Estimating county-level regional price parities from public data

Author: AnnElizabeth McMahon

Contact: amcmahon@doc.gov

Date: March 2024

Abstract: Using publicly available data, I estimate experimental county-level regional price parities (RPP) for all counties in the United States. I use micro-data to approximate utilities, housing, and goods county-level RPPs and combine these components into RPP All Items estimates using personal consumption expenditure weighting. I exploit the relational nature of RPPs and relevel my RPP All Items estimates which allows my estimates to closely align with published RPP data from the Bureau of Economic Analysis when aggregated. Using the experimental RPP estimates, I discuss high-level findings such as within metropolitan area price differentials and conclude with potential areas for future methodological improvement.

I thank numerous individuals for helpful comments, particularly Ben Bolitzer, Jane Dokko, and Jed Kolko. I also thank the Bureau of Economic Analysis for their excellent work in developing and producing regional price parity data. This work would not be possible without their years of diligent and thoughtful research. Additionally, I thank Feeding America for estimating food price differences and extensively detailing their methodology.

Introduction

Published by the Bureau of Economic Analysis (BEA), regional price parities (RPPs) are valuable economic data series which measure price level differences across geographic areas in a given year. As detailed by BEA, RPPs allow for place-to-place comparisons and are expressed as a percentage of the overall national price level. By construction, RPP for the entire United States is equal to 100 which enables comparison between places. In 2022, Alabama's state-wide RPP was 87.8 while Alaska's was 102[1]. From this data, we calculate that for the average Alaskan consumer, Alaska's prices were 16% higher than those in Alabama and 2% above prices nationally. Moreover, the concept of differing prices by place is well known. Most consumers could easily identify that a day in New York City would cost more than an identical day in rural Texas. RPP data is valuable because it quantifies place-based price differences by using a weighted average of goods and services price levels.

Currently, RPP data is available at three geographic cuts: state (SARPP), portion areas (PARPP) and metropolitan areas (MARPP). The available geographic cuts of RPP data are extensive, covering all fifty states, the District of Columbia and 384 metropolitan areas. The PARPP data has two groupings, metropolitan portion area and non-metropolitan portion area. For instance, the PARPP metropolitan area for South Carolina covers all of counties in South Carolina that are in a metropolitan area such as Charleston and Spartanburg while the non-metro portion covers all counties of South Carolina which are outside a metropolitan area.

Additionally, RPP data on goods, housing, utilities, and other services is available for each geographic cut. This provides users with a wide array of data. For example, users can compare relative housing prices between metropolitan areas or goods prices between states. These four RPP components also aggregate into an additional series, RPP All Items. This overarching data series contains "all the detailed consumption goods and services used in the estimation of the RPPs[2]" and is a wholistic look at price differences by place.

While county-level RPP estimates are created as part of BEA's overall RPP methodology, they are not currently published as more work is needed to verify their reliability. This leaves a data gap, particularly for assessing price differences between non-metro counties within a state. PARPP non-metro data aggregates across sizeable areas and may mask residents' differing economic experiences when it comes to prices, particularly in resort towns. More granular RPP data could be useful for people making decisions about policy and planning. For example, a recent GAO report stated "41% of U.S. Coast Guard units are located in remote areas or high vacation rental towns[3]" which can make it difficult for service members to find affordable housing.

---

[1] Source: Bureau of Economic Analysis

[2] BEA Interactive Tables, RPP data

[3] [Coast Guard: Better Feedback Collection and Information Could Enhance Housing Program | U.S. GAO]

This working paper documents the construction of experimental county-level RPPs using public data. With a simplistic methodology, I harness existing public data to estimate county-level price differences and explore varying economic experiences at the local level.

The rest of the paper proceeds as follows. Section 1 details the construction of Utilities, Housing, and Goods county-level RPP estimates using micro-data. Section 2 details the releveling of these three components and compares their aggregations to the equivalent BEA data. Section 3 details how the components combine into an estimate for RPP All Items via consumer expenditure weighting. Section 4 discusses the releveling of RPP All Items estimates and compares them to BEA data. Section 5 highlights some high-level findings from the experimental data set, and the paper concludes with possible research applications and future areas for methodological improvement.

Section 1: Component construction

In this section, I detail the construction of Utilities, Housing and Goods county-level RPP estimates using micro-data. While RPP Other Services is a component of RPP All Items, I am unable to estimate it directly at the county level due to the lack of available data. In section 3, I integrate Other Services data into my estimation of RPP All Items.

Utilities

To use similar methodology to BEA, I utilize Energy Information Administration (EIA) data for annual state-level energy prices. From EIA's State Energy Data System (SEDS), I pull residential energy prices by fuel type by state for 2021. The fuel types I use are propane, natural gas, electricity, wood, distillate fuel oil and kerosene and I average the latter two fuel type prices into the category 'oil'. Units across the fuel types are the same, dollars per million BTU. I also pull national prices for each of these five fuel types, again averaging distillate fuel oil and kerosene into the category 'oil'. Joining the state prices and national prices, I calculate the premium or discount for each state by fuel type. For example, a state's electricity price could be discounted but its natural gas price could be at a premium when compared to the respective national prices.

Using the American Community Survey (ACS) 5-year 2021 data, I pull the distribution of home units heated by natural gas, propone, electricity, oil or wood at the county-level. Combining both the EIA data and ACS data, I use the county-level distribution of homes by fuel type as weights when averaging state-level premiums and discounts. This produces my Utilities RPP estimate.

The underlying assumption of my approach is that fuel type is a main reason why utilities prices vary from one county to another. If county A primarily uses natural gas while county B primarily uses electricity, local pricing will be a function of whether a state has a premium or discount on those fuel types. Additionally, most consumers are not able to easily switch between home heating fuel types without incurring substantial costs. As consumers are dependent on trends in construction and local heating infrastructure, they are effectively utility 'price takers' based on fuel type.

Housing

Using the same approach as utilities, I utilize ACS 5-year data from 2021 to pull the distribution of homes by bedroom type and median gross rent by bedroom type at the county-level. I use rents instead of housing prices to avoid the forward-looking expectations that are present in housing prices. Bedroom types range from studio to 5 bedrooms or more.

I construct a national average rent for each bedroom type and then calculate the premium or discount of a county's rent compared to the national rent. For example, a county's one-bedroom median rent may be discounted compared to the average rent of one-bedrooms nationally, but there may be a premium on four-bedroom rents. To construct each county's Housing RPP, I average each county's rent premiums and discounts, weighted by bedroom types within that county.

There are 16 counties with no rent data for any bedroom type, with most of these counties being under 1,000 people. Using Census's 2010 county adjacent file, I use the estimated Housing RPP of an adjacent county with the most similar population size in absolute terms.

Goods

For estimating Goods RPP at the county-level, capturing all goods was not possible due to the lack of available micro-data. Therefore, I focus on differences in food prices by place. Food likely drives a substantial portion of placed based goods price variation and there is a wide literature about food prices' impact on food insecurity and health outcomes.

Feeding America, a non-profit working to end hunger, estimates the cost of a meal by county as an input into their food insecurity calculation. Using scanner data, they construct a relative price index that "allows for comparability between counties"[4] and they control for the basket of food they are comparing between places. Using this cost-of-food index, they take "the average dollar amount spent on food per meal by food-secure individuals"[5] and "adjust the national average cost per meal by a relative food cost index to derive a local estimate"[6]. Feeding America's food expenditure data is derived from Census's Current Population Survey; the full methodology is available on the Feeding America website.

From Feeding America's 2021 county-level 'cost-per-meal' estimates, I calculate each county's premium, or discount compared to the national cost-per-meal. I use the national cost-per-meal as calculated by Feeding America.

Section 2: Components releveled

Once each component is constructed, I relevel it to better fit within the aggregation structure of BEA's data. I do this to fully integrate BEA's published RPP data into my estimates since BEA's

---

[4] Map the Meal Gap 2023 Technical Brief.pdf (feedingamerica.org)
[5] Map the Meal Gap 2023 Technical Brief.pdf (feedingamerica.org)
[6] Map the Meal Gap 2023 Technical Brief.pdf (feedingamerica.org)

methodology and data sources are more sophisticated than my approach. This is also mathematically feasible since I am applying a level shift to the data and the variable of interest is a relative price relationship rather than the price itself.

