A Nontraditional Data Approach to the CPI Gasoline Index

CPI Crowd-Sourced Motor Fuels Data Analysis Project[1]

John Bieler, David Popko, Sarah Niedergall, and Ilmo Sung[2]

December 2019

Abstract

The Bureau of Labor Statistics (BLS) has traditionally relied on manual, in-store observation of prices for goods and services sampled in the Consumer Price Index (CPI). However, given the trend of declining response rates in recent years and the high cost of manual collection, the need to explore other channels and sources of price data has become increasingly paramount. One current study is the Crowd-Sourced Motor Fuels Data Analysis project, which may lead to replacement of the current CPI gasoline sample. These motor fuel price data are obtained via web scraping with the site's permission. These data, which are web scraped from the company's free public database, offer an opportunity to improve the efficiency of price collection for gasoline. Moreover, an equally important benefit arises from using this source, in that it greatly increases the number of available price observations beyond what the current CPI survey methodology and sampling frame can provide.

Since expenditure information is unavailable, the International CPI Manual (ILO, et al., 2004) recommends an unweighted geometric means or a Jevons price index as the aggregator of choice. A Jevons index typically relies upon single price observations for sampled items; however, in this case we used an average price (calculated at each station, for each grade of gasoline) to measure the month-to-month price change. We also created indexes that utilized county level weights, which may allow for a more accurate measurement of price changes within an area.

Our results show that both average prices and price indexes calculated based on the crowd-sourced gasoline data behave similarly to CPI data. Currently, CPI data collectors record certain features of gasoline that may reflect how it is priced. We cannot collect this information from the crowd-sourced app so we are ignoring possible differences due to these gasoline item characteristics, which other than fuel grade, are rather homogenous. Nevertheless, our results suggest that no significant bias arises as a result of disregarding this information. As a source of monthly pricing information, the gasoline data from this crowd-sourced app is a suitable replacement for the CPI field-collected gasoline data. The experimental indexes created have a long-term trend similar to the official CPI for gasoline. Additionally, the larger sample size of the crowd-sourced data captures, to a greater degree, the price change variance exhibited within a geographic zone.

JEL Codes: E00, E31

Keywords: price measurement, price index

---

[1] *The authors thank staff at the Bureau of Labor Statistics for their feedback to this research, especially David Friedman, Rob Cage, Crystal Konny, Nicole Shepler, Kelley Khatchadorian, Anya Stockburger, Brendan Williams, and Mark Bowman. The views expressed herein are those of the authors and not necessarily those of the Bureau of Labor Statistics, or the US Department of Labor.*

[2] *John Bieler (Bieler.john@bls.gov), Sarah Niedergall (Niedergall.sarah@bls.gov), and David Popko (Popko.david@bls.gov) are economists and Ilmo Sung (Sung.ilmo@bls.gov) is a data scientist in the Office of Prices and Living Conditions, U.S. Bureau of Labor Statistics.*

Using Crowd-Sourced Gasoline Prices in the CPI

Introduction

The Bureau of Labor Statistics (BLS) mostly relies on manual collection of its prices for commodities and services, but momentum has increased from both forces external to and within BLS to research and utilize alternative data. It is crucial to employ alternative data for sampling, replacing and/or supplementing current data collection to make our price indexes more accurate, timely, and relevant. To the BLS, alternative data include everything from transaction data and corporate data, to collecting prices via web scraping, Application Programming Interfaces (APIs), or crowd-sourced applications. Since the 1980's, the Consumer Price Index (CPI) has used secondary source data for index calculation, sample frames and comparisons, and supplementing collected data to support hedonic modeling. Secondary source data is the umbrella term we use to describe any data that are gathered by someone other than BLS data collectors. What is new in the current times is the variety and volume of the data sources.

The BLS has other and more pertinent reasons to turn to alternative data. Declining respondent participation rates, restricted access to knowledgeable employees, and a tight budgetary environment in recent years has resulted in a need to explore alternatives to traditional data collection. With these obstacles in mind, the CPI Program has launched an initiative to obtain and analyze alternative price data for inclusion in the index. Alternative data sources provide an opportunity to address many of the challenges encountered by the CPI over the past few decades. In addition to addressing the challenges mentioned above, the adoption of alternative data sources could also potentially increase sample sizes, reflect consumer substitution patterns more quickly, reduce or eliminate respondent burden, help address non-response problems in the CPI's surveys, and reduce collection costs. Over the next several years, the CPI program will rely more on alternative data, in the form for transaction level data, corporate level data, or secondary source data. This project works towards this goal by exploring potential improvements in the accuracy, collection, and composition of the motor fuel and gasoline indexes. In addition, the CPI program is looking for data collection that is cost-effective in order to alleviate the financial burden of manual collection.

Our objective is to examine the web-scraped data from a popular crowd-source website and mobile application to determine if it can be used as a reliable monthly input for the construction of CPI gasoline indexes. We seek to understand whether these price observations have distinct dynamics, their advantages and disadvantages, and whether they could be a dependable source of information for index calculation and publication. We have been collecting daily motor fuel price data from the company's *free* public database since June 2017[3]. This source of alternative data provides us with the opportunity to measure the daily price fluctuations that consumers encounter at the gas pump. We created a number of indexes based largely on the Jevons index methodology. The ultimate goal of this study was to compare these experimental indexes to the published CPI gasoline index, and to determine which one more accurately reflects the price movement of gasoline fuels. Overall, we found that the CPI for gasoline and the experimental crowd-sourced price indexes were very similar, but these alternative data may better capture the fluctuations of gasoline prices due to the greater number of price observations and extended geographic coverage.

---

[3] The CPI has obtained verbal permission to extract data from the company's website for research purposes. We may also share the data with other federal statistical agencies that abide by our confidentiality agreement.

Currently, even when collecting information from websites, CPI data collectors manually enter data into a hand-held tablet with a customized app for data entry. A list of price quotes is downloaded to the tablet for each outlet in the CPI sample every month. The data collector records the new month's price, along with any changes to the item's characteristics, and electronically submits the data to BLS Washington headquarters for review.

The CPI is exploring using web scraping to automate data collection from these websites instead. Many international government statistical agencies have begun using web scraped data for price index calculation. For example, the UK's Office for National Statistics is planning to use web scraped data for about 25 categories by 2023.[4] Statistics Canada has also commenced a CPI Alternative Data Sources (ADS) Initiative in recent years, and hopes to replace 20 percent of their field collected prices with alternatively sourced data.[5]

The most popular example of a consumer price index created using web scraped data is happening in academia. MIT's BPP has demonstrated the benefits of using web scraping to collect massive amounts of data for the purposes of price measurement.[6] Employing only a set of web collected price data, they have been able to build and maintain customized datasets that fit specific measurement and research needs.

Sample selection for alternative data is just one issue facing the CPI, along with item replacement, publication timeliness when dealing with multiple data sources, and creating or updating security/confidentiality guidelines. The principal upfront cost is planning and building an IT infrastructure that will meet all or most data source needs. These concerns have been documented, and are currently being addressed by CPI leadership. [7]

This paper reviews BLS efforts to replace its traditional CPI gasoline survey with price observations obtained from the crowd-sourced company's website and app. To begin, we review the CPI motor fuel sample and how the data are currently collected before describing the gasoline alternative data. Next, we discuss the index methodologies used to calculate experimental gasoline indexes using the crowd-sourced data. Then, we review geography and the degrees of timeliness that are possible with these alternative data throughout the index methodology section, before concluding.

CPI Data for Gasoline

The CPI monitors and publishes a monthly motor fuel index that measures the average price change over time of gasoline and other motor fuels. Price information is collected continuously for each category of motor fuel throughout a given month, with each sampled quote is only collected once during the month. The data are sent to the BLS national office and reviewed for accuracy by CPI analysts. The collected prices are per-gallon unit prices and include excise, sales, and other taxes paid by consumers.

---

[4] https://www.ons.gov.uk/economy/inflationandpriceindices/articles/usingalternativedatasourcesinconsumerpriceindices/may2019#introduction

[5] https://unstats.un.org/unsd/bigdata/conferences/2017/presentations/day2/session2/case2/3%20-%20Jon%20Wylie%20-%20Statistics%20Canada.pdf

[6] http://www.thebillionpricesproject.com/our-research/

[7] Konny, C., Williams, B., and Friedman, D. 2019. Big Data in the U.S. Consumer Price Index: Experiences and Plans. A chapter in *Big Data for the 21st Century Economic Statistics*. Forthcoming from University of Chicago Press. Retrieved from: https://www.nber.org/chapters/c14280.

The CPI outlet sample for all grades of gasoline currently consists of approximately 1,300 outlets, which are selected from the Telephone Point-of-Purchase Survey (TPOPS) conducted by the Census Bureau for BLS.[8] In this survey respondents report how much they spend at particular stores for particular items – in this instance, how much they spent on gasoline and where.

The data from the TPOPS are then used to select the retail establishments at which the CPI will then monitor the prices of sample selected goods and services.[9] Each outlet selected for gasoline pricing is assigned a price quote for each grade of gasoline, as well as for other motor fuels such as diesel and alternative fuels. These quotes will then be tracked as to their respective price fluctuations over the subsequent months. The current CPI sample has approximately 4,300 price quotes for gasoline products nationally, and just over 800 price quotes for other motor fuels. This equates to an average of 55 quotes per metro area in the CPI geographic sample. All grades of gasoline are usually available each month for collection; fewer outlets offer information on diesel and alternative fuels.

