0.303	0.276	-8.9%	61812
AL	0.325	0.444	0.421	-5.2%	1325
AR	0.427	0.477	0.459	-3.8%	1107
CA	0.027	0.051	0.053	3.7%	5477
FL	0.057	0.088	0.075	-14.6%	3064
GA	0.251	0.431	0.396	-8.1%	1436
ID	0.419	0.528	0.519	-1.7%	1054
IL	0.118	0.195	0.183	-6.2%	1699
LA	0.259	0.322	0.289	-10.2%	1519
MA	0.029	0.016	0.024	47.6%	1219

MI	0.157	0.153	0.146	-4.6%	1439
MS	0.722	0.797	0.779	-2.3%	1211
MT	0.854	0.932	0.932	0.0%	1028
NC	0.198	0.272	0.261	-4.0%	1564
NM	0.286	0.373	0.364	-2.4%	1176
NY	0.045	0.054	0.054	0.0%	2224
OH	0.259	0.277	0.304	9.7%	1476
OR	0.175	0.196	0.176	-10.2%	1017
PA	0.133	0.173	0.160	-7.5%	1560
SC	0.177	0.277	0.232	-16.2%	1120
TN	0.182	0.186	0.186	0.0%	1176
TX	0.089	0.116	0.116	0.0%	3315
VA	0.149	0.305	0.322	5.6%	1173
WA	0.212	0.347	0.331	-4.6%	1209

The samples in these cases range for the most part from small to very small as they deal with only about one tenth of the state's original population and many states have well fewer than 1000 observations to begin with. It is suspected that this is why some states with low populations or low rural counts have anomalous findings. However, because the effect holds for the national statistic and there remain many states with sufficient observations, the data is useful. One must just be aware of the sample sizes when viewing these data.

These statistics show that, for the most part, the RPP adjustment decreases the proportion of rural dwellers in the lowest income bracket. This both supports previous findings as well as the theory regarding real income levels. The expectation would be that the proportion in the bottom 10 percent would be identical to the population in the total population. While this does not necessarily happen in the states, most states do approach the expected proportion. However, after the adjustment most states rural populations still seem to experience higher than expected incidence of very low incomes. This suggests

that while the problem of rural incomes may be overstated at first glance, the problem still persists after adjustment.

This is not to say that it should be an objective to ensure that the rural component of the bottom decile is no higher than the overall rural proportion. Lower real incomes in rural areas should be a signal that agglomeration economies have not yet reached a sufficient scale and that further redistribution is economically efficient. The forces keeping lower paid households in place are not entirely clear but may have to do with variables that cannot be directly priced. In many cases people enjoy a rural lifestyle or the communities that form outside of major centers. While this is difficult to attribute a dollar benefit to, it is equally incorrect to suggest that this evidence means that more people should move into cities. That is one possible interpretation, but with the gap relatively narrow it is also reasonable to suggest that the non-economic value derived from rural dwelling is sufficient to believe that things are nearly in equilibrium. Crudely, the suggestion is that $Real\ Wage_r + Lifestyle\ Value_r = Real\ Wage_u + Lifestyle\ Value_u$ where r represents rural and u represents urban. This is only included to recognize that wages may not be the only deciding factor in location for many people. The lifestyle values will introduce a taste gradient in the population, which is too complex to solve in this paper, but it is useful in practical applications to think of the household as maximizing a utility equation which includes both wage and lifestyle. Alternatively high fixed costs to relocation may keep the bottom decile in place.

6.0 CANADIAN APPLICATION

It is expected that the Canadian data would reflect closely what has been observed in the American example previously. Canada shares a number of important characteristics with the United States that are important in this example. First, Canada and the United States are both free market, open economies. Most prices are determined on a global stage, and the rest are determined by the market. There is little direct government price setting in either country. Second, both countries share a similar urban/rural split. In 2021, the World Bank published that 81.65% of the Canadian population was urban, and 82.87% of the American population. These are remarkably similar. Because the largest price differences are noticed in major cities this is important.

The United States' most expensive cities are, however, considerably more expensive than Canada's most expensive. Mercer's research shows 19 American cities that have a higher cost of living that the most expensive Canadian city (Toronto) in 2023. This is despite Canada as a whole having a 1% higher cost of living than the United States according to the World Bank in 2022. These two facts together suggest that there are much smaller price discrepancies within Canada as in the United States.