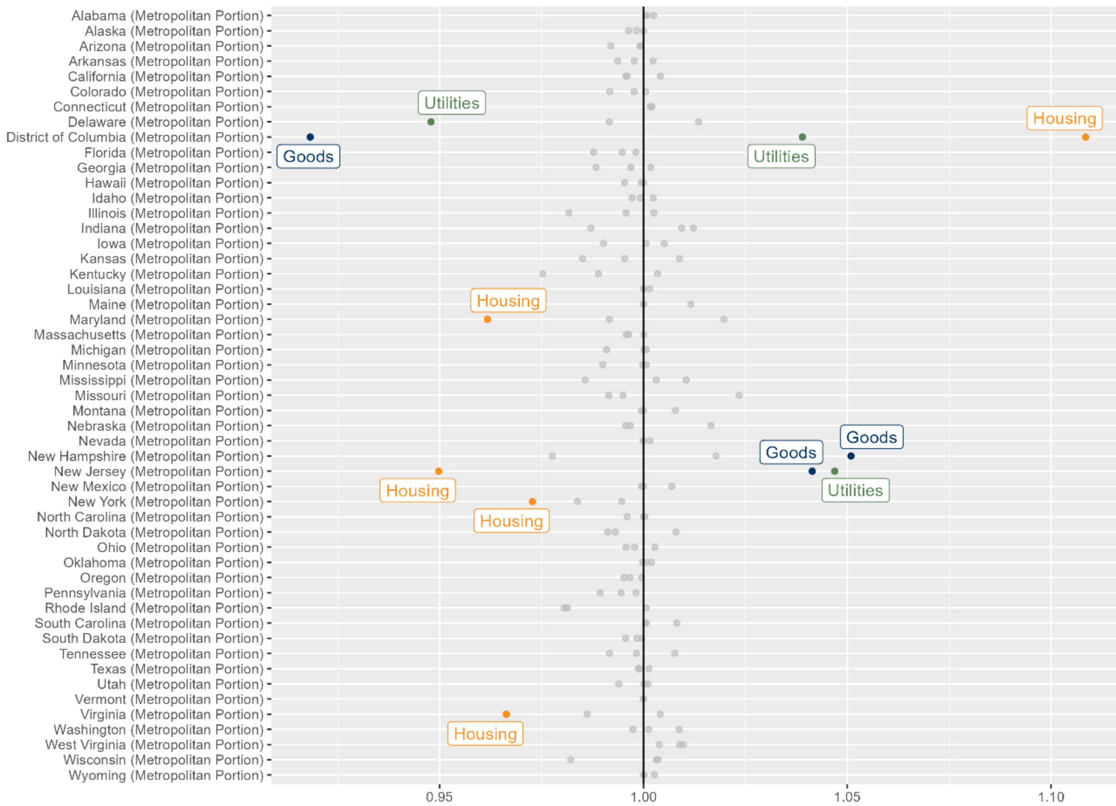
To relevel, I aggregate each of my three components (utilities, housing, goods) via population weighting to the lowest published geography. For counties in a metro, I aggregate my data to each respective metropolitan area (MARPP). For counties outside a metro, I aggregate to the non-metro portion area (PARPP) for that state. I then divide my estimate by BEA's data to get a releveling factor for each component and apply these releveling factors back to the counties. Note that I use the 2020 Decennial Census for my county-level population counts and for all BEA data, I utilize 2021 vintages. Histograms of the releveling factors are available in the Appendix, see Figures A1-A6.

After this procedure, my releveled estimates respectively aggregate via a population weighted average to MARPP and PARPP non-metro data exactly. This is expected; these geographic cuts generated the releveling factors hence aggregations of my releveled estimates match the equivalent BEA data. When aggregating my estimates to the PARPP metro and SARPP geographic levels, many of my estimates are similar to BEA's data however there are some outliers.

As seen in Figure 1, I aggregate my county-level estimates to the PARPP metro level and compare them to BEA's data by component. A quotient of one means the BEA data and my estimate match exactly. If the quotient is greater than one, my estimate is smaller than BEA's data; if the quotient is less than one, my estimate is larger than BEA's. From this comparison, the majority of my estimates are off by less than 2.5% and are shown in grey in Figure 1. The colored dots represent outliers by component type, with District of Columbia leading among the outliers. To see non-releveled estimates aggregated to PARPP metro and compared to BEA data, see Figure B1 in the Appendix.

Figure 1: Difference between BEA PARPP metro portion data and releveled estimate

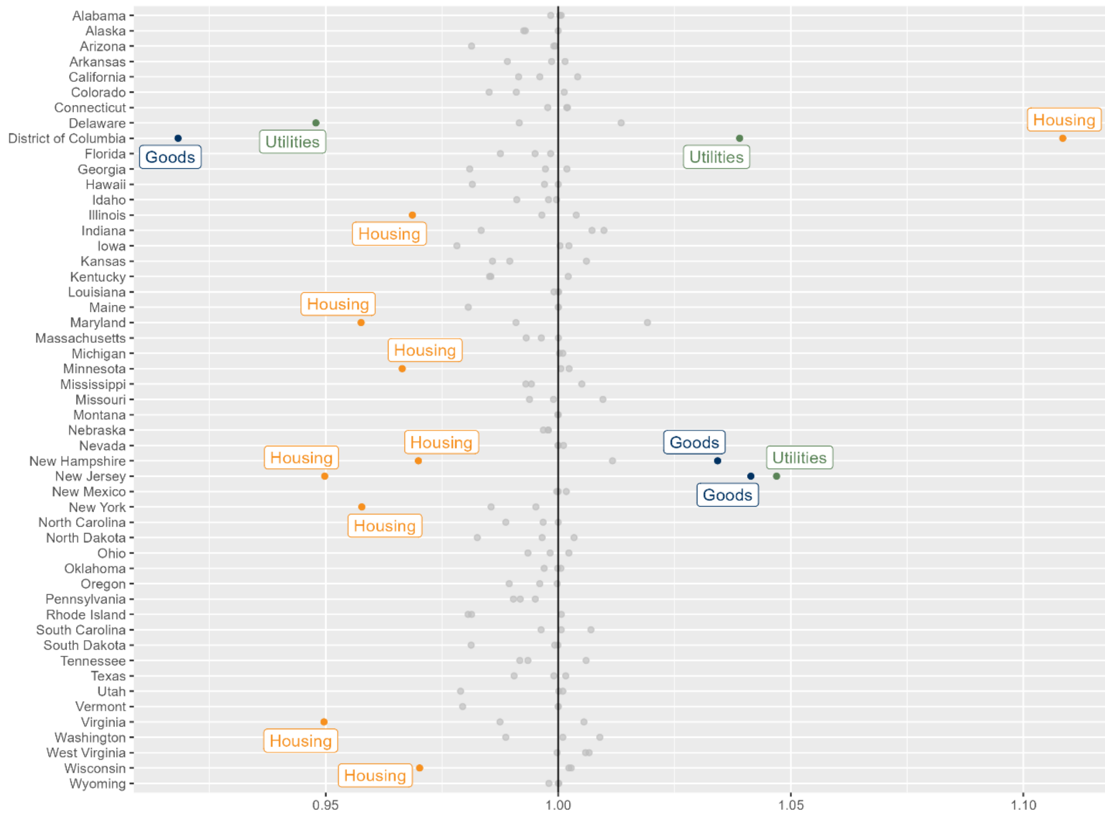
BEA data divided by experimental estimate.
Differences of more than 0.025 from 1 highlighted



As seen in Figure 2, outliers are present when comparing BEA SARPP data and my estimates. As in the previous comparison, RPP components for the District of Columbia lead among the outliers. It is unclear how much of the overall outlier variation is due to random versus systemic error. On the latter, systemic error could be due to how the components were constructed, either methodological or from data source choices. To see the non-releveled SARPP estimates compared to BEA data, see Figure B2 in the Appendix.