The Consumer Expenditure (CE) survey provides data on expenditures, income, and demographic characteristics of consumers in the US, and is used by the CPI to weight each item in the CPI and the areas within that item.[10] Both the biennial aggregation weights and the relative importance values for motor fuels are estimated from CE data. The CPI currently collects gasoline prices in 75 cities across the United States.[11]

CPI Item Title	CPI Item Code	Relative Importance
Transportation	SAT	16.348
Public transportation	SETG	1.113
Private transportation	SAT1	15.235
Motor fuel	SETB	3.762
Gasoline (all types)	SETB01	3.671
Regular	SETB011/SS47014	N/A
Midgrade	SETB012/SS47015	N/A
Premium	SETB013/SS47016	N/A
Other motor fuels	SETB02	0.091
Diesel	SETB021	N/A
Alternative Fuels	SETB022	N/A

Table 1. Relative Importance of the CPI for Motor Fuels as of December 2018

For elementary area indexes, the CPI uses the geometric mean (equation 1) index formula, which approximates a cost-of-living-index, to calculate within area indexes for gasoline, using the TPOPS outlet weights. Next, a Laspeyres (equation 2) formula is employed to aggregate the area indexes into a national index using the CE expenditure weights. The CPI publishes 32 area indexes for gasoline, 23 larger cities and nine regional indexes defined by the nine Census divisions.

[8] https://www.bls.gov/respondents/cpi/tpops/home.htm

[9] TPOPS was retired as a survey in September of 2019; since that time, BLS has used data reported in the Consumer Expenditure Survey (CE) to select the establishments from which its sample is derived. The entire scope of this research was carried out under the TPOPS framework.

[10] https://www.bls.gov/cex/home.htm

[11] https://www.bls.gov/opub/mlr/2016/article/the-2018-revision-of-the-cpi-geographic-sample.htm

Equation 1: The CPI employs a weighted geometric mean formula to calculate price change at the elementary area level. For each area *a*, the price relative *R* from time *t-1* to time *t* is calculated by taking the weighted geometric mean across all price observations in *a* under the following relation:

$$R_{[t,t-1],a} = \prod_{s,g \in a} \left(\frac{P_{s,g,t}}{P_{s,g,t-1}}\right)^{\frac{W_{s,g}}{\Sigma W}}$$

Where:

$$P_{s,g,t} = observed\ price\ of\ fuel\ grade\ g\ at\ station\ s\ at\ time\ t$$

And:

$$W_{s,g} = weighted\ expenditure\ for\ grade\ g\ at\ station\ s$$

This relative is multiplied by the index level I_{t-1} at time *t-1,* to obtain the index level at time *t.*

Equation 2: Using the results of the single area indexes calculated in equation 1, the CPI uses a Laspeyres methodology to obtain aggregate level indexes at the regional or national level. For each area *a* in *A*, we apply the following formula to calculate the aggregate level relative:

$$R^L_{[A,t-1,t]} = \frac{\sum_{a \in A} I_{a,t} W_a}{\sum_{a \in A} I_{a,t-1} W_a} = \frac{I^L_{A,t}}{I^L_{A,t-1}}$$

Where:

$$I_{a,t} = index\ level\ for\ gasoline\ at\ time\ t\ in\ area\ a$$

And:

$$W_a = CE\ reported\ expenditure\ weight\ in\ area\ a$$[12]

Crowd-Sourced Gasoline Data

The BLS collects data by downloading price information from the crowd-sourced website and app on a daily basis. With the company's permission, we have been collecting daily motor fuel price data for regular, midgrade, premium, and diesel from their website since June 2017. The automated data collection we implemented is different from usual web scraping programs, where one captures information on a web browser or html source files. Instead, our computer program automatically pulls prices and gas station information displayed on the site's mobile app. This method of web scraping data provides us with the opportunity to measure the daily price fluctuations that consumers encounter at the gas pump. The table below lists the variables collected, with a brief description. In terms of alternative data, the number of variables we collect from this company is rather small.

#	Variable	Variable Description

[12] Under CPI methodology, this would refer to the aggregation weight, derived from CE but then inflation adjusted to a pivot month (December 2017 in this case).

1	Zip Code	Gas Station Zip Code
2	Station ID	Gas Station ID
3	Name	Name of Gas Station.
4	Price	Price of gas at time of collection.
5	Fuel Type	Type of fuel for price collected.
6	Address	Address of Gas Station.
7	Longitude	Longitude of Gas Station
8	Latitude	Latitude of Gas Station
9	Posted Time	The time the data was posted to website or mobile app
10	Collection Time	The time the data was collected from website or mobile app
11	Hours	The number of hours between the posted time and collection time.

Table 2. Crowd-Sourced Data Variables and Descriptions

Between November 2017 and May 2019 we collected 120 million gasoline observations from the crowd-sourced website and app. After the data are collected, we remove all duplicate observations. As a result of duplicate removal, the number of observations decreased to 98 million, or an 18% reduction.

Figure 1. Crowd-Sourced Data Observations, Nov. 2017-May 2019

In order to calculate a gasoline index that conforms to the CPI's geographic structure, we mapped each gas station to the CPI's Primary Sampling Units (PSUs), which are defined as core-based-statistical-areas. We were then left with 56 million observations that mapped to the CPI geographic structure, or about 57% of our original sample after removing duplicates. The sample was restricted to this geographic structure in order to compare our experimental indexes with the published CPI numbers. The data outside CPI PSUs will be used to create rural county gasoline indexes in a future endeavor.

Though specifics are not given, the crowd-sourced company does have a system in place to remove any information that they believe is inaccurate. By offering incentives such as discounts on fill-ups, they want to encourage as many people as possible to post gas prices, which should make the occurrence of inaccurate price data rare. The website and mobile app also asks users to report incorrect prices or suspicious accounts that are posting inaccurate prices. Average prices per gas station were calculated, and the standard deviations were found for each fuel grade in each PSU. Any average price that was outside ±3 standard deviations was removed from index calculation. On average, one percent of the data were removed as outliers each month.

After outlier removal, arithmetic average prices for each fuel type were calculated per gas station in a given month, which allowed for equal weighting of gas stations within a PSU. Each month contained approximately 2.5 million observations. We calculated about 140,000 average prices for gasoline each month – 50,000 for regular fuel, 44,000 for midgrade fuel, and 46,000 for premium fuel.

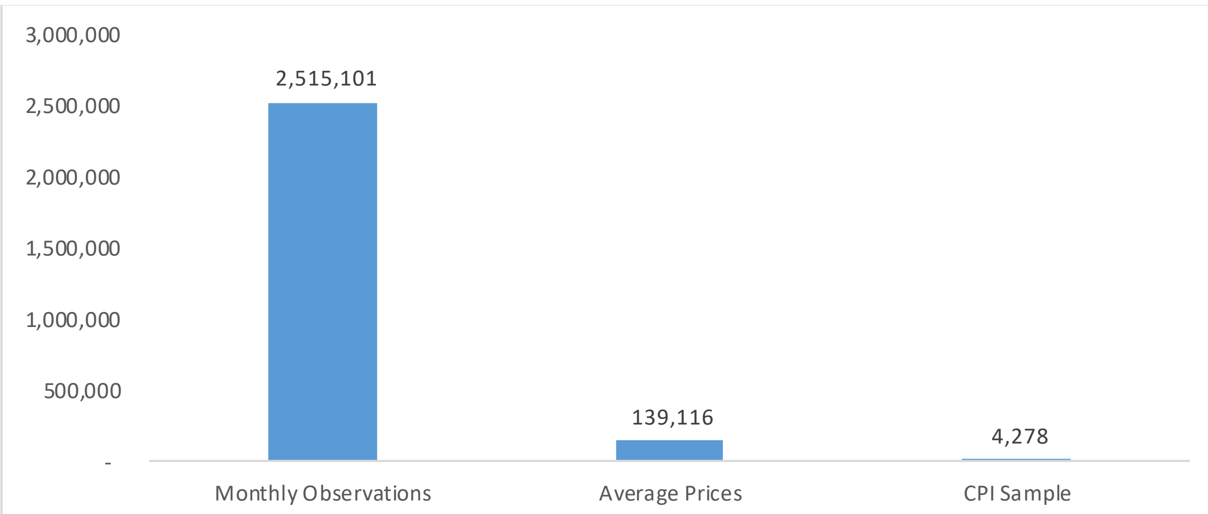

Figure 2. Crowd-Sourced Data and CPI Sample Comparison

The crowd-sourced data are significantly larger than the CPI in terms of volume of data points. However, the CPI uses a systematic sampling methodology based on expenditure types, so that the sample includes both large and small volume gas stations. The web scraped price data, meanwhile, are a convenient sample based on what users are observing and entering into the company's website and mobile app.

Index Methodologies and Results

It was determined that the process of including the alternative gasoline data into the CPI system needed to occur in two phases. First, the crowd-sourced data was transformed in order to render it suitable for the CPI system. This data transformation was done in two steps: 1) the arithmetic mean of prices for a given month, fuel type, and PSU was calculated, and 2) after the initial month, average prices were updated using a Jevons relative for each subsequent month.

The second phase of this process used those prices to calculate an aggregate gasoline index using a geometric mean formula while adhering to the current CPI geographic structure. The following sections examine this approach in greater detail.