Part of this phenomenon may be natural. The United States has a much larger population. This causes large agglomerations in a few cities either for historical or industry related reasons. In these cities the prices are very high and land rents go up with population, as discussed in the literature review. This causes the upward movement of the price level to a larger degree than in Canada. However, domestic policies are no doubt also relevant to this discussion.

The United States have largely allowed and encouraged the redistribution of people seeking better jobs, usually moving from low earning states to higher earning

states. In Canada, policies such as equalization and other stimulus programs targeted at specific areas keep the population and wages in these periphery areas artificially high. It also prevents people from redistributing within Canada to large cities, meaning that Canada's major cities are potentially underpopulated from an economic perspective. This underpopulation and diversion of resources to periphery regions diverts funds away from these cities so that locally set prices do not go as high as they otherwise might. In whole, the regional policies increase prices in the periphery and decrease them in the cores. This obviously results in lower price discrepancies. For this reason, it is suggested that income regional price differences do not likely bias the Gini coefficient in any significant way in Canada.

7.0 CONCLUSION

This report has taken a novel approach to the study of inequality by not only analyzing the effect of subnational price level differentials, but also by taking this down to the level of the CBSA. It has shown that regional price level differentials provide a relevant and important lens through which to study inequality. The paper provided theoretical justification to show that it is logical that regional price levels would change some things about our standard understandings of inequality. It has also supported the theory, despite some faults in the dataset, by showing that purchasing power adjustments can change the Gini coefficient, which is one of the most often cited statistics in relation to inequality. This effect was small in most states, but the effect was more pronounced in states with large price level differentials within them such as California.

The RPP adjustment also altered the makeup of the lowest earning bracket as low-cost and rural dwellers tended to be overrepresented at the bottom of nominal income distributions. Following the RPP adjustment rural representation tend to become nearer to their proportion of the entire population, but they still have high occurrence. It has been posited that this has to do with non-measurable values put on rural lifestyle by individual households and that even with perfect mobility (a very strong assumption), it would not reach their overall proportion. Fixed costs may also be to blame.

With the information provided in this paper we provide two recommendations for policy makers. The first is that regional price levels may have important implications for the distribution of benefits and whether certain benefits should have the same nominal value when distributed to individuals over a large geography. By over-allocating in real terms to low cost and rural dwellers governments may be inefficiently addressing inequality by missing the real lowest earners who more often live in cities and high-cost

areas than nominal figures may indicate. Further implications of this potential over-allocation should be the subject of further research. Secondly, policy makers should ensure that they have strong data on regional price levels within their jurisdiction. Multiple previous authors share the concern that many CPI and RPP calculations overrepresent major population centers in their data collection which causes biased results. Data collection and distribution regarding price levels needs to continue to be improved in the US and Canada.

This research also lends itself to be extended to the price levels for other subpopulations as well such as women and older peoples.

8.0 BIBLIOGRAPHY

Aaberg, Rolf. (2009). "Ranking Intersecting Lorenz Curves." *Social Choice and Welfare 33* (2), 235-259.

Aten, Bettina H. (2017). "Regional Price Parities and Real Regional Income for the United States." *Social Indicators Research, 131* (1), 123–143.

Aten, Bettina. (2018). "Real Personal Income and Regional Price Parities." *World Bank.* https://pubdocs.worldbank.org/en/475111539796688272/pdf/ICP-TF-OGP03-DOC-S03-SubNationalPPPs-BA2.pdf.

Atkinson, Anthony. (1970). "On the Measurement of Inequality." *Journal of Economic Theory,* 2, 244-263.

Bonsall, Nettie. (2021). "Richer and Fairer: New Brunswick Income Trends, 1976-2019." *Center for the Study of Living Standards.*

Chiu, Henry. (2007). "Intersecting Lorenz Curves, the Degree of Downside Inequality Aversion, and Tax Reforms." *Social Choice and Welfare 28,* 375-399.

Chu, Cyrus and Yi-Ting Wang. (2021). *"Gini Coefficient Versus Top Income Shares – Pattern Change Differences." Economics Letters* 201.