Figure 2: Difference between BEA SARPP data and releveled estimate

BEA data divided by experimental estimate.
Differences of more than 0.025 from 1 highlighted



Section 3: Combining into 'All Items'

To combine the releveled, component RPP estimates into a single RPP All Items estimate for each county, I use a simple consumption weighting model. It is important to note this is a substantially more simplistic method than BEA's approach of a Geary multilateral price level index via a simultaneous solver.

Leveraging personal consumption expenditure (PCE) data from BEA for 2021, I calculate shares of consumer expenditure on goods, housing, utilities, and other services[7] by state. The share of consumer expenditure is represented as *c* in the equation below. My RPP component estimates are written as *rpp* with the corresponding subscript to denote housing (*H*), utilities (*U*) or goods (*G*).

$$c_H rpp_H + \ c_U rpp_U + c_G rpp_G + \ c_O BEA_O = RPP\ All\ Items$$

[7] Share of consumer expenditure on other services is household consumption expenditure on services – housing - utilities.

As noted earlier, I did not estimate county level RPP for Other Services. Therefore, I use the lowest available level of geography of BEA RPP data (MARPP and PARPP non-metro, respectively) to include other services in my equation, denoted with subscript (*o*).

Section 4: 'All Items' releveled

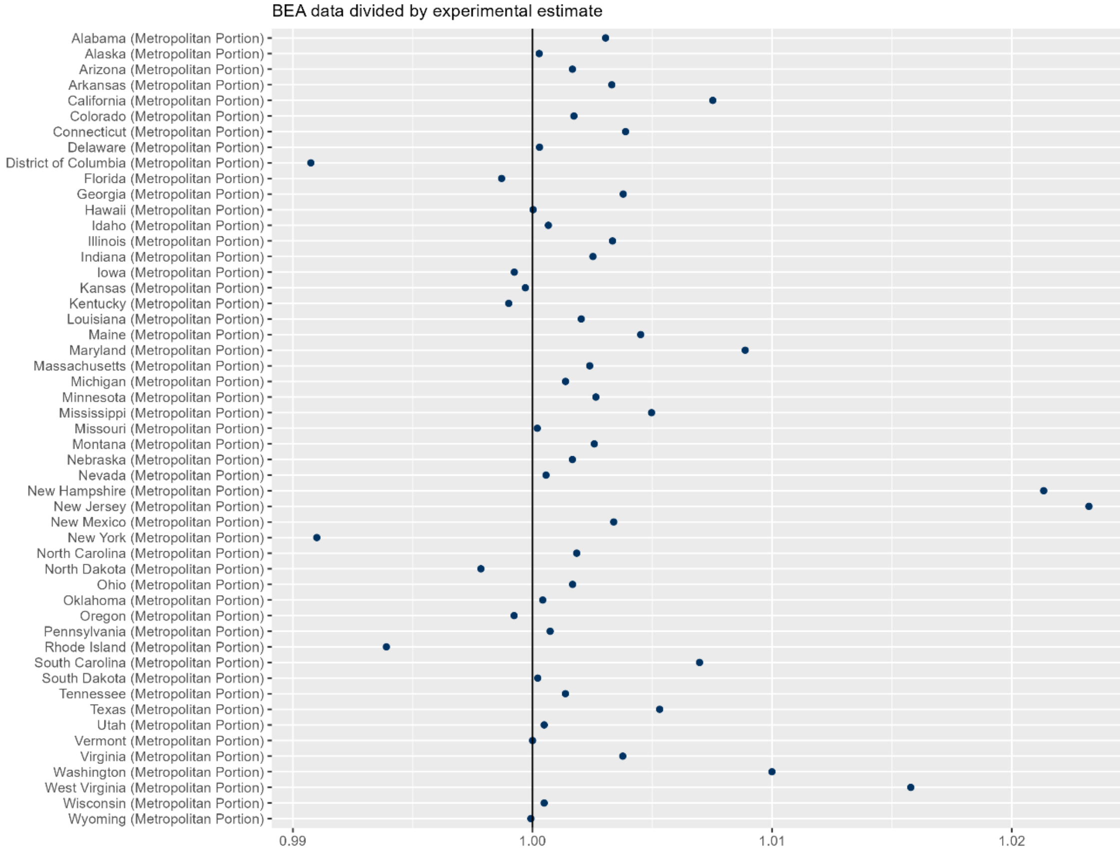
After calculating RPP All Items for each county, I apply the same releveling approach that was used for the components. After aggregating my RPP All Items estimates to either the MARPP or PARPP non-metro level via population weighting, I divide my estimate by BEA's data to yield a releveling factor. I then apply this factor back to the county-level estimates.

Because RPP All Items is releveled using MARPP and PARPP non-metro data, aggregations of my estimates are an exact match compared to BEA data. Additionally, the releveling factors for RPP All Items have a noticeably smaller distribution than the releveling factors for the components. Histograms of the releveling factors for RPP All Items are available in the Appendix, see Figures C1-C2.

Figure 3 shows the aggregation of releveled RPP All Items estimates compared to BEA's equivalent PARPP metro portion area data. As in previous Figures, a quotient of one represents an exact match while a quotient greater than one shows my estimates are smaller than BEA's. As seen on the x-axis, the majority of the quotients are between 0.99 and 1.01.

Figure 3: RPP All Items, difference between BEA PARPP metro portion area data and experimental estimates

BEA data divided by experimental estimate

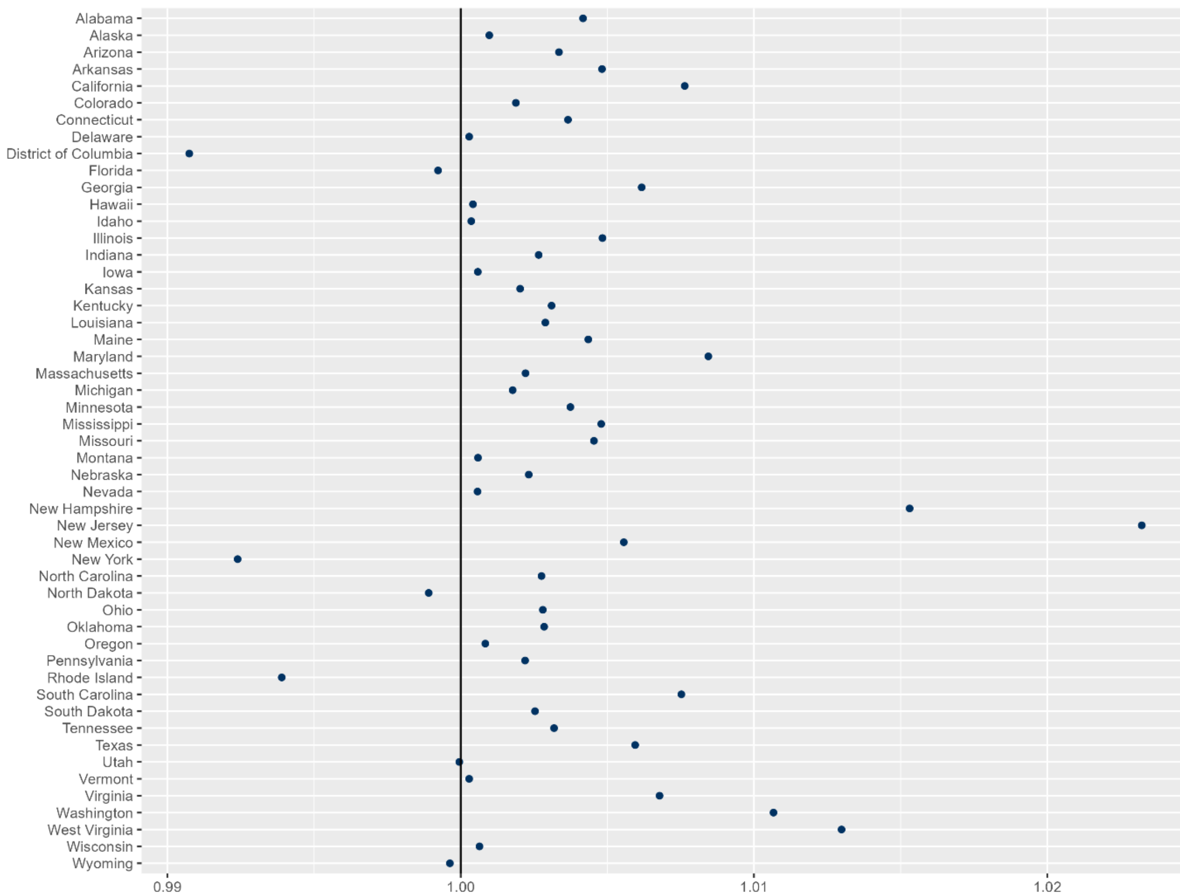


As seen in the Appendix in Figure D1, prior to releveling, the majority of experimental estimates were smaller than BEA's estimates. This suggests that the releveling is a necessary step given the methodology for constructing RPP All Items is simplistic, as detailed in Section 3. Yet once releveled, the aggregated experimental estimates still retain their relational properties and are quite close to BEA's data.