Experimental Crowd-Sourced Data Index Methodology

We used the CPI sample of fuel types and PSUs as a means of inputting prices from the crowd-sourced data. The BLS research team transformed the crowd-sourced data into a suitable format for the CPI database structure; following this, we could recalculate the CPI for gasoline using the web scraped collected prices.

Using the CPI sample as a structure to input the crowd-sourced price data allows us to revert to the original CPI in the event that the crowd-sourced data was rendered unavailable. Furthermore, the CPI sample also consists of observations based on out-of-town gasoline expenditures. Using the CPI sample allows us to input prices for out-of-town trips. This issue is discussed in greater detail in the future research section of this paper. Finally, the CPI sample already contains weights that reflect the proper proportion of expenditures between the different fuel types within an area.

In the base month of November 2017 (the first full month of crowd-sourced prices we have without duplicates), we used the average price across all gas stations for each fuel type within a given CPI PSU to calculate our experimental indexes. For example, if an observation for regular gasoline was sampled in New York City in the CPI, we calculated the arithmetic average price for regular gasoline across all gas stations within New York City in the crowd-sourced data, and replaced the CPI price with a crowd-sourced price. Fuel types are equally weighted within a PSU according to the Jevons index methodology. Note that the crowd-sourced gas station prices are arithmetic averages taken over the course of the month.

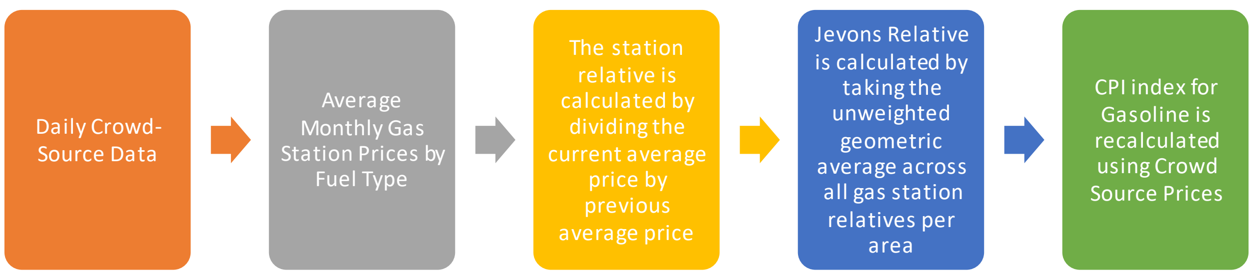

Figure 3. Experimental Index Calculation Flow

Next, a Jevons (see equation 3) relative was calculated using the average price data for all outlets for a particular fuel type and area during the time period studied. The previous month's price for the PSU was then updated using the calculated Jevons relative to calculate the current month's price. After price creation, we then recalculated the index for gasoline using the crowd-sourced data.

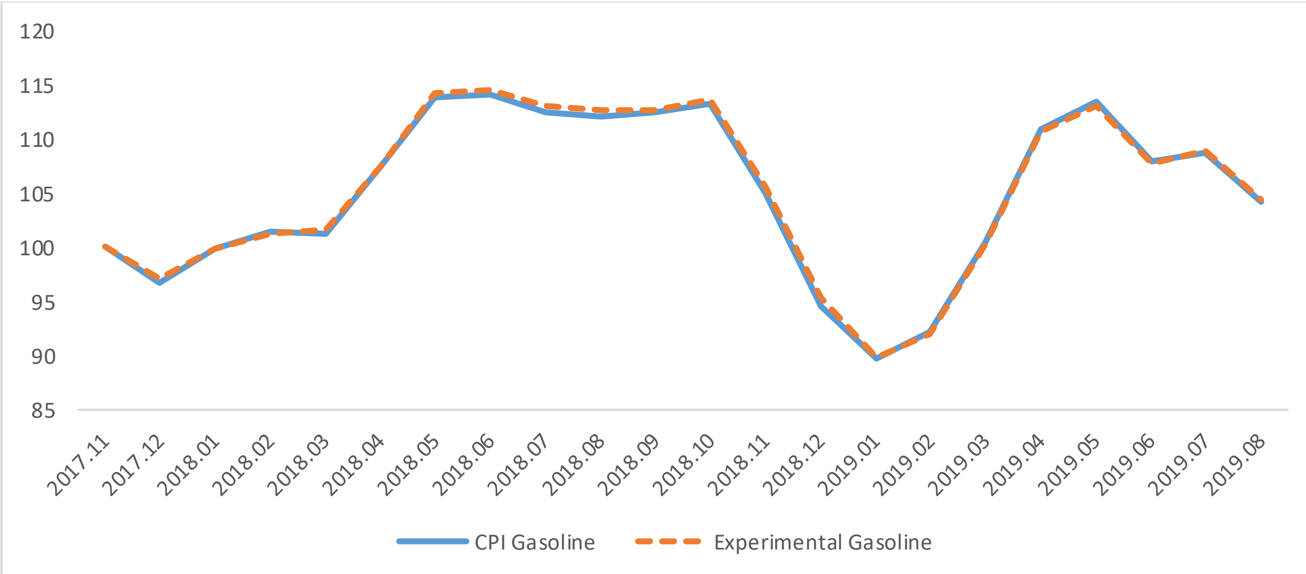

Figure 4. U.S. CPI for Gasoline and Experimental Crowd-Sourced Data Index, Nov. 2017-Aug. 2019

A Wilcox sign rank test was used to compare the index relatives between the official CPI gasoline index and the experimental index at the U.S. level and found no significant difference. This test was also performed on area level indexes, which can be found in Appendix 1.

Table 3. Wilcoxon signed rank test results, CPI for Gasoline versus experimental indexes, U.S. level, Nov. 2017-Aug. 2019

Area Code	Area Name	Number of Observations	Probability
0000	U.S. Level	20	0.98544

Equations 3 and 4: Our research applications employ a Jevons aggregation methodology at the single area level for each fuel type. The Jevons methodology is similar to the geometric means method outlined in equation 1; however, the Jevons geometric mean is an unweighted one. For each area, the gasoline price relative *R* is calculated as follows:

$$R^{J}_{[G,t-1,t]} = \prod_{g \in G} \left(\frac{P_{g,t}}{P_{g,t-1}} \right)^{\frac{1}{n}}$$

Where:

$$P_{g,t} = the\ price\ observation\ at\ station\ g\ observed\ at\ time\ t$$

However, for our research purposes, we use a modified Jevons approach; rather than taking a geometric mean across each individual price observation, we take the sample arithmetic mean of all prices reported for a given station in a given month, as illustrated below:

$$R^{J}_{[G,t-1,t]} = \prod_{g \in G} \left(\frac{\bar{P}_{g,t}}{\bar{P}_{g,t-1}} \right)^{\frac{1}{n}}$$

Where:

$$\bar{P}_{g,t} = the\ sample\ mean\ of\ all\ prices\ observations\ observed\ at\ station\ g\ observed\ in\ month\ t$$

As such, the area price index relative becomes a geometric mean taken across the sample mean prices of all stations within a given CBSA (area) boundary, weighted only according to respective fuel grade.

County-level Weighting

An additional weighting approach attempted to account for geography-based variations in price change at a level more granular level than that found in traditional CPI methodology. This work stemmed from a previous research effort using the crowd-sourced data, entitled *Does Location Matter? A Case-study in the Effect of Geography on Gasoline Price Inflation*[13]. The study used a fixed-effects regression model calculated for a single area, to evaluate CPI sampling methodology. The regression accounted for geographic (as represented by county) variables, as well as numerous economic considerations (home value, population density, etc.). Overall, the regression found that the above variables were significant in explaining price change variation at a cross sectional level. However, controlling for said variables in an experimental sampling methodology, did not yield area level indexes that were significantly different from that calculated using the baseline CPI sampling methodology.

We attempt to expand on the research in *Does Location Matter?* However, rather than controlling for geographic considerations in a sampling capacity, we attempt to observe the effect of weighting at a more granular geographic level.[14] Under current CPI practice, area level weights are calculated at large, without controlling for populations (or expenditure patterns) of respective counties within a given CBSA. This methodological choice is largely the result of sample-size constraints; the resources required to effectively sample and track prices of goods and services at a level more granular than the above are too costly. Using the crowd-sourced data, however, we are able to remove this constraint. The additional observations effectively reduce the cost of collection to zero. As such, we are able to construct a weighting scheme that accounts for price-change fluctuations *within* a geographic area, as well as across geographic areas.

We estimated weights for each county in all 32 CPI areas, and subsequently constructed indexes aggregating first at this county level, then the PSU/Area level, and finally on a nationwide basis. Using CE household level data, we calculated these weights on both plutocratic (weighted by dollar gasoline expenditure amounts multiplied by CE population weight[15]) and democratic (weighted by the number of cars owned by a household multiplied by population weight) bases.

[13] Popko, David. Sung, Ilmo. *Does Location Matter? A Case-study in the Effect of Geography on Gasoline Price Inflation.* Joint Statistical Meetings, 2019, Denver, CO

[14] For additional information on generalizing the findings in *Does Location Matter?,* and on price movement variation within and across areas, please see appendix 2.

[15] Population weight is equal to the estimated number of households represented by an individual surveyed consumer household.