Diamond, Rebecca & Moretti, Enrico (2021). "Where is Standard of Living the Highest? Local Prices and the Geography of Consumption." *NBIR Working Paper Series 29533*.

Dixit, A. K., & Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. The *American Economic Review*, 67(3), 297–297.

Efron, Bradley. (1982). "Bias of the Jackknife Variance Estimate." In *The Jackknife, the Bootstrap and Other Resampling Plans,* 21-26.

Emery, Herbert & Guo, Xiaolin (2020). "Using an Engel Curve Approach to Infer Cost of Living Experienced by Canadian Households." *Canadian Public Policy* 46 (3), 397-413.

Frankfurt, Harry. (1987). "Equality as a Moral Ideal." *Ethics 98* (1), 21-43).

Fujita, Masahisa, Krugman, Paul, & Venables, Anthony. (2001). *The Spatial Economy: Cities, Regions, and International Trade*. MIT Press.

Geloso, Vincent and Youcef Msaid.(2018). "Adjusting Inequalities for Regional Price Parities: Importance and Implications." *Journal of Regional Analysis and Policy 48* (4), 1-8.

Hamilton, B. W. (2001). “Using Engel’s Law to Estimate CPI Bias.” *American Economic Review*, 91 (3), 619–630.

Karagiannis, Elias & Kovacevic. (2000). “A Method to Calculate the Jackknife Variance Estimator for the Gini Coefficient.” *Oxford Bulletin of Economics and Statistics,* 119-122.

Krugman, Paul (1991). “Increasing Returns and Economic Geography.” *Journal of Political Economy 99* (3), 483-499.

Krugman, Paul (1991). “Space: The Final Frontier.” *Journal of Economic Perspectives 12* (2), 161-174.

Locke, Wade, Doug May, Noel Roy, Michael Sullivan, & Craig Wiseman. (2015). “Promoting Tax Fairness in Newfoundland and Labrador Through The Personal Income Tax System.” *Collaboration for Applied Research in Economics Research Report.*

De Maio, Fernando (2007). “Income Inequality Measures.” *Journal of Epidemiology and Community Health 61* (10), 849-852.

Mehran, Farhad. (1976). “Linear Measures of Income Inequality.” *Econometrica (44)* no. 4, 805-809.

Oxfam. (2023). “Survival of the Richest: How we Must Tax the Super Rich now to Fight Inequality.”

Piketty, Thomas (2014). Capital: in the Twenty-First Century. (A. Goldhammer, Trans.) *Belknap Press of Harvard University Press.*

Piketty, Thomas. (2003). “Income Inequality in France, 1901-1998.” *Journal of Political Economy 111* (5), 1004-1042.

Piketty, Thomas. (2003). “Income Inequality in the United States 1913-1998.” *The Quarterly Journal of Economics 118* (1), 1-41.

Sheshinski, Eytan. (1971). “Relation Between a Social Welfare Function and the Gini Index of Income Inequality.” *Journal of Economic Theory 4,* 98-100.

USAFacts.org. (2023). “How much do Americans Spend on Housing?” usafacts.org. https://usafacts.org/data-projects/housing-costs.

Veall, Michael & Saez, Emmanuel. “The Evolution of High Incomes in Northern America: Lessons from Canadian Evidence.” *The American Economic Review 95* (3), 831-849.

Veall, Michael (2012). “Top Income Shares in Canada: Recent Trends and Policy Implications.” *The Canadian Journal of Economics 45* (4), 1247–1272.

CURRICULUM VITAE

Joseph Drost

Post-secondary education

University of New Brunswick, Master of Arts in Economics

Thesis: “Price Consideration in the Measurement of Inequality.”

September 2022 – Expected May 2024

University of New Brunswick, Bachelor of Arts, Joint Honours in Economics and Political Science

Honours thesis: “The Universal Declaration of Human Rights and the United Nations Declaration on the Rights of Indigenous Peoples: Two Case Studies in Human Rights Approaches"

September 2018-April 2022

Conference presentations:

“Boost NB New Brunswick Economic Snapshot”

Shaping the Debate Conference

Fredericton, NB

February 2023

“The Universal Declaration of Human Rights and the United Nations Declaration on the Rights of Indigenous Peoples: Two Case Studies in Human Rights Approaches"

UNB Arts Matters Conference

Fredericton, New Brunswick

April 2022