Figure 4 is the aggregation of the releveled, experimental estimates compared to the SARPP BEA data and shows the comparison quotient. Similar to the previous figure, the majority of experimental estimates are within 1% of BEA's data. Figure D2 in the Appendix shows raw RPP All Items estimates compared to BEA SARPP data.

Figure 4: RPP All Items, difference between BEA SARPP data and experimental estimates

BEA data divided by experimental estimate

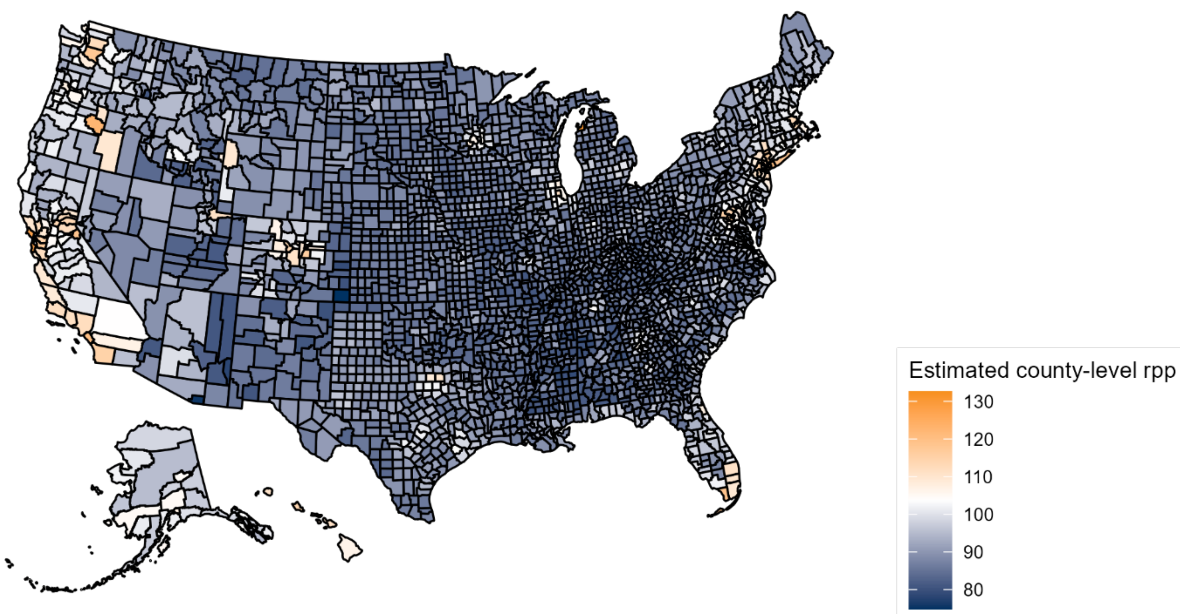


Section 5: What the estimates show

With RPP All Items constructed, I turn to the experimental estimates themselves. Figure 5 shows RPP All Items estimates for all 3,143 US counties. Visually, portions of California's coast, the Upper Northwest, Southern Florida, Hawaii, and parts of New England emerge as more expensive than the national average. While there is no precise county level data to compare to, these trends match my priors about the geography of prices. Two additional comparisons at the state level data are available in the Appendix, Figures E1-E2.

Figure 5: United States, RPP All Items

Experimental county-level rpp estimates, relative to national



Ranking all counties from most expensive to least, 8 of the top 10 most expensive counties are in a metropolitan area as seen in Table 1. While cities are typically more expensive than their suburban or rural counterparts, prices in these counties are substantially high with New York, NY (Manhattan) leading at 32.6% above the national average. Yet of these 10 counties, they tend to be in the same metropolitan areas. Seven of the ten are in the New York, San Francisco or Washington, DC metropolitan areas. When extending to the top 20 most expensive counties, 13 of the counties are in these same three metropolitan areas.

Table 1. Most expensive counties

Experimental RPP All Items ranking

Rank	County	Metropolitan Area	RPP All Items
1	New York County, New York	New York-Newark-Jersey City, NY-NJ-PA	132.6
2	Leelanau County, MI		130.9
3	San Mateo County, CA	San Francisco-Oakland-Hayward, CA	127.7
4	Marin County, CA	San Francisco-Oakland-Hayward, CA	127.2
5	Arlington County, VA	Washington-Arlington-Alexandria DC-VA-MD-WV	125.2
6	San Francisco County, CA	San Francisco-Oakland-Hayward, CA	123.3
7	Crook County, OR		122.6
8	Nassau County, NY	New York-Newark-Jersey City, NY-NJ-PA	120.6
9	Orange County, CA	Los Angeles-Long Beach-Anaheim, CA	120.3
10	Falls Church, VA	Washington-Arlington-Alexandria DC-VA-MD-WV	120.2

Table 2 shows the top five largest price differentials within a metropolitan area. In other words, I compute the price differential between a metro's most expensive and least expensive county. For example, among counties in the Washington, D.C. metro area, Arlington County, VA is 37% more expensive than Madison County, VA. Most metropolitan areas have smaller price gaps than this; 86% of metro areas have a price differential of less than 10% between their most and least expensive counties. However, there are 54 metros with a price differential of 10% or more.

Table 2. Largest 'within metro' price differentials

Price differentials by metropolitan areas, ranked

Rank	Metropolitan Area	Most Expensive County	Least Expensive County	Price Differential
1	Washington-Arlington-Alexandria DC-VA-MD-WV	Arlington County, VA	Madison County, VA	37.0%
2	New York-Newark-Jersey City, NY-NJ-PA	New York County, NY	Pike County, PA	28.8%
3	Nashville-Davidson-Murfreesboro-Franklin, TN	Williamson County, TN	Smith County, TN	27.8%
4	Atlanta-Sandy Springs-Roswell, GA	Fulton County, GA	Heard County, GA	26.5%
5	Charlotte-Concord-Gastonia, NC-SC	Mecklenburg County, NC	Chester County, SC	22.6%

Price differentials also exist in states' non-metropolitan counties. Table 3 compares the most and least expensive non-metropolitan counties within a state. Numerous factors can drive a state's non-metro county price variation; one county may have luxury resorts while another, hundreds of miles away, could be agriculture heavy. What is clear is that these price differentials highlight non-metro residents within the same state may face substantially different price experiences.