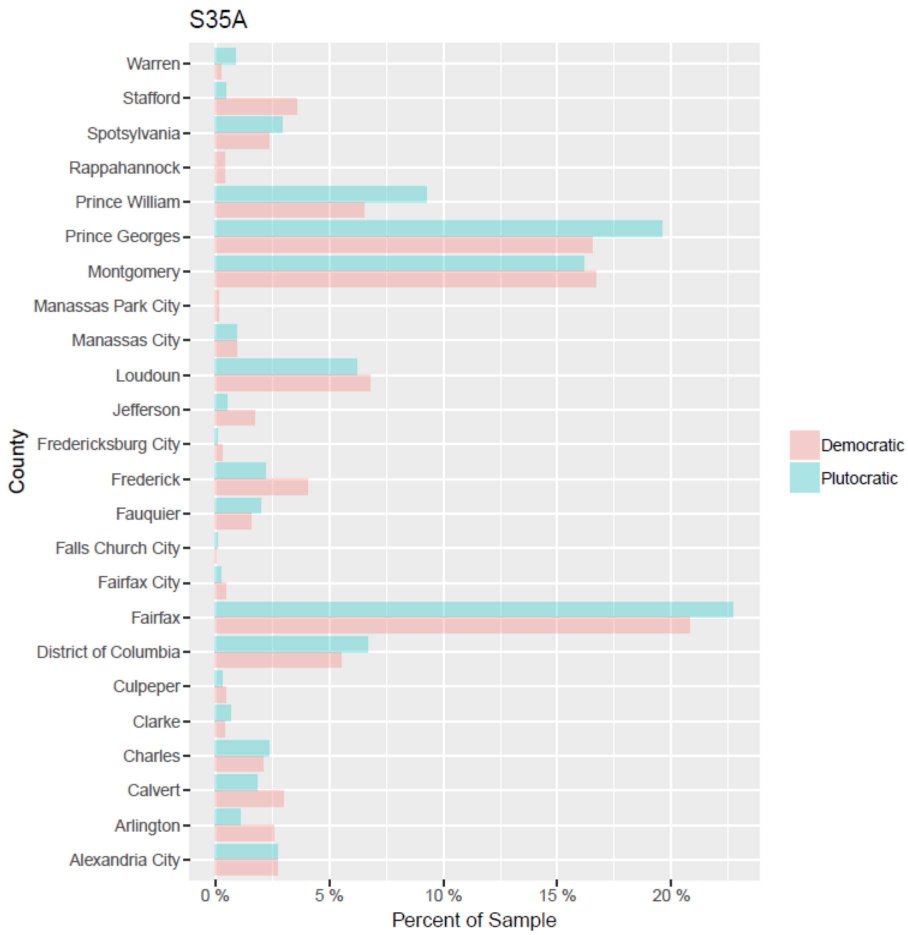

Figure 5. Comparison of county level weights for the Washington-Arlington-Alexandria, DC-MD-VA-WV CBSA, constructed on a plutocratic and democratic basis.

Using these county level weights as described above, we then proceeded with the index methodology outlined in *Alternative Data for CPI Motor Fuels*[16] that weighted observations according to fuel grade. Using this approach, we constructed indexes first at the county level, and then aggregated them into area indexes using the county level weights, and finally a national index. We then used a paired t-test to compare these index relatives with relatives calculated with area weights produced at-large.

[16] Bieler, John. Niedergall, Sarah. Popko, David. Sung, Ilmo. *Alternative Data for CPI Motor Fuels, p. 6, 2019*

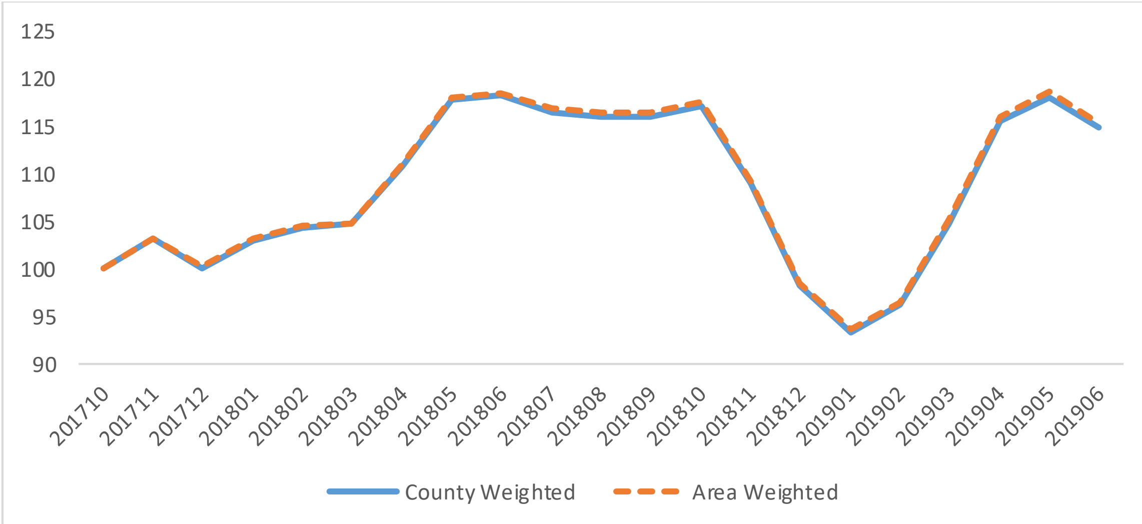

Figure 6. Comparison of trajectories for experimental gasoline indexes calculated with county level weighting (plutocratic), vs standard area level weighting.

Our paired t-test results did not show significant differences between index relatives dependent on weighting at the county or area level. Table 2 in the appendix shows the sum of squared error between the CPI index and the experimental index.

Area and Region-level Variance of Long Term Index Relatives

We briefly performed an analysis of long term relatives across individual areas – to answer whether gasoline prices move differently across geography on a long term basis. Our long term relative was taken across all stations in each of the eligible CPI CBSAs, that observed a price in the first month (November 2017) and the final month at the time of this writing (September 2019). We took the variance across all stations within each of these CBSAs. We also performed these analyses at the region level.

On average, non-self-representing areas – CBSAs which comprise a number of smaller cities, showed greater variance across their respective stations, than their single city counterparts. We suspect that this is because these non-self-representing areas are indeed classified over wider geographic regions. Similarly the “West” census region showed the greatest variance across station relatives by far. In addition to having greater geographical coverage than any of the other three regions – which all show similar variance levels – the Western region contains Alaska and Hawaii, which may contribute to this difference.

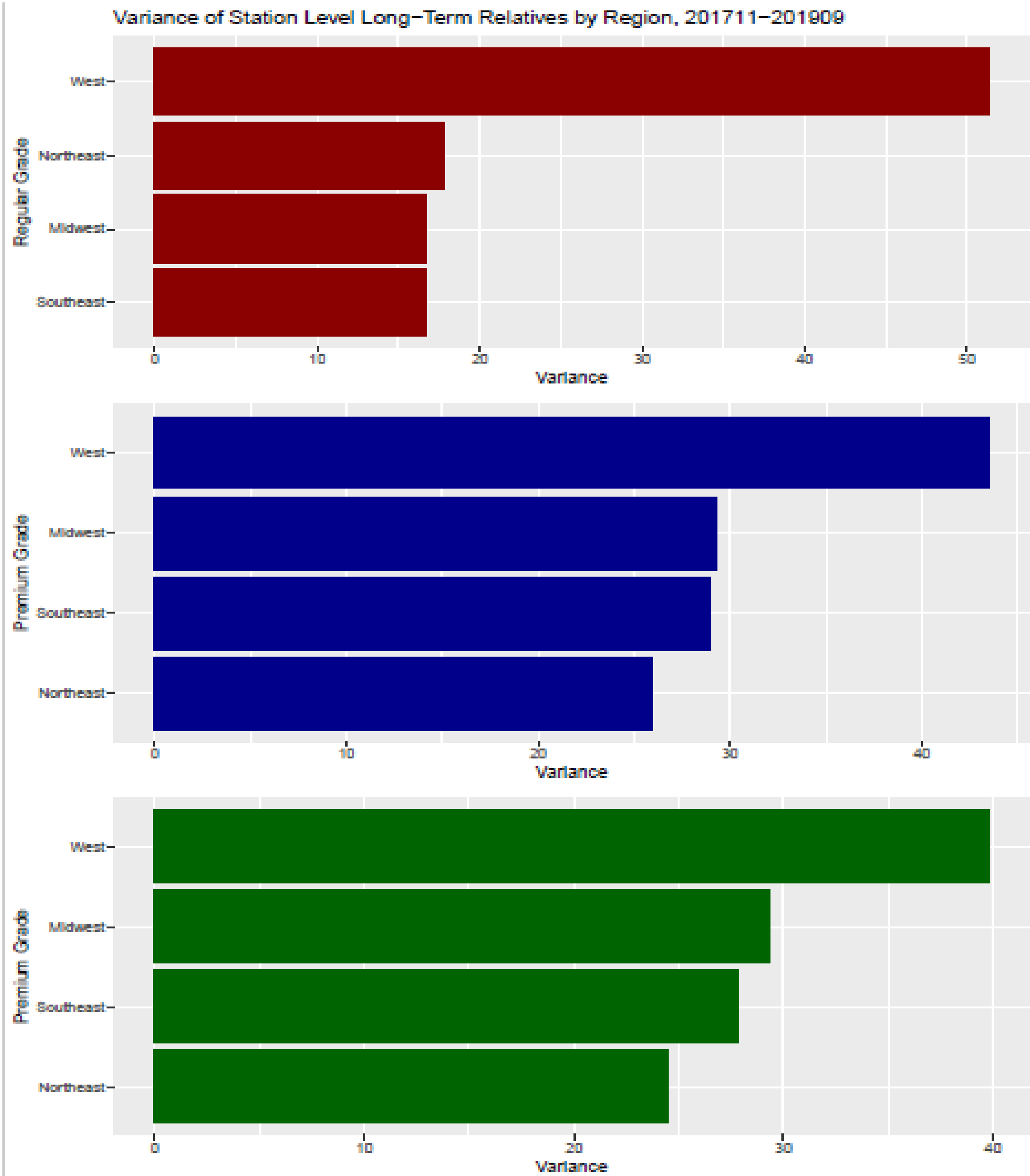

Figure 7. Long-term Variances of Gas Stations by fuel grade per region, Nov. 2017-Aug. 2019

Future Research

Accounting for Gasoline Expenditures on Out-of-Town Trips

The sheer volume of data available from the crowd-sourced company's website and app – relative to the CPI's currently sample – makes it conducive for exploring possible future improvements in CPI index methodology. We also explored the improvement of accounting for out-of-town gasoline expenditures in area index estimation.