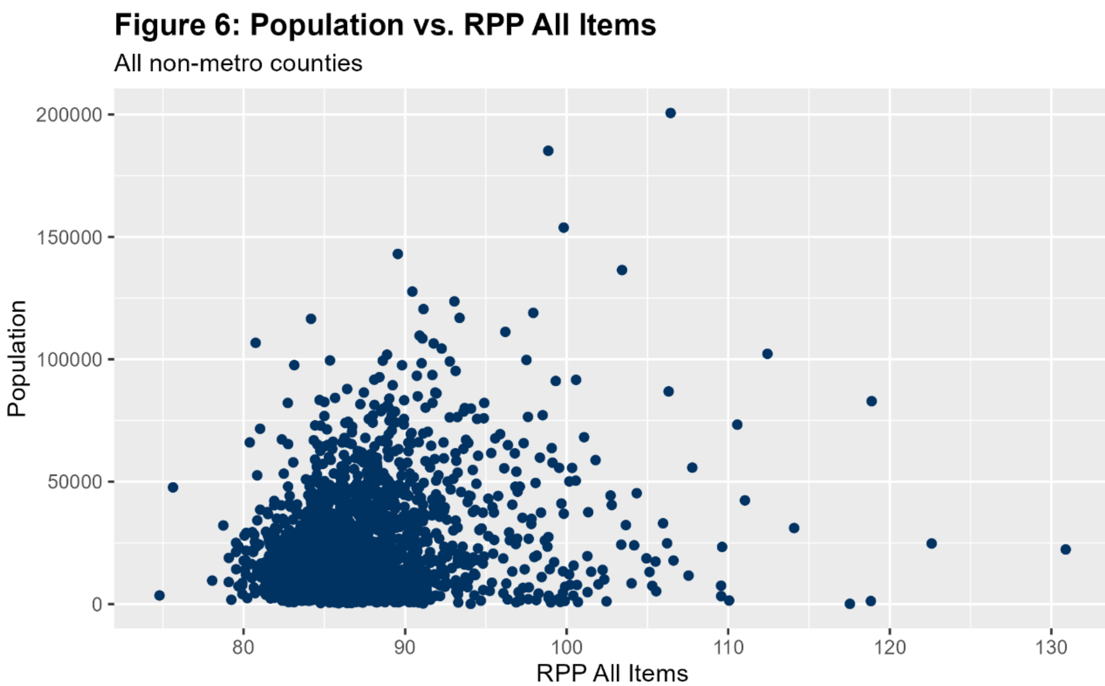
Some counties with notable resort towns appear among the top 5 largest differentials, as seen in Table 3. Monroe County, FL includes Key West and Summit County, CO has multiple ski towns. Leelanau, MI is located on the shore of Lake Michigan and has a variety of summer home rentals and wineries. Of all the states, Massachusetts has the smallest non-metro price differential at 3.7%. Dukes County, MA and Nantucket, MA are the only two Massachusetts counties not in a metropolitan area, are geographically close and are both popular vacation spots.

Table 3. Largest 'non metro' price differentials

Price differentials by state, ranked

Rank	State	Most Expensive Non-Metro County	Least Expensive Non-Metro County	Price Differential
1	Michigan	Leelanau County	Huron County	55.2%
2	Colorado	Summit County	Baca County	52.5%
3	Florida	Monroe County	Dixie County	38.1%
4	Oregon	Crook County	Malheur County	36.7%
5	Utah	Summit County	Emery County	36.6%

Most non-metro counties have an RPP All Items estimate below the national average but not all. As seen in Figure 6, there are a handful of non-metro counties of varying population sizes where RPP All Items estimates are equal to or greater than 100. While the majority of non-metro counties are cheaper than the national average, those that are more expensive could be so for a variety of reasons. They may include resort towns, be located in micropolitan areas or could be extremely rural, where goods can be expensive despite cheap housing.



Conclusion

The purpose of this working paper is to construct RPP All Items estimates at the county-level using public data. With simplistic methodology and exploiting the relational nature of the data, RPP estimates can be constructed from micro-data while closely fitting within the aggregation structure of BEA RPP data.

My experimental estimates suggest there is substantial county-level price variation both within metropolitan areas and across non-metro counties. This illustrates that residents within the same metropolitan area may face considerably different prices and thus may have different purchasing power. Similarly, residents of non-metros counties face varying price differentials. For example, one non-metro county may have high housing costs due to being a resort town, another may have low housing costs, but expensive goods given its remote location.

This experimental dataset can have a variety of research applications to further understanding of economic variation across metro/non-metro counties and within metropolitan areas. From calculating county-level cost-of-living adjusted wages to grouping counties by similar price profiles, this dataset may be able to aid research on important and challenging questions.

There is also a host of future methodological work which could improve the experimental estimates. Additional information such as housing structure type and water costs could be integrated into the RPP component estimates. Sensitivity tests could assess whether ACS 1-year data could be substituted for ACS 5-year data when constructing the housing and utilities components. Currently county-level data is aggregated via population weighting; expenditure weighting may be an additional or better weighting option. Presently MARPP or PARPP non-metro data generate the releveling factors, respectively, but with a simultaneous solver PARPP metro and SARPP data could also be incorporated into the releveling factors.

Bibliography

Aten, B. H., Figueroa, E. B., & Martin, T. M. (2012). Regional price parities for states and metropolitan areas, 2006-2010. *Survey of Current Business*, *92*(8), 229-242.

Bittner, Amy (2023). *Wyoming Cost of Living Index for the Fourth Quarter of 2022.* Wyoming Administration & Information, Economic Analysis Division http://eadiv.state.wy.us/wcli/NewsRelease-4Q22.pdf

Bureau of Economic Analysis. (2023, April). *Methodology for Regional Price Parities, Real Personal Consumption Expenditures, and Real Personal Income* April 2023. https://www.bea.gov/system/files/methodologies/Methodology-for-Regional-Price-Parities_0.pdf

Dewey, Jim (2023). *The 2022 Florida Price Level Index.* Florida Department of Education. https://www.fldoe.org/core/fileparse.php/7507/urlt/fpli.pdf

Government Accountability Office. (2024). Coast Guard: better feedback collection and information could enhance housing program (GAO Publication No. 24-106388). United State Government Accountability Office. https://www.gao.gov/assets/d24106388.pdf

Hake, M., Engelhard, E., & Dewey, A. (2023). Map the Meal Gap 2023: An Analysis of County and Congressional District Food Insecurity and County Food Cost in the United States in 2021.

Appendix

Figure A1: Distribution of housing releveling factors (MARPP)

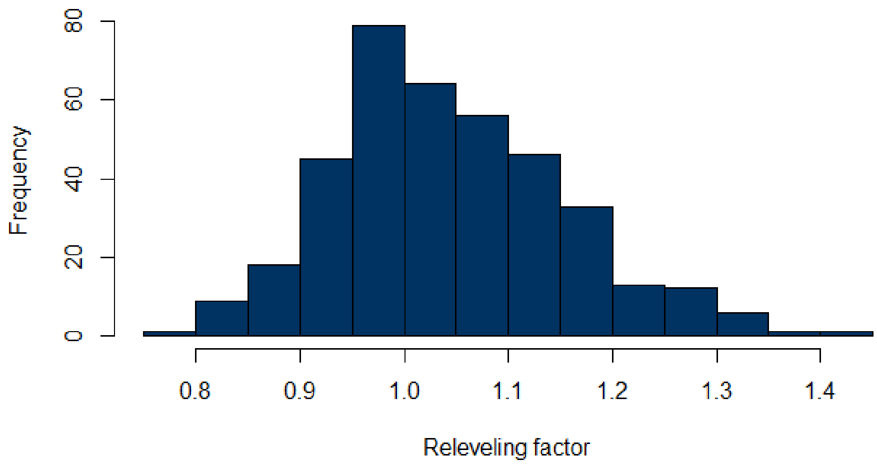


Figure A2: Distribution of goods releveling factors (MARPP)

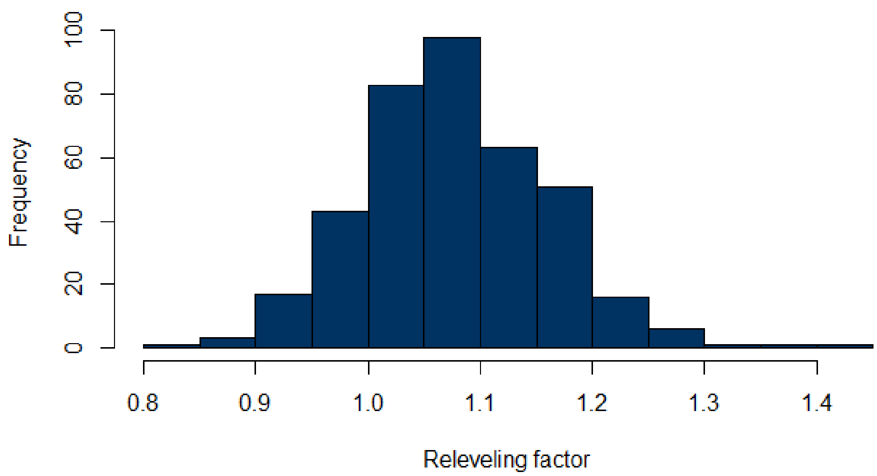


Figure A3: Distribution of utilities releveling factors (MARPP)

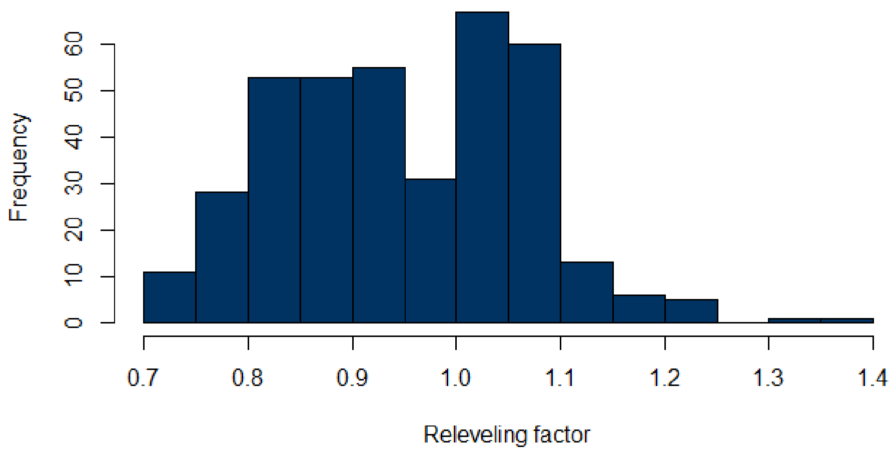


Figure A4: Distribution of housing releveling factors (PARPP non-metro)

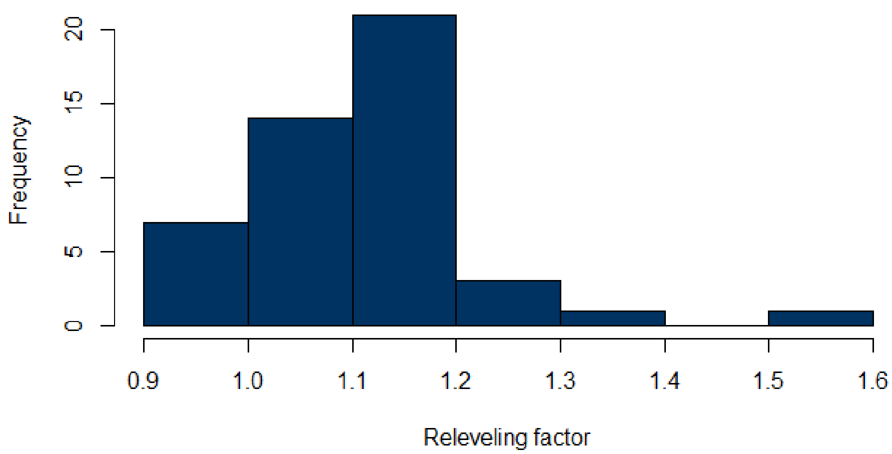


Figure A5: Distribution of goods releveling factors (PARPP non-metro)

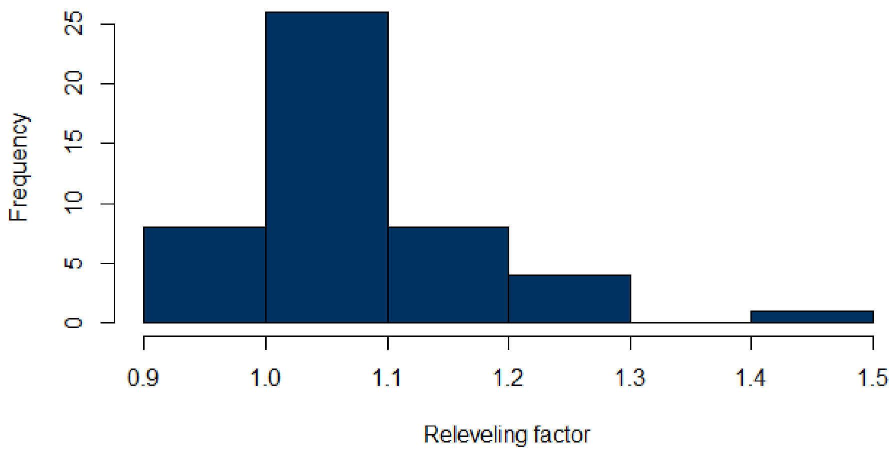


Figure A6: Distribution of utilities releveling factors (PARPP non-metro)

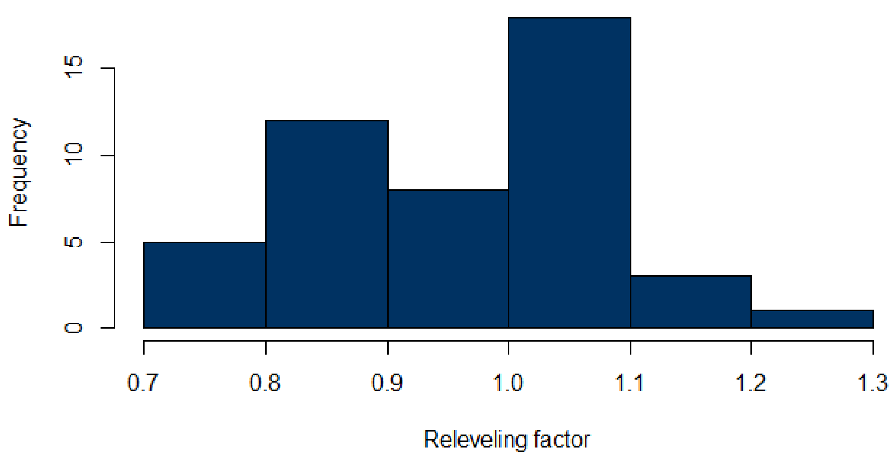


Figure B1: Difference between BEA PARPP metro portion area data and raw estimates