The CPI for a given item – in our case, gasoline – is weighted geographically, across 32 areas in the United States; of the 75 cities in which the CPI collects prices, 23 larger cities represent themselves as areas; the remaining cities are bundled into 9 regionally based areas. Each area's experimental crowd-sourced gasoline price index currently estimates the change of price in gasoline *within that area*. By definition, this is different than an index estimating the change of price in gasoline experienced *by consumers living within said area*. The latter would require accounting for consumers' gasoline purchases made outside a given area, such as those made while on trips.

The CE collects household's expenditures on gasoline, differentiating between gasoline purchased in a consumer's area of residence, and that purchased on trips (along with the trip destination). On average, these out-of-town reports account for approximately 7% of total gasoline expenditures reported in the CE, yet approximately 20% of respondents on average report these out-of-town purchase each quarter. As such, to achieve a representative price index for price change experienced by consumers in a given area, it would be necessary to account for the weight of these purchases. Up until now, this has been impossible from a collection standpoint; the resources simply do not exist to track prices at all of consumers' trip destinations. With a crowd-sourced based sample, these collection costs are effectively reduced to zero, rendering an index of this sort possible.

Equation 5: We propose a methodology that treats the gasoline price relative in month t of a given area a as the weighted geometric mean of a gasoline-in-town relative, and a gasoline-out-of-town relative:

$$REL_{t,a} = \left(REL_{t,a,in}{}^{\frac{W_{in}}{W_{in}+W_{out}}}\right)\left(REL_{t,a,out}{}^{\frac{W_{out}}{W_{in}+W_{out}}}\right)$$

The in-town-relative would be calculated according to the crowd-sourced data/CPI methodology outlined above. Similarly, the out-of-town relative would calculate a relative for each out of town destination reported by CE respondents in an area[17]. These relatives would be weighted in accordance with the relative expenditure weights by each respondent in the area, to comprise a final out-of-town gasoline relative for a given area:

$$REL_{t,a,out} = \prod_{d \in D} rel_d^{\left(\frac{w_d}{\sum_{d \epsilon D} w_d}\right)}$$

Where d and D signify respectively each destination in the set of all trip destinations reported by consumers of that area. In this manner, we explored an effective use of the large volume of data from the crowd-sourced website and mobile app.

Experimental Daily Index

We also calculated experimental daily indexes for gasoline using the crowd-sourced data. Figure 6 compares the results of our experimental daily indexes to the CPI monthly index for gasoline. The BLS research team again used a Jevons relative to calculate the individual fuel type indexes. However, we used the CPI TPOPS weights to calculate the gasoline daily index, which allows for the weighting for each individual fuel type within each area. On average, respective fuel types are weighted approximately to 85% regular, 10% premium and 5% midgrade across all areas, though minor variations exist across areas.

[17] Destinations outside the scope of CPI areas could be treated in one of two ways: (1) calculate a relative for each destination, defined by City/State/County, regardless of its inclusion in the CPI scope, or (2) creating a catch-all "area" for all areas outside the CPI scope.

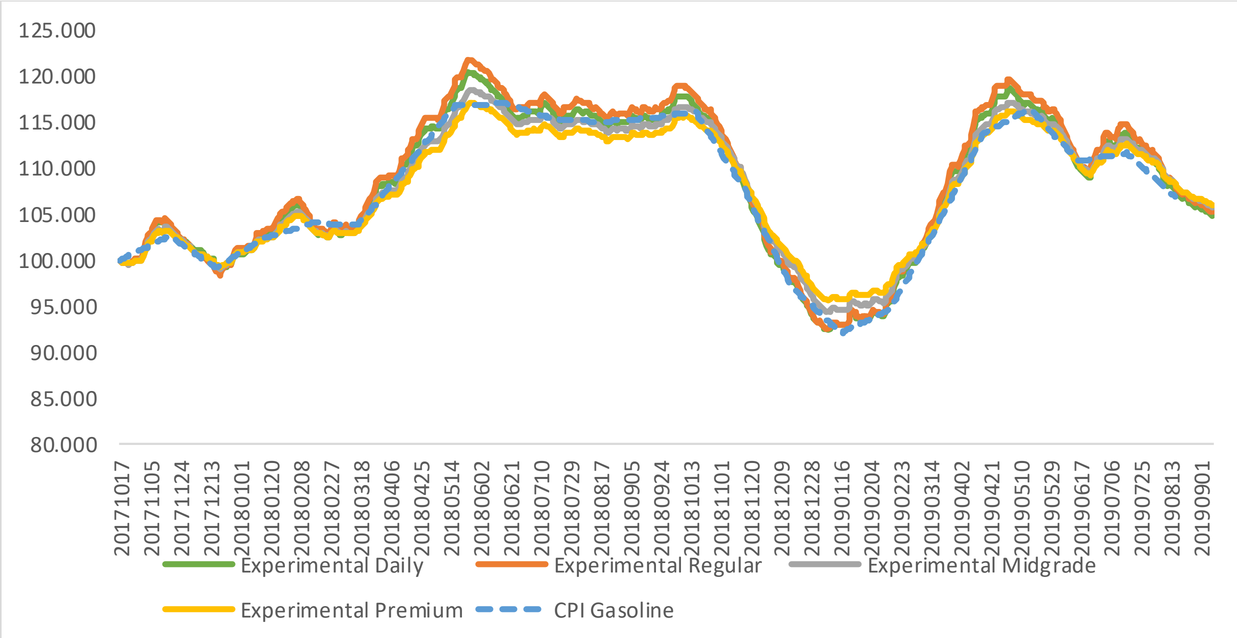

Figure 8.Experimental Indexes and U.S. CPI for Gasoline

Conclusion

The CPI has been constructed primarily using data collected by BLS staff. Web scraped data have the potential to address many of the problems faced in recent years including lower response rates and higher collection costs.

This paper reviewed the results of an experimental Jevons index and experimental county weights using crowd-sourced data. CPI has the choice of remaining with the weighting information extracted from the TPOPS survey, calculating an index with equal weighting for the price relatives within an area, or employing these county weights. Looking at the indexes and price changes, we found that the experimental crowd-sourced indexes and CPI move in a similar pattern. By performing statistical tests, we can conclude that these indexes are not different. These results are very promising.

We found that the crowd-sourced data are a viable source for gasoline price data, and recommend replacing the CPI gasoline prices with prices from this company. Gasoline is relatively simple to price via traditional data collection, but with the crowd-sourced data, the CPI can expand its sample and coverage. Future research is planned, and we will continue to evaluate results in order to make a decision on use for official index calculation in the upcoming year.

Appendix 1

Table 4. Descriptive Statistics of Crowd-Sourced data, Nov. 2017-Aug. 2019

Month	Grade	N	Mean	Std Dev	Minimum	Maximum
201711	Regular	59,741	2.5514576	0.3028627	1.94	4.769
201712	Regular	59,671	2.4732454	0.3057961	1.69	4.79
201801	Regular	59,875	2.5469586	0.2951578	1.94	4.34
201802	Regular	59,615	2.5817516	0.3323173	1.69	4.75
201803	Regular	59,842	2.5891745	0.340588	2.02	4.49
201804	Regular	60,044	2.7447399	0.3323332	1.99	4.79
201805	Regular	60,093	2.9244437	0.3214268	1.89	4.88
201806	Regular	59,855	2.9344972	0.3355906	1.99	4.93
201807	Regular	59,802	2.888884	0.33166	2.09	4.93
201808	Regular	59,718	2.8752498	0.3230829	2.1	4.89
201809	Regular	59,642	2.8741001	0.3329934	2.05	4.93
201810	Regular	59,578	2.8929676	0.3752693	2.21	4.99
201811	Regular	59,561	2.6795102	0.4392769	1.8866667	4.99
201812	Regular	59,780	2.416614	0.4713839	1.6158065	5.312
201901	Regular	59,782	2.2931156	0.4473711	1.6266667	5.255
201902	Regular	59,559	2.3581609	0.3912051	1.67	5.2488889
201903	Regular	60,000	2.5756508	0.3336012	1.79	4.99
201904	Regular	60,316	2.8490011	0.419852	1.85	5.2933333
201905	Regular	60,024	2.9124549	0.457127	2.14	5.8071429
201906	Regular	60,028	2.7708705	0.4502034	1.69	5.866
201907	Regular	60,069	2.8052265	0.3960812	1.99	5.49