BEA data divided by experimental estimate

Figure B2: Difference between BEA SARPP data and raw estimates

BEA data divided by experimental estimate

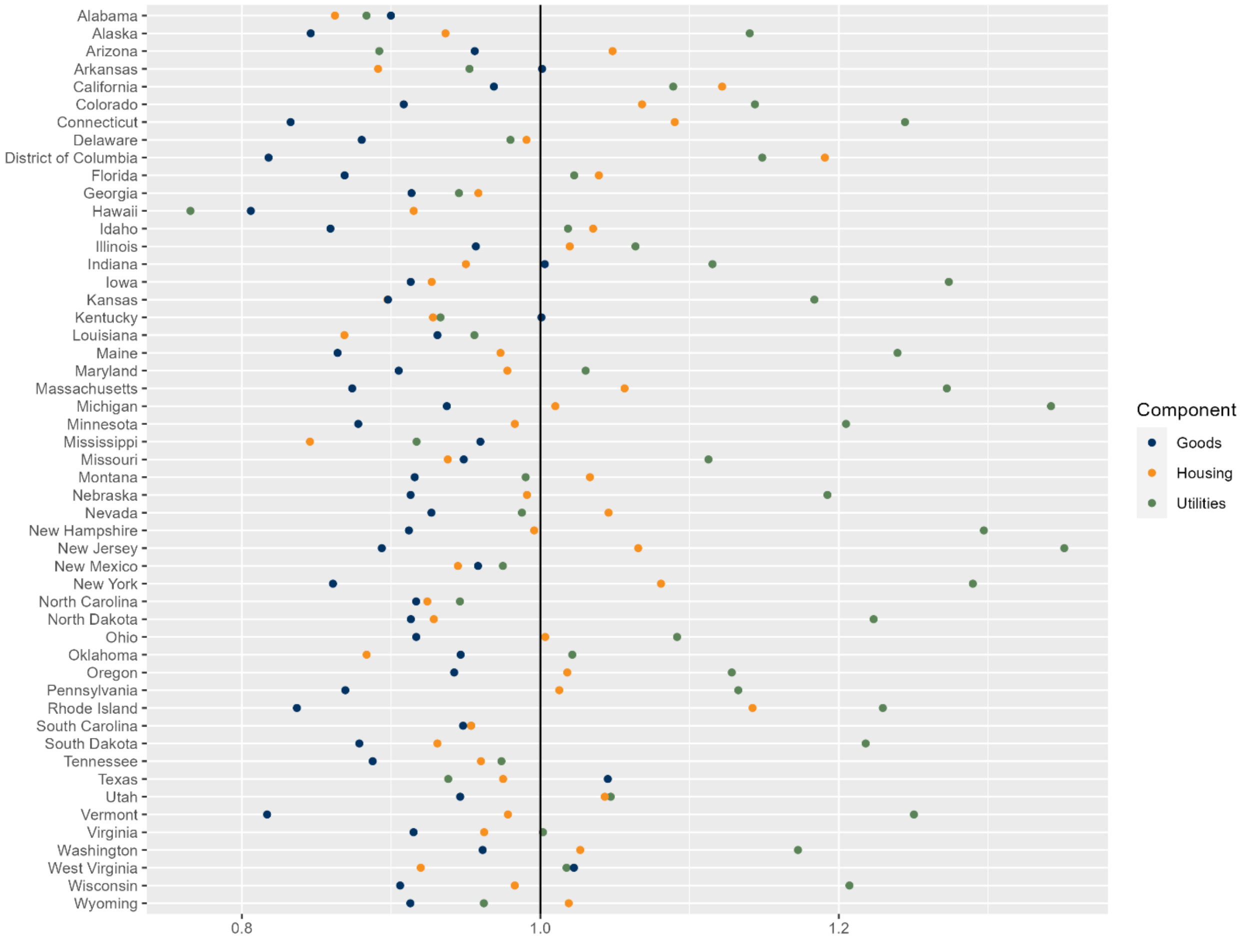


Figure C1: Distribution of all items releveling factors (PARPP non-metro)

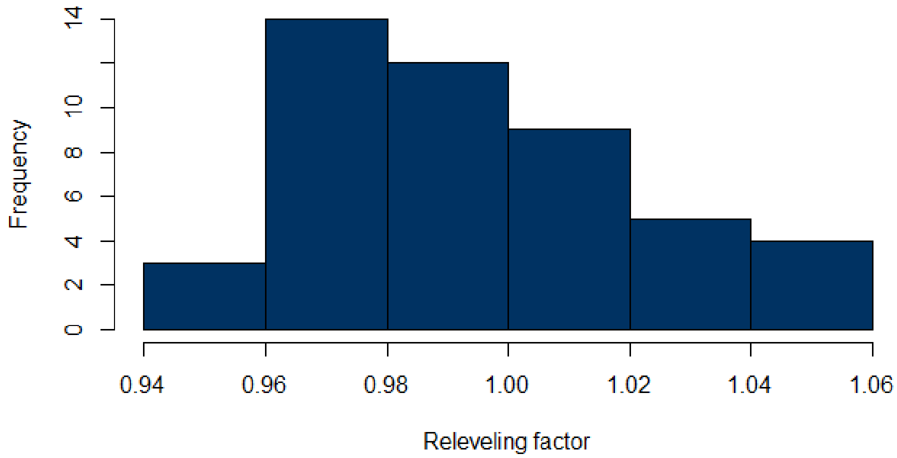


Figure C2: Distribution of all items releveling factors (MARPP)

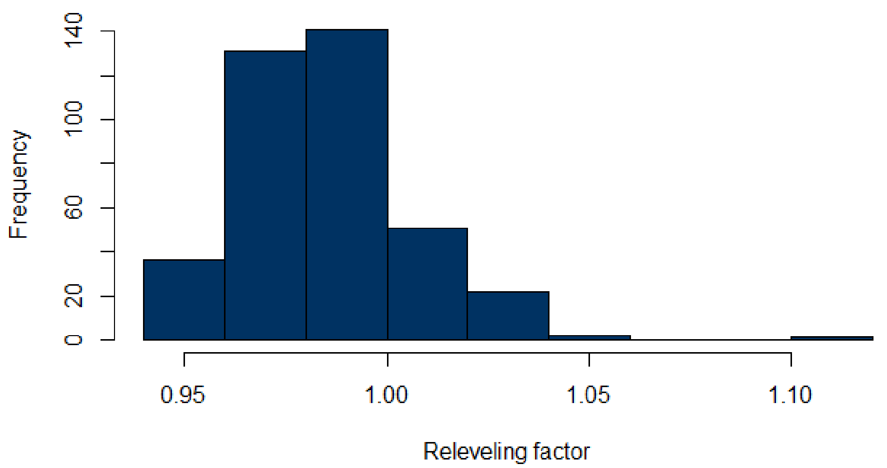


Figure D1: RPP All Items, difference between BEA PARPP metro portion area data and raw estimates