201908	Regular	59,884	2.6837729	0.4060372	1.89	5.6054545
201711	Premium	52,917	3.0720719	0.2781083	2.19	4.69
201712	Premium	52,729	3.0075021	0.2761734	1.75	4.5941667
201801	Premium	52,822	3.0639923	0.272587	2.15	4.598
201802	Premium	52,182	3.1053314	0.2919481	2.25	4.79
201803	Premium	52,989	3.1113666	0.3009115	2.23	4.99
201804	Premium	53,302	3.2491138	0.2996686	2.27	4.99
201805	Premium	53,440	3.4242152	0.2940198	2.39	5.09
201806	Premium	53,089	3.4602444	0.2978662	2.4	5.19
201807	Premium	52,903	3.416683	0.2951787	2.438	5.19
201808	Premium	52,728	3.4060816	0.2896549	2.4166667	5.0691667
201809	Premium	52,526	3.4072914	0.2936408	2.41	5.09
201810	Premium	52,658	3.4318815	0.3242315	2.5355556	5.29
201811	Premium	51,650	3.2532789	0.3755949	2.09	5.19
201812	Premium	51,238	3.0029406	0.4083724	1.6736364	4.99
201901	Premium	51,386	2.8704505	0.3867681	1.6955556	4.97
201902	Premium	51,568	2.9158193	0.3442658	1.7318182	4.965
201903	Premium	52,652	3.1026433	0.3059433	2.09	5.24
201904	Premium	52,661	3.3744179	0.3815084	2.38	5.415
201905	Premium	51,518	3.4579614	0.4077815	2.4314286	5.4233333
201906	Premium	51,365	3.3357436	0.400101	2.1841176	5.39
201907	Premium	51,268	3.3557151	0.3626946	2.2761111	5.19
201908	Premium	50,708	3.2578221	0.3624046	2.0745455	5.19

201711	Midgrade	52,023	2.8360091	0.2860627	2.01	4.5855556
201712	Midgrade	51,821	2.7702946	0.2886764	1.75	4.49
201801	Midgrade	52,087	2.8248314	0.283632	2.06	4.4370588
201802	Midgrade	51,271	2.8664571	0.3142672	2.02	4.49
201803	Midgrade	52,115	2.8724342	0.3241507	1.94	4.69
201804	Midgrade	52,636	3.0113228	0.3203622	2.2442857	4.89
201805	Midgrade	52,980	3.1849763	0.3107065	2.3	4.95
201806	Midgrade	52,181	3.2216616	0.3161748	2.3	4.9453846
201807	Midgrade	52,026	3.1777511	0.3136373	2.32	4.95
201808	Midgrade	51,877	3.1656666	0.3064508	2.19	4.95
201809	Midgrade	51,905	3.1656067	0.3139979	2.04	4.99
201810	Midgrade	51,575	3.190896	0.3519355	2.23	4.9528571
201811	Midgrade	50,661	3.0092169	0.4125036	1.83	4.89
201812	Midgrade	50,404	2.7548339	0.4501541	1.64	4.89
201901	Midgrade	50,745	2.6214958	0.4304939	1.665	4.8557143
201902	Midgrade	50,382	2.6687168	0.3783243	1.7733333	4.89
201903	Midgrade	51,721	2.8583421	0.3266609	1.94	4.7366667
201904	Midgrade	51,911	3.1311543	0.4077838	2.29	5.325
201905	Midgrade	50,723	3.214557	0.4384146	2.29	5.4357143
201906	Midgrade	50,376	3.0907946	0.430085	2.15	5.19
201907	Midgrade	50,502	3.1105763	0.3838789	2.04	5.09
201908	Midgrade	49,971	3.0106124	0.3890372	2	5.09

Between December 2017 and August 2019, there was an average of 428 gas stations added each month (or 0.8% of the priced stations). During that same time, there was an average of 2,102 stations reported as missing (or 3.9% of the priced stations). The prices of these missing stations were imputed indefinitely in our experimental indexes, or until a price was reported in later months. The average length of time an observation went missing was 7.78 months (out of a possible 22 months). Of those that went missing, 91% returned in a later month.

Figure 9. Gas Station Turnover for experimental indexes using data from crowd-sourced website and app, Nov. 2017-Aug. 2019

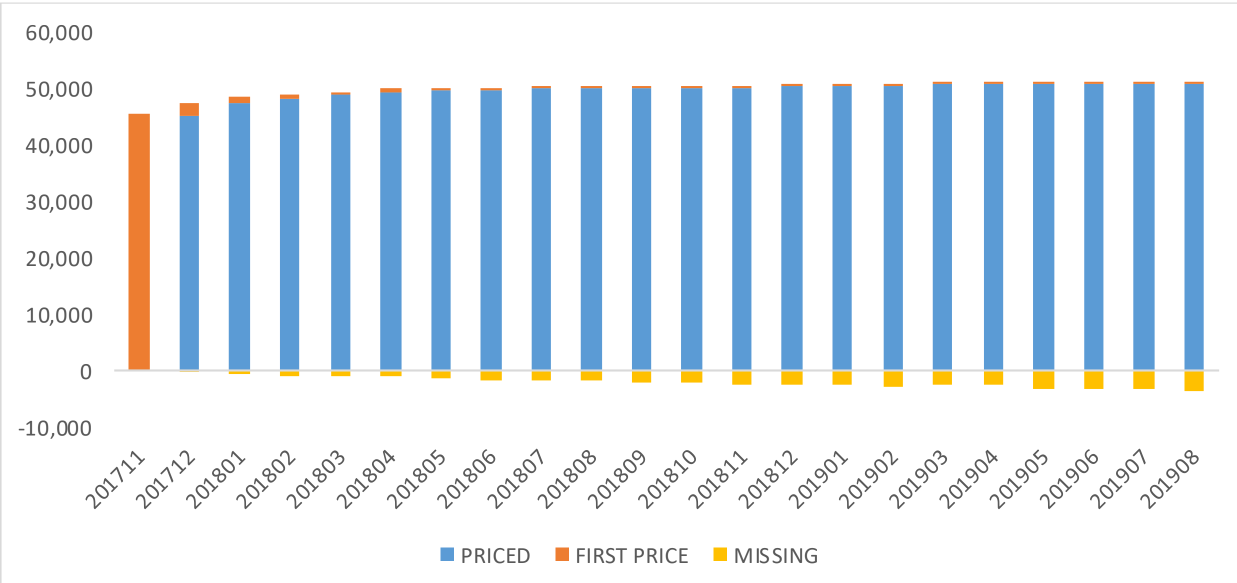

Table 5. Wilcoxon signed rank test, CPI for gasoline vs experimental indexes, Area level comparisons, Nov. 2017-Aug. 2019

Area Code	Area Name	Number of Observations	Probability
S49A	Los Angeles-Long Beach-Anaheim, CA	20	1
S11A	Boston-Cambridge-Newton, MA-NH	20	0.98544
S48A	Phoenix-Mesa-Scottsdale, AZ	20	0.98544
S49D	Seattle-Tacoma-Bellevue, WA	20	0.98544
N370	West South Central - Size Class B/C	20	0.95633
S12B	Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	20	0.95633
S49C	Riverside-San Bernardino-Ontario, CA	20	0.95633
N110	New England - Size Class B/C	20	0.92728
N240	West North Central - Size Class B/C	20	0.92728
S35B	Miami-Fort Lauderdale-West Palm Beach, FL	20	0.92728
S49G	Urban Alaska	20	0.92728
S23A	Chicago-Naperville-Elgin, IL-IN-WI	20	0.89832
S48B	Denver-Aurora-Lakewood, CO	20	0.89832
S37A	Dallas-Fort Worth-Arlington, TX	20	0.86949

S49F	Urban Hawaii	20	0.86949
S24B	St. Louis, MO-IL	20	0.84082
S35D	Tampa-St. Petersburg-Clearwater, FL	20	0.84082
S24A	Minneapolis-St. Paul-Bloomington, MN-WI	20	0.78413
S35A	Washington-Arlington-Alexandria, DC-VA-MD-WV	20	0.78413
N120	Middle Atlantic - Size Class B/C	20	0.75617
S23B	Detroit-Warren-Dearborn, MI	20	0.75617
S35C	Atlanta-Sandy Springs-Roswell, GA	20	0.75617
S12A	New York-Newark-Jersey City, NY-NJ-PA	20	0.72851
S35E	Baltimore-Columbia-Towson, MD	20	0.70118
N480	Mountain - Size Class B/C	20	0.67422
S49B	San Francisco-Oakland-Hayward, CA	20	0.67422
N230	East North Central - Size Class B/C	20	0.62151
N350	South Atlantic - Size Class B/C	20	0.62151
S49E	San Diego-Carlsbad, CA	20	0.43043
S37B	Houston-The Woodlands-Sugar Land, TX	20	0.36828
N499	Pacific Size Class B/C Other Than Urban Hawaii And Urban Alaska	20	0.32998
N360	East South Central - Size Class B/C	20	0.27736

Appendix 2: Price Movement Variation Within and Across Areas

We attempt to show that, on at least a cross-sectional basis, price movement varies within individual CPI areas. Our work builds off a previous study of price movements across a single CPI area. In *Does Location Matter? A Case-study in the Effect of Geography on Gasoline Price Inflation*, we see an in depth analysis of price movement across a single CPI area – the Washington DC metro area. The study employed a fixed-effects regression model to analyze price movements across all stations in the DC area. Independent variables included geographical considerations – e.g. county, home-value and population density of the surrounding station – as well as attributes of the stations themselves – e.g. previous month price level of a given station, distance from a station to its nearest neighbor, and brand. Ultimately, the work found that all of the above attributes were significant in explaining price change variation across the area *to a degree*; when indexes were constructed with samples stratified on the above variables, the relatives showed now significant differences from those of the baseline production index. Similarly, when including a significant fixed effect variable – the movement of the price of crude oil (i.e. the underlying commodity from which gasoline is derived) – the effect and significance of all the above attributes was greatly diminished.

Does Location Matter? gives us a straightforward conclusion; that at a given cross section in time, one can observe variation in gasoline price change across an area, and attribute such variation to the geography and station characteristics. However, all of these variations are conditional on movement in the price of the underlying crude oil, and fluctuation in oil price is the principal determinant of movement in gasoline prices. Thus, it validated the CPI's current practice of sampling at the PSU level.

Methodology

Our first objective is to replicate the findings in *Does Location Matter?,* at both the national level, and across all areas. We do this by performing a two-way fixed-effects ANOVA on set of price relatives from the crowd-sourced website and app, for each CPI area, as well as for the set of all CPI areas. The purpose of this is to observe whether cross-sectional price variation (i.e. variation among price changes in a given month) is more attributable to brand, or geography. We control for the fixed-effects of fuel grade (regular, midgrade or premium), and month of observation.

The national level ANOVA compares variation in price across each of the 32 CPI areas on the 2010 geographic structure, and across the top percentile of brands. The 61 brands found in the top percentile account for 81 percent of total observed stations; all other stations are grouped into an "OTHER" category. At the area level, we perform our ANOVA across counties within each area, and the top decile of brands.

Results

National Level

Our national level ANOVA examines price change variation from November 2017 to June 2019, and how much of it is attributable to brand and geography, as represented by CPI area. Both variables were shown to be significant in the model. However, when compared with our time fixed effect – i.e. the overall movement of gasoline prices in a given month – the portion of variation explained by both geography and brand is miniscule. This is consistent with findings in Does Location Matter?, which posited that brand and geography are significant in explaining price movement variation, but only conditional on the overall trend (i.e. commodity price movement) of gasoline price movements in a given month.

Table 6. Fixed Effects ANOVA of gas-station brand, and CPI area. Fixed effects control for month of observation (CP), and fuel grade

ANOVA Results -- National					
Var.	DF	Sum of Sq.	Mean Sq.	F-Stat	P-Value
Brand	61	1.697615	0.02783	25.45899	0.00E+00
Area	31	3.573651	0.115279	105.4587	0.00E+00
CP	20	963.244	48.1622	44059.36	0.00E+00
GRADE	2	0.111223	0.055612	50.87414	8.09E-23
Residuals	477300	521.7465	0.001093		

Area Level

We perform 31 fixed-effects two-way ANOVAs for each CPI Area (we exclude S49F, Urban Hawaii, because the area contains only a single county). Brand only explained variation in area level price movement to a significant degree in 2 of the 31 areas we tested. Geography, as denoted by county, explained price movement variation to a significant degree in 24 of the 31 areas. This once again confirms findings in *Does*

Location Matter, which found that at the area of a single area, geography played a larger role in determining price change variation than brand. Nevertheless, both variables – once again – explained only a miniscule portion of price change variation when compared to month of observation (CP). This is consistent with national level findings. Nevertheless, as the main report demonstrates, these variations are cross-sectional in nature – they are dependent on the period of observation. Constructing a weighted index controlling for these factors does not appear to preserve the variations, and does not appear to affect relatives.

Table 7. Fixed-effects two-way ANOVA results across brand (top decile by representation for each area), and counties within each area

Area Level ANOVAs						
AREA	**Variable**	**DF**	**Sum of Sq**	**Mean Sq.**	**F-Stat**	**P-Value**
N110	name	30	0.007577	0.000253	0.419357	0.99781
N110	COUNTY	5	0.006328	0.001266	2.10145	0.062097
N110	CP	19	62.62578	3.296094	5472.684	0
N110	GRADE	2	0.000422	0.000211	0.350151	0.704584
N110	Residuals	41535	25.01574	0.000602		
N120	name	66	0.02094	0.000317	0.620753	0.993375
N120	COUNTY	20	0.072617	0.003631	7.103886	2.12E-20
N120	CP	19	122.3643	6.440228	12600.54	0
N120	GRADE	2	0.000831	0.000416	0.813094	0.443487
N120	Residuals	100400	51.31515	0.000511		
N230	name	107	0.126652	0.001184	1.37932	0.005723
N230	COUNTY	49	0.177098	0.003614	4.211652	3.4E-21
N230	CP	19	644.1099	33.90052	39504.11	0
N230	GRADE	2	0.053912	0.026956	31.41172	2.29E-14
N230	Residuals	199776	171.4381	0.000858		
N240	name	49	0.038621	0.000788	1.056821	0.365831
N240	COUNTY	18	0.068274	0.003793	5.085803	7.87E-12
N240	CP	19	160.8229	8.464361	11349.28	0
N240	GRADE	2	0.020033	0.010016	13.4304	1.48E-06
N240	Residuals	44256	33.0064	0.000746		
N350	name	196	0.140458	0.000717	0.974544	0.587207
N350	COUNTY	60	0.299544	0.004992	6.789198	4.7E-53
N350	CP	19	575.8019	30.30536	41212.52	0
N350	GRADE	2	0.06501	0.032505	44.20378	6.39E-20
N350	Residuals	273666	201.2385	0.000735		
N360	name	118	0.103735	0.000879	1.174328	0.095037
N360	COUNTY	34	0.167743	0.004934	6.590352	8.16E-30
N360	CP	19	312.1272	16.42775	21944.33	0
N360	GRADE	2	0.026516	0.013258	17.71005	2.04E-08
N360	Residuals	122351	91.59321	0.000749		

N370	name	190	0.164336	0.000865	1.034579	0.356779
N370	COUNTY	45	0.216943	0.004821	5.76656	6.52E-32
N370	CP	19	455.8933	23.99439	28700.78	0
N370	GRADE	2	0.061641	0.030821	36.8659	9.84E-17
N370	Residuals	166603	139.2832	0.000836		
N480	name	27	0.020032	0.000742	0.902594	0.609713
N480	COUNTY	5	0.058061	0.011612	14.12687	7.72E-14
N480	CP	19	126.1662	6.640327	8078.337	0
N480	GRADE	2	0.012403	0.006202	7.544723	0.000529
N480	Residuals	53992	44.38098	0.000822		
N490	name	39	0.132977	0.00341	7.829786	7.41E-43
N490	COUNTY	11	0.167208	0.015201	34.90613	3.25E-75
N490	CP	19	69.84325	3.675961	8441.276	0
N490	GRADE	2	0.00711	0.003555	8.163142	0.000285
N490	Residuals	47031	20.4808	0.000435		
S11A	name	120	0.026081	0.000217	0.455219	1
S11A	COUNTY	8	0.013382	0.001673	3.50338	0.00047
S11A	CP	19	76.28721	4.015116	8409.481	0
S11A	GRADE	2	0.010606	0.005303	11.10681	1.5E-05
S11A	Residuals	87990	42.01093	0.000477		
S12A	name	110	0.136153	0.001238	2.113413	9.04E-11
S12A	COUNTY	28	0.329731	0.011776	20.10718	8.6E-101
S12A	CP	19	329.2951	17.33132	29592.43	0
S12A	GRADE	2	0.066001	0.033	56.34665	3.42E-25
S12A	Residuals	266705	156.2004	0.000586		
S12B	name	60	0.028414	0.000474	0.785581	0.886546
S12B	COUNTY	14	0.054642	0.003903	6.474511	2.92E-13
S12B	CP	19	135.1881	7.115161	11802.96	0
S12B	GRADE	2	0.035918	0.017959	29.79145	1.16E-13
S12B	Residuals	92589	55.81529	0.000603		
S23A	name	53	0.038936	0.000735	0.796698	0.855698
S23A	COUNTY	17	0.039549	0.002326	2.522919	0.000499
S23A	CP	19	500.1035	26.32124	28544.87	0
S23A	GRADE	2	0.088269	0.044134	47.86274	1.66E-21
S23A	Residuals	152568	140.683	0.000922		
S23B	name	24	0.010405	0.000434	0.471862	0.986497
S23B	COUNTY	7	0.018464	0.002638	2.870788	0.00537
S23B	CP	19	336.415	17.70605	19270.19	0
S23B	GRADE	2	0.018792	0.009396	10.22621	3.62E-05
S23B	Residuals	101603	93.35601	0.000919		
S24A	name	56	0.016542	0.000295	0.458744	0.99983