BEA data divided by raw experimental estimate

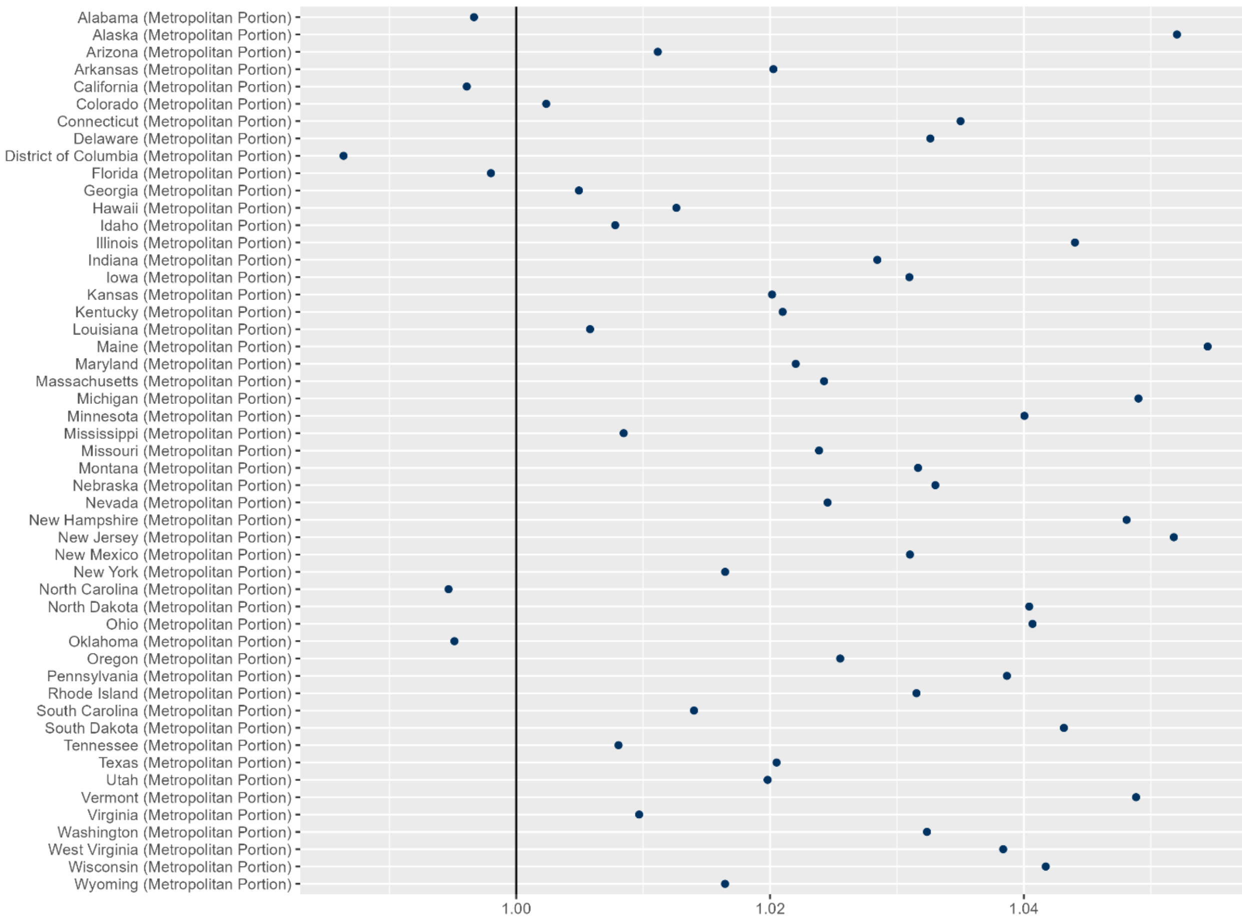


Figure D2: RPP All Items, difference between BEA SARPP data and raw estimates

BEA data divided by raw experimental estimate

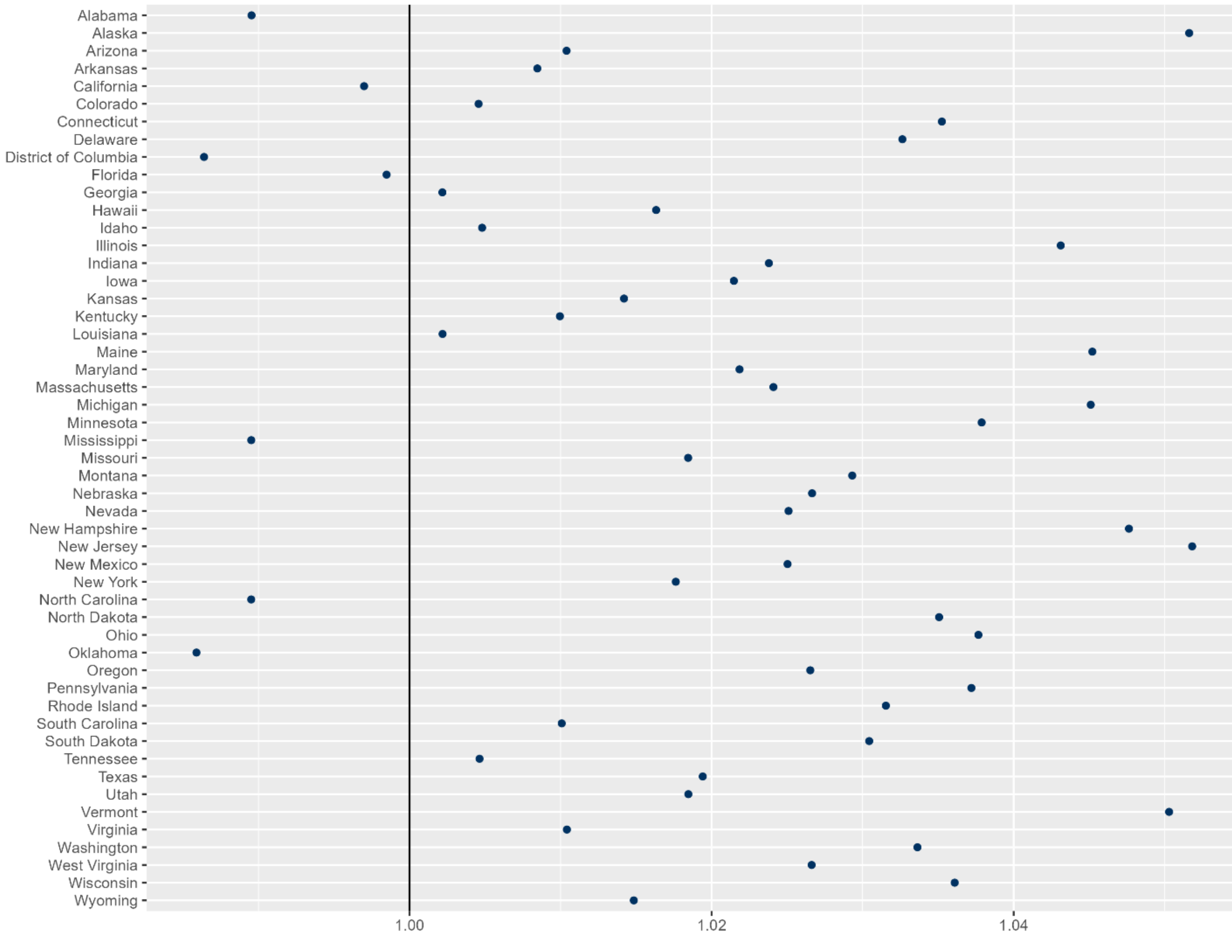
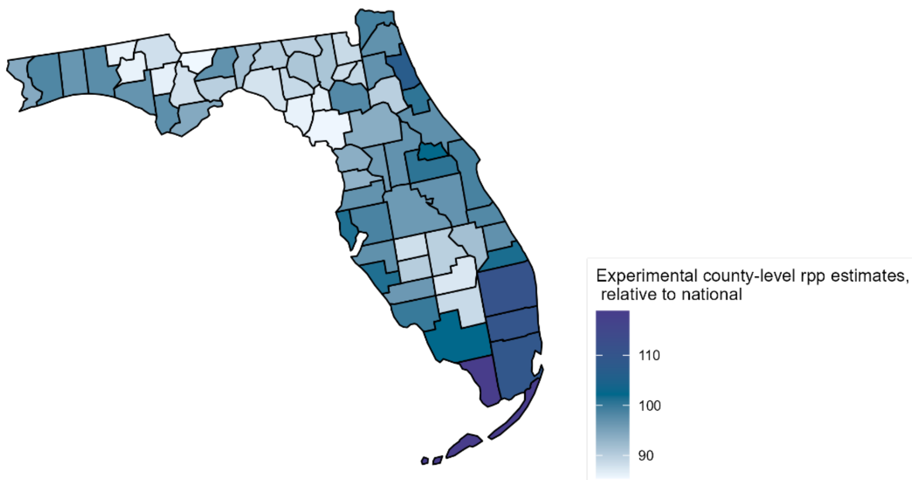


Figure E1 shows estimated RPP All Items for Florida, while directly below it an image from the Florida Department of Education's report on the state's 2022 Price Level Index for the relative cost of personnel in schools. Despite substantial differences, the data are still loosely comparable since there is likely a correlation between costs of school personnel and cost of living. Though the levels differ, similar geographic patterns emerge.

Figure E1: Florida, RPP All Items



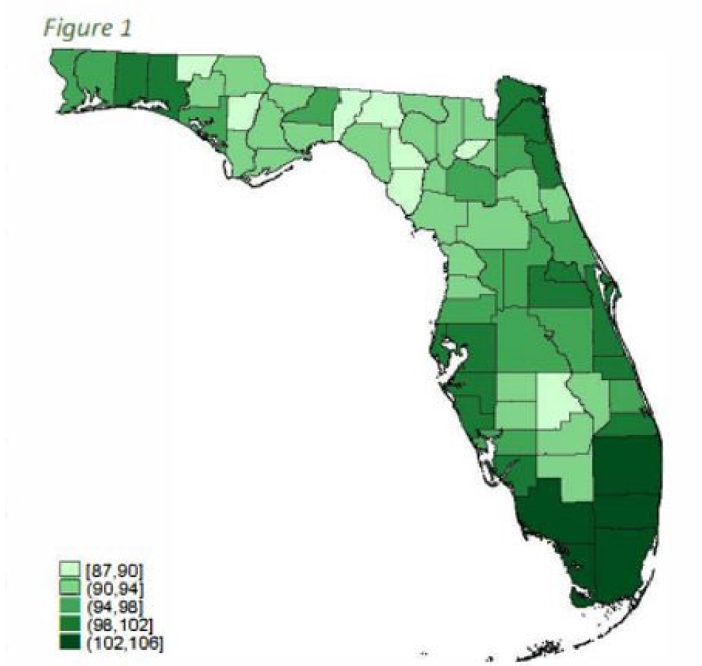


Figure E2 shows estimated RPP All Items for Wyoming, while directly below it an image from a Wyoming Economic Analysis Division report on Cost-of-Living Index for the Fourth Quarter of 2022. While the Wyoming Economic Analysis Division data is relative to the statewide average, it is still loosely comparable. Though the levels differ, similar geographic patterns emerge.

Figure E2: Wyoming, RPP All Items

Experimental county-level rpp estimates, relative to national