S24A	COUNTY	17	0.010019	0.000589	0.915217	0.555311
S24A	CP	19	179.3047	9.437087	14655.74	0
S24A	GRADE	2	0.004656	0.002328	3.615061	0.026921
S24A	Residuals	65856	42.40582	0.000644		
S24B	name	28	0.009423	0.000337	0.453576	0.994136
S24B	COUNTY	14	0.026468	0.001891	2.548176	0.00117
S24B	CP	19	243.9993	12.84207	17308.69	0
S24B	GRADE	2	0.020339	0.010169	13.70626	1.12E-06
S24B	Residuals	59346	44.03136	0.000742		
S35A	name	50	0.015978	0.00032	0.481469	0.999284
S35A	COUNTY	22	0.096801	0.0044	6.629229	3.03E-20
S35A	CP	19	129.0123	6.790122	10230.23	0
S35A	GRADE	2	0.027049	0.013525	20.37665	1.42E-09
S35A	Residuals	90686	60.19113	0.000664		
S35B	name	28	0.019032	0.00068	1.011957	0.446835
S35B	COUNTY	4	0.027468	0.006867	10.2237	2.84E-08
S35B	CP	19	145.4877	7.657247	11400.05	0
S35B	GRADE	2	0.004083	0.002042	3.039686	0.047854
S35B	Residuals	96374	64.73298	0.000672		
S35C	name	69	0.026504	0.000384	0.504918	0.999803
S35C	COUNTY	32	0.110523	0.003454	4.540088	2.3E-16
S35C	CP	19	377.9623	19.89275	26149	0
S35C	GRADE	2	0.01743	0.008715	11.45591	1.06E-05
S35C	Residuals	156445	119.0149	0.000761		
S35D	name	27	0.022911	0.000849	1.239644	0.182058
S35D	COUNTY	6	0.003401	0.000567	0.828016	0.547916
S35D	CP	19	160.3318	8.438518	12327.82	0
S35D	GRADE	2	0.023857	0.011928	17.42626	2.72E-08
S35D	Residuals	62803	42.98929	0.000685		
S35E	name	27	0.011703	0.000433	0.821186	0.728552
S35E	COUNTY	8	0.002912	0.000364	0.689554	0.701216
S35E	CP	19	102.6535	5.402817	10236.32	0
S35E	GRADE	2	0.026052	0.013026	24.67958	1.94E-11
S35E	Residuals	48123	25.39972	0.000528		
S37A	name	91	0.061505	0.000676	0.770325	0.949082
S37A	COUNTY	14	0.032232	0.002302	2.624001	0.000809
S37A	CP	19	649.8655	34.20345	38982.86	0
S37A	GRADE	2	0.05849	0.029245	33.33183	3.36E-15
S37A	Residuals	172318	151.1913	0.000877		
S37B	name	120	0.063549	0.00053	0.579739	0.999935
S37B	COUNTY	10	0.038373	0.003837	4.200814	7.49E-06

S37B	CP	19	543.6707	28.61425	31324.95	0
S37B	GRADE	2	0.11817	0.059085	64.68234	8.29E-29
S37B	Residuals	188560	172.243	0.000913		
S48A	name	20	0.021244	0.001062	2.182219	0.001684
S48A	COUNTY	1	0.001175	0.001175	2.413214	0.12032
S48A	CP	19	178.814	9.411261	19334.79	0
S48A	GRADE	2	0.042179	0.02109	43.32744	1.57E-19
S48A	Residuals	61137	29.7586	0.000487		
S48B	name	33	0.021202	0.000642	0.891308	0.646393
S48B	COUNTY	10	0.021274	0.002127	2.951297	0.001031
S48B	CP	19	191.3233	10.06965	13969.3	0
S48B	GRADE	2	0.02249	0.011245	15.60003	1.69E-07
S48B	Residuals	49333	35.56125	0.000721		
S49A	name	58	0.176096	0.003036	9.427161	1.74E-80
S49A	COUNTY	1	0.002372	0.002372	7.364526	0.006653
S49A	CP	19	339.8475	17.88671	55538.05	0
S49A	GRADE	2	0.017771	0.008885	27.58916	1.05E-12
S49A	Residuals	146359	47.13671	0.000322		
S49B	name	46	0.079069	0.001719	4.361058	3.72E-21
S49B	COUNTY	4	0.030217	0.007554	19.1659	9.09E-16
S49B	CP	19	92.72896	4.880472	12382.38	0
S49B	GRADE	2	0.006524	0.003262	8.276695	0.000255
S49B	Residuals	55528	21.88616	0.000394		
S49C	name	42	0.065265	0.001554	3.652198	1.42E-14
S49C	COUNTY	1	0.000222	0.000222	0.52173	0.470107
S49C	CP	19	155.1971	8.168268	19197.93	0
S49C	GRADE	2	0.007254	0.003627	8.524625	0.000199
S49C	Residuals	62535	26.60717	0.000425		
S49D	name	27	0.06596	0.002443	5.223659	2.25E-17
S49D	COUNTY	2	0.014724	0.007362	15.74218	1.46E-07
S49D	CP	19	78.85615	4.150324	8874.473	0
S49D	GRADE	2	0.00241	0.001205	2.576192	0.076072
S49D	Residuals	58349	27.28807	0.000468		
S49E	name	41	0.030645	0.000747	2.173952	2.05E-05
S49E	COUNTY	1	0.000433	0.000433	1.260329	0.261595
S49E	CP	19	97.60695	5.137208	14941.55	0
S49E	GRADE	2	0.007755	0.003877	11.27749	1.27E-05
S49E	Residuals	41916	14.41157	0.000344		
S49G	name	5	0.001761	0.000352	1.370503	0.232
S49G	COUNTY	1	0.002786	0.002786	10.83674	0.001
S49G	CP	19	15.64199	0.823263	3202.777	0

S49G	GRADE	2	3.64E-05	1.82E-05	0.070773	0.931674
S49G	Residuals	6472	1.663605	0.000257		

Table 8. T-test results comparing CPI Gasoline relatives from November 2017 - June 2019 (base period = 201710), for each CPI area

AREA	Estimate	T-Stat	P-value	DF	Low	High	Method	Alternative
N110	-5.3E-05	-0.39861	0.694625	19	-0.00033	0.000226	Paired t-test	two.sided
N120	0.000325	0.378424	0.709311	19	-0.00147	0.002122	Paired t-test	two.sided
N230	-0.00041	-0.42296	0.677076	19	-0.00243	0.001616	Paired t-test	two.sided
N240	-0.00018	-0.13263	0.895882	19	-0.00303	0.002665	Paired t-test	two.sided
N350	-0.00029	-0.43775	0.666506	19	-0.00165	0.001078	Paired t-test	two.sided
N360	-0.00026	-0.477	0.638802	19	-0.00139	0.000872	Paired t-test	two.sided
N370	0.000199	0.287335	0.776966	19	-0.00125	0.001647	Paired t-test	two.sided
N480	0.000377	0.132007	0.896367	19	-0.00561	0.006362	Paired t-test	two.sided
N490	0.000926	1.094459	0.287435	19	-0.00085	0.002698	Paired t-test	two.sided
S11A	2.76E-05	0.2927	0.772924	19	-0.00017	0.000225	Paired t-test	two.sided
S12A	-0.00026	-0.93376	0.362143	19	-0.00083	0.00032	Paired t-test	two.sided
S12B	0.000306	1.274874	0.217724	19	-0.0002	0.000808	Paired t-test	two.sided
S23A	-5.4E-05	-0.26993	0.790126	19	-0.00048	0.000367	Paired t-test	two.sided
S23B	-8.8E-05	-0.60307	0.553591	19	-0.00039	0.000216	Paired t-test	two.sided
S24A	0.000261	0.110673	0.913036	19	-0.00467	0.005196	Paired t-test	two.sided
S24B	3.38E-05	0.074863	0.941106	19	-0.00091	0.00098	Paired t-test	two.sided
S35A	6.7E-05	0.162898	0.872319	19	-0.00079	0.000928	Paired t-test	two.sided
S35B	-1.6E-05	-0.59737	0.557311	19	-7.2E-05	4E-05	Paired t-test	two.sided
S35C	-0.00038	-1.38418	0.182347	19	-0.00094	0.000192	Paired t-test	two.sided
S35D	-0.00061	-0.86059	0.400194	19	-0.00211	0.00088	Paired t-test	two.sided
S35E	-0.00013	-0.12606	0.90101	19	-0.00227	0.002014	Paired t-test	two.sided
S37A	2.37E-05	0.233814	0.81763	19	-0.00019	0.000236	Paired t-test	two.sided
S37B	4.65E-05	0.616919	0.544611	19	-0.00011	0.000204	Paired t-test	two.sided
S48A	3.37E-05	1.022728	0.319283	19	-3.5E-05	0.000103	Paired t-test	two.sided
S48B	-0.00311	-3.60359	0.001893	19	-0.00492	-0.0013	Paired t-test	two.sided
S49A	-1.2E-06	-0.58326	0.566577	19	-5.6E-06	3.18E-06	Paired t-test	two.sided
S49B	-7.8E-05	-0.64706	0.525333	19	-0.00033	0.000175	Paired t-test	two.sided
S49C	-1.7E-05	-0.30439	0.764139	19	-0.00014	0.000102	Paired t-test	two.sided
S49D	8.63E-05	0.433131	0.669797	19	-0.00033	0.000503	Paired t-test	two.sided
S49E	-0.00023	-0.39244	0.699101	19	-0.00148	0.001012	Paired t-test	two.sided
S49F	0	#NUM!	#NUM!	19	#NUM!	#NUM!	Paired t-test	two.sided
S49G	1.9E-05	0.443654	0.662304	19	-7.1E-05	0.000109	Paired t-test	two.